

Executive Summary

Updated Recommendations to the State Board of Education for a State Accountability Index

January 2009

Pete Bylsma, EdD, MPA

This summary about the proposed accountability index includes updated policies that reflect recent stakeholder input. Revised results have been generated using these policies and are included in this document. Appendix A provides a detailed description of the policy changes. The full document about the index is available on the State Board of Education Web site.

CREATING THE ACCOUNTABILITY INDEX

The legislature requires the State Board of Education to develop a statewide accountability system that will help improve academic performance among all students in the state. Part of that requirement is to “adopt objective, systematic criteria” to identify schools and districts for recognition and for receiving additional state support. To meet this requirement, the Board has developed 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 supporting continuous improvement efforts. Schools and districts in most need are given “Priority” status, making them eligible to receive more significant 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 two years.

Various principles guided the development of the index. The accountability system will (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) focus at both the school and district levels, (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, and (12) provide multiple ways to reward success.

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 students from non-low income families, low-income students, and by all students compared to similar schools (controlling for the percentage of students who are learning English, have a disability, live in low-income homes, and are mobile), and the level of improvement in the achievement of all students 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					

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 all the tested grades). **The index is the simple average of all 20 ratings.** The index ranges from 1.0 to 7.0. Table 2 shows how each of the five outcomes are measured using the four indicators and the benchmarks that produce the ratings. Tier assignments are determined based on the index score. Schools and districts fall into five tiers, 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 6th (Priority) tier.

Table 2: Benchmarks and Ratings for Outcomes and Indicators

	READING	WRITING	MATH	SCIENCE	EXT. GRAD. RATE ¹
ACHIEVEMENT (NON-LOW INCOME)	<u>% MET STANDARD</u>				<u>RATE</u>
	<u>RATING</u>				<u>RATING</u>
ACHIEVEMENT (LOW INCOME)	90.1 - 100%				> 95
	80.1 - 90%				90.1 - 95%
	70.1 - 80%				85.1 - 90%
	60.1 - 70%				80.1 - 85%
	50.1 - 60%				75.1 - 80%
	40 - 50%				70 - 75%
	< 40%				< 70%
ACHIEVEMENT VS. PEERS²	<u>DIFFERENCE IN LEARNING INDEX</u>				<u>DIFFERENCE IN RATE</u>
	<u>RATING</u>				<u>RATING</u>
	> .20.....				> 12
	.151 to .20.....				6.1 to 12
	.051 to .15				3.1 to 6
	-.05 to .05				-3 to 3
	-.051 to -.15				-3.1 to -6.....
	-.151 to -.20				-6.1 to -12.....
IMPROVEMENT³	<u>CHANGE IN LEARNING INDEX</u>				<u>CHANGE IN RATE</u>
	<u>RATING</u>				<u>RATING</u>
	> .15.....				> 6
	.101 to .15				4.1 to 6
	.051 to .10.....				2.1 to 4
	-.05 to .05				-2 to 2
	-.051 to -.10				-2.1 to -4.....
	-.101 to -.15				-4.1 to -6.....
< -.15				< -6.....	

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

¹ This outcome only applies to schools and districts that are authorized to graduate students.

² This indicator adjusts the outcomes using statistical methods (multiple regression) to control for four student characteristics beyond a school's control: the percentage of low-income, ELL, special education, and mobile students. (Mobile students are those who are not continuously enrolled from October 1 through the testing period.) Scores are the difference between the actual level and the predicted level. 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). District calculations also control for the level of current expenditures.

³ Measured in terms of the change from the previous year.

The proposed index does not include AYP results. Feedback from all stakeholders revealed a lack of confidence in the current AYP results for accountability purposes. The index is *more valid* because it is based on the performance of all students in more subjects, is more differentiated than a “Yes/No” system, does not count students multiple times, and addresses several unintended consequences created by the current system. The index is *more inclusive* because it uses a smaller minimum number for reporting (10 students across the entire school/district), includes the results of all students, regardless of how long they have been attending school or district, includes both writing and science (this helps prevent a narrowing of the curriculum), and uses the Learning Index to measure performance across the range of assessment results (this reduces the focus on 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. Finally, 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 for reporting.

INITIAL RESULTS

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.

Table 3: Tier Ranges and Preliminary Results (2007)

Tier	Index Range	# of Schools	% of Schools	# of Students¹	# of Districts	% of Districts	# of Students¹
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
Acceptable	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 (eligible for Innovation Zone) ²	1.00 – 2.99	TBD	TBD	TBD	TBD	TBD	TBD

¹Approximate number (some schools did not provide enrollment data).

²To be determined after in-depth analyses of the data and local conditions of those in the struggling tier.

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 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 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 3.45 is the

index score, which puts the school in the middle of the Acceptable tier. The average ratings are color-coded, and a set of “stars” indicate the rating so the overall results can be seen at a glance. These types of results could be made public on the state Web site (the format for presenting the results must still be determined). Results presented in this “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.

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
Non-low inc. ach.	*****	*****	***	*	*****	
Low-inc. ach.	****	****	*	*	*****	
Ach. vs. peers	**	**	**	**	*****	
Improvement	*	****	*	****	*****	

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). A deeper analyses would also occur for districts that have an index number in the Struggling tier to determine if they merit receiving extra support.

Other tables and charts can illustrate school and district results as well. For example, Figure 1 shows the distribution of the number of schools in each tier for an actual district. Table 5 shows all the results across multiple years in a hypothetical district.

Figure 1: Accountability Results in “Actual” District, 2007

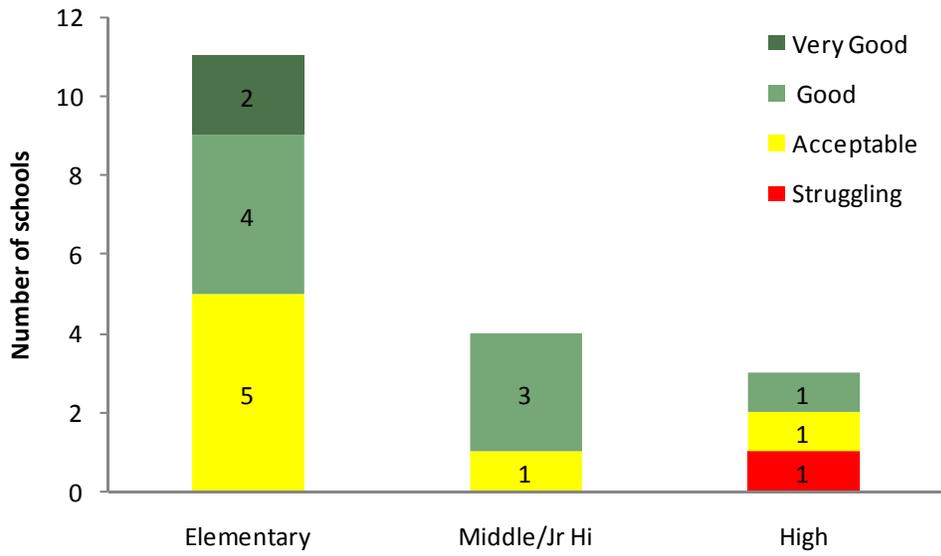


Table 5: Showing Accountability Results Over Time (Hypothetical District)

<i>Indicator/Outcome</i>	<i>YEAR</i>			
	2004	2005	2006	2007
Non-low inc. ach.	2.75	3.00	3.20	3.80
Reading	****	*****	*****	*****
Writing	****	****	****	*****
Math	**	**	**	****
Science	*	*	*	*
Ext. grad. rate	NA	***	****	****
Low-income ach.	2.00	2.00	2.40	2.40
Reading	***	*****	*****	*****
Writing	***	***	*****	*****
Math	*	*	*	**
Science	*	*	*	*
Ext. grad. rate	NA	*	**	*
Ach. vs. peers	4.00	4.00	4.00	4.00
Reading	****	****	****	****
Writing	****	****	****	****
Math	****	****	****	****
Science	****	****	****	****
Ext. grad. rate	NA	****	****	****
Improvement	5.67	5.25	4.60	3.60
Reading	*****	*****	****	****
Writing	NA	*****	*****	****
Math	*****	*****	****	****
Science	*****	*****	****	****
Ext. grad. rate	NA	NA	*****	**
INDEX	3.47	3.47	3.55	3.45

HANDLING SPECIAL CASES

The accountability system needs to be flexible enough to accommodate some special situations. These include holding alternative schools accountable, possibly excluding some ELL results from the index calculations, and not counting 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). 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. Moreover, alternative schools may need to be held accountable through more than just an index score because in many cases, they have intentionally targeted student populations facing significant challenges.

Two options should be considered for holding alternative schools accountable. First, these schools would receive the normal index score using the calculations used for all schools. Alternative schools will likely be over-represented in the Struggling tier, and these schools are examined more closely (see the next section) to determine if they need additional support. A second approach is to have the alternative schools that serve highly mobile and at-risk students (including those in correctional facilities) volunteer to be held accountable using other outcomes over an extended period of time (e.g., 3 years). This approach is used in California. For example, a school could choose to be evaluated by the number of credits earned, attendance rates, and gains on pre-post tests given during the year. This system is very complex, but it is viewed in California as being a more valid system for these types of schools. OSPI could use concepts in this approach in its analysis of alternative schools that fall in the Struggling tier.

Results for ELL students are currently included in AYP determinations in the student's second year of enrollment in a U.S. public school. OSPI has requested that ELL results not be included until an ELL student has been enrolled in a U.S. public school for three years or until an ELL student achieves an intermediate level of English proficiency on the WLPT, whichever comes first.¹ This request is based on research that shows it takes many years for an ELL student to acquire "academic" proficiency in English, the state assessments are given entirely in English, no translated versions are administered, and the students must be able to read and write English in order to understand and respond to the test items. Moreover, testing these students in English violates widely-adopted testing standards and ethics because of threats to validity and mistreatment of human subjects. However, the U.S. Education Department has denied OSPI's request to change the way ELL students are included in AYP calculations.

Nevertheless, computations for the proposed state accountability system could exclude the results for ELL students who had not achieved intermediate English proficiency (Level 3 composite) on the WLPT or for three years in a U.S. public school, whichever comes first, whenever a test requires reading and writing in only English. WLPT results would also need to be made public on the OSPI Report Card, which is not current OSPI practice. This would provide more accountability for progress among ELL students. This policy would still include the results of a very large percentage of ELL students. About 70% of ELL students statewide enter school in kindergarten, and they will have attended school for three years before taking the state assessment for the first time in grade 3. Of the ELL students who were enrolled in grades assessed by the WASL/WAAS (grades 3-8 and 10), more than 81% had reached the intermediate level of the WLPT in 2008 and would have their scores included in the accountability calculations. While very few ELL students would have their results excluded, this policy would increase the fairness and validity of the accountability results.

Schools and districts that perform at very high levels are not able to improve much more. To avoid "penalizing" these schools for a lack of improvement, the ratings for this indicator will not be included in the index calculations under certain conditions. Without this policy, schools and 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. Specifically, a school or district can request that the improvement indicator not be used to compute the accountability index when a Learning Index reaches 3.85 out of 4.00 and remains at or above that level for two consecutive years. (A

¹ The composite score from the annual Washington Language Proficiency Test (WLPT) reflects proficiency in reading, writing, speaking, and listening.

school or district needs to improve by more than .15 to receive a rating of 7, which is impossible when their Learning Index is at or greater than 3.85.) The first year the Index falls above 3.85, the school/district would get a rating based on its improvement. If the Index stays at or above 3.85, the maximum rating is not possible, so the indicator would not be calculated.² The same policy applies to the extended graduation rate outcome. A school or district can request the extended graduation rate results not be used when computing the accountability index when the rate reaches or exceeds 94% and remains at that level for two consecutive years (the graduation rate must improve by more than 6 percentage points to earn a rating of 7). Of the schools with graduation data, 11% had a rate that was at least 94% in two consecutive years.

IDENTIFYING “PRIORITY” SCHOOLS AND DISTRICTS (LOWEST TIER)

Various quantitative and qualitative data will be used to determine which schools and districts that fall in the Struggling tier should be placed in the Priority tier and be eligible to receive significant support. The data fall in four categories:

- **Contextual Data:**
 - Type of school
 - Changes in student population
 - Programs served by the school
 - Level of student mobility
- **Assessment Results (WASL/WAAS/WLPT)**
 - 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
 - Problems with data that generate the index (e.g., reporting errors related to graduates)

Each year, the process begins when OSPI computes the index 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. This reduces the number of schools and districts that require a deeper analysis. OSPI staff then review the index results for each school and district in the Struggling tier and sort them into two categories:

² 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), no district reached this level in 2006, and no school or district reached this level in any subject in 2007.

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

(1) Schools/districts that *remain in the struggling tier* are those that have not been in this 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 (and additional information for high schools and districts using August results).

OSPI staff then conduct a deeper analysis using available data for the schools and districts that have the possibility of being placed in the Priority tier. This may require contacting the district and/or local ESD to get more information. Based on this review, schools and districts are sorted again into the same two categories. Those placed in the Priority tier are notified of the possible designation and the reasons why this designation is possible. The district/school is given a chance to avoid the Priority designation by providing more information, including what explains the low index results. Appeals are made to OSPI with local school board approval. OSPI then reviews the additional information, and then recommend a final Priority list to the State Board of Education for review and approval. Schools and districts that are placed in the Priority tier would be offered additional state support, which would be tailored to meet their specific needs.

INTEGRATING THE SYSTEMS

Federal law requires states to have a single accountability system. Many states combine their state accountability system with the federal NCLB system. The State Board of Education will work with OSPI in the coming months to prepare a unified accountability plan for the U.S. Education Department that will recommend using the proposed state accountability system in place of the current federal system. A new administration may provide more flexibility to states that design alternative systems. The proposed system has many desirable features that could make it a viable alternative to the current rules used to measure AYP. A new method for determining AYP and what constitutes being “in improvement” still must be determined, and the consequences of not making AYP could remain the same or be different.

The assessment and participation results will continue to be disaggregated for all student subgroups and be made public, as required by federal law. These disaggregated data will be used in the process of determining which schools and districts are in need of improvement and what type and level of support should be provided by the state.

RECOGNITION

The Board intends to provide recognition based on sustained exemplary performance. The accountability system will provide multiple ways to reward success and will rely on criterion-referenced measures using the results from the accountability matrix.

Three options can be considered: providing recognition for (1) each of the 30 cells of the matrix, (2) the 20 “inner” cells of the matrix, and (3) the 10 “average” cells of the matrix. The advisors recommended providing recognition in all 30 cells because they believe people are motivated to improve the most when they can experience success. A minimum rating is required to receive recognition—5.50 in the 20 “inner” cells and 5.25 in the “averaged” cells (see Table 6). Any cell with a rating of 6.00 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 recognition is given only for sustained exemplary performance.

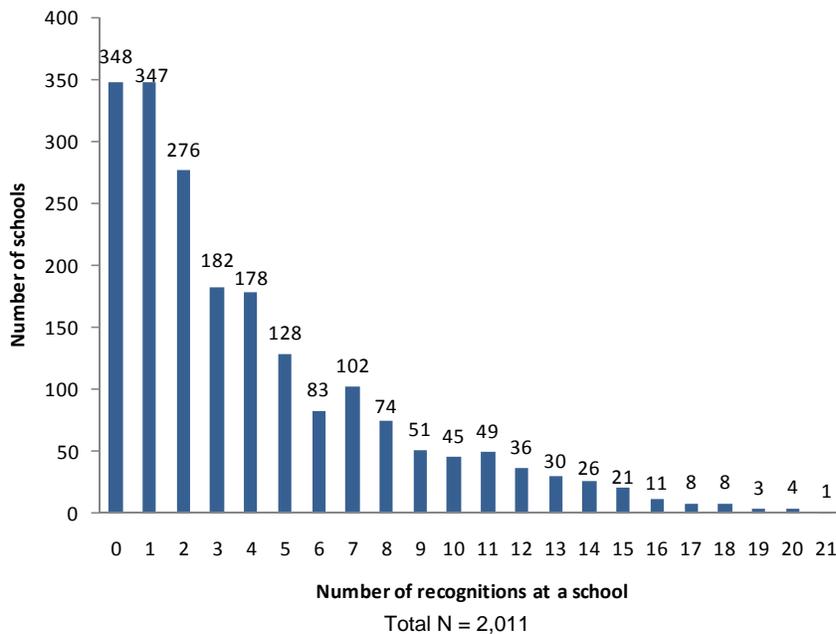
Table 6: Minimum 2-Year Average Required for Recognition**

Indicator	Reading	Writing	Math	Science	Grad Rate	Average
Non-low inc. ach.	5.50	5.50	5.50	5.50	5.50	5.25
Low-inc. ach.	5.50	5.50	5.50	5.50	5.50	5.25
Ach. vs. peers	5.50	5.50	5.50	5.50	5.50	5.25
Improvement	5.50	5.50	5.50	5.50	5.50	5.25
Average	5.25	5.25	5.25	5.25	5.25	5.25

**Any cell of the matrix with a 2-year average rating of 6.00 or above would be recognized “with honor.”

Figure 2 shows the number of schools that would have received awards if the proposed system was in place in 2007 for the 2,011 schools and all 30 cells were eligible to receive recognition. The largest number of schools would have received recognition in just one or two of the 30 areas, and 348 schools (17%) would not have received any recognition. At the other extreme, 242 schools (12%) would have received recognition in 10 or more areas, and one school would have received recognition in 21 of the 30 cells of the matrix. The largest number of schools (57%) met the criteria for reading achievement among their non-low income students. Achievement in math, science, and among low-income students had far fewer schools meeting the minimum criteria. Less than 4% had an overall average of 5.25 on the accountability index over the 2-year period. Of the 149 schools that had a 2-year index average of less than 2.50 (i.e., those in the Struggling tier), 71% would not have received any recognition in any of the 30 cells, 22% would have received recognition in one cell, and 7% would have received recognition in two or three areas (most often in writing improvement).

Figure 2: Number of Schools with Recognition, by Number of Recognitions (2007)



This system of recognition will supplement and could replace some types of recognition currently in place. The federal government provides funding to schools for three awards, primarily those

receiving Title I funds. OSPI also provides awards for improvement but no extra funding as part of its recognition. Schools and districts that receive recognition in the proposed system will not be compensated monetarily, except possibly for schoolwide bonuses for all school staff based on “multiple measures of student performance.” These bonuses have been recommended by the Basic Education Finance Task Force in its December 2008 report to the Legislature. The proposed recognition system could be used as a basis for these rewards. For example, staff in schools that have a 2-year average in overall improvement of at least 5.25 could be given the schoolwide bonus. In 2007, about 12% of the schools statewide met this criterion. 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 will need to remain flexible. It must adapt to changes in NCLB and graduation requirements, the assessment system, and other factors that may impact the results. Moreover, a number of issues must still be resolved before the index can be implemented effectively. For example, further review of the results should occur to ensure the index measures the achievement and improvement the Board intends. Various OSPI and State Board 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 measuring improvement and holding alternative schools accountable need further review. Finally, some method of measuring community and legislative support needs to be incorporated to ensure “reciprocal accountability.”

APPENDIX A

RECOMMENDED CHANGES IN THE ACCOUNTABILITY INDEX

This appendix lists eight recommended changes to the accountability index, the reasons for the changes, and the effect the changes have on the proposed system and results. Changes 1-4 have been made and are reflected in the Executive Summary; changes 5-8 are presented for Board consideration.

1. Change the first indicator, achievement by *all* students, to be achievement by *non-low income* students.

Rationale: Using results for non-low income students separate from those from low income families (used as the second indicator) means no student is double counted.⁴ In the original indicator using “all” students, low-income students end up being counted in both indicators. The suggested change was suggested by some stakeholders and reflects the belief that all students have equal value and no group of students is more important than any other group. This policy will reveal more clearly the size of the achievement gap based on family income level, which is the strongest predictor of student achievement. It would also reveal when low-income students perform *better* than their better-off classmates.

Effect of the Change: This change has relatively little impact on higher SES schools and districts because the “all” students group does not include many low-income students. It has a larger impact on those with higher percentages of low-income students because it would not double-count as many students, who tend to perform and graduate at lower levels. A comparison of the results using the original “all students” indicator and the “non-low income” indicator shows the index a small increase in the index, with slightly larger increases (in the .2 to .3 range on a 7-point scale). The overall effect is relatively small because this indicator counts for only 25% in the index. Finally, the policy would reduce the relationship between the index and SES to even lower levels.

2. Change the scale from 5 points (0-4) to 7 points (1-7).

Rationale: This change provides greater differentiation, or “spread”, to the results (like a + or – when giving letter grades). It also avoids the comparison with the grade point average. Several comments were made about the need for more points on the scale and starting the ratings at 1 rather than at 0.

Effect of the Change: Changes were made to the benchmarks, it is more difficult to achieve the highest ratings, and there are more tiers.

Table A1 shows the new benchmarks and ratings using the 7-point scale, which can be compared to the original set using a 5-point scale, shown in Table A2. A few minor changes were made in the highest and lowest benchmarks (see highlighted numbers).

⁴ Under the current AYP rules, student results are reported for all students as well as in 8 subgroups: the 5 race/ethnic groups (American Indian, Asian/Pacific Islanders, African American, Hispanic, and White), ELL students, low-income students, and students with disabilities. As a result, some students are counted as many as five times, while others are counted only once or twice. These 9 groups are measured in terms of their achievement and “participation rate” in reading and math, and recent federal regulations required accountability for each of the groups for the graduation rate. If any group does not meet the annual grade-level target, the school or district does not make AYP.

Table A1: Revised Benchmarks and Ratings (7-pt scale)

	READING	WRITING	MATH	SCIENCE	EXT. GRAD. RATE	
ACHIEVEMENT (NON-LOW INC.)	<u>% MET STANDARD</u>			<u>RATING</u>	<u>RATE</u>	<u>RATING</u>
	90.1 - 100%			7	> 95	7
	80.1 - 90%			6	90.1 - 95%	6
	70.1 - 80%			5	85.1 - 90%	5
ACHIEVEMENT (LOW INCOME)	60.1 - 70%			4	80.1 - 85%	4
	50.1 - 60%			3	75.1 - 80%	3
	40 - 50%			2	70 - 75%	2
	< 40%			1	< 70%	1
ACHIEVEMENT VS. PEERS	<u>DIFFERENCE IN LEARNING INDEX</u>			<u>RATING</u>	<u>DIFFERENCE IN RATE</u>	<u>RATING</u>
	> .20			7	> 12	7
	.151 to .20			6	6.1 to 12	6
	.051 to .15			5	3.1 to 6	5
	-.05 to .05			4	-3 to 3	4
	-.051 to -.15			3	-3.1 to -6	3
	-.151 to -.20			2	-6.1 to -12	2
	< -.20			1	< -12	1
IMPROVEMENT	<u>CHANGE IN LEARNING INDEX</u>			<u>RATING</u>	<u>CHANGE IN RATE</u>	<u>RATING</u>
	> .15			7	> 6	7
	.101 to .15			6	4.1 to 6	6
	.051 to .10			5	2.1 to 4	5
	-.05 to .05			4	-2 to 2	4
	-.051 to -.10			3	-2.1 to -4	3
	-.101 to -.15			2	-4.1 to -6	2
	< -.15			1	< -6	1

Table A2: Initial Benchmarks and Ratings (5-pt scale)

	READING	WRITING	MATH	SCIENCE	EXT. GRAD. RATE	
ACHIEVEMENT (ALL STUDENTS)	<u>% MET STANDARD</u>			<u>RATING</u>	<u>RATE</u>	<u>RATING</u>
	86-100%			4	> 95	4
	70-85.9%			3	85-95%	3
ACHIEVEMENT (LOW INCOME)	55-69.9%			2	75-84.9%	2
	40-54.9%			1	65-74.9%	1
	< 40%			0	< 65%	0
ACHIEVEMENT VS. PEERS	<u>DIFFERENCE IN LEARNING INDEX</u>			<u>RATING</u>	<u>DIFFERENCE IN RATE</u>	<u>RATING</u>
	> .20			4	> 12	4
	.10 to .20			3	5.01 to 12	3
	-.099 to .099			2	-5 to 5	2
	-.20 to -.10			1	-5.01 to -12	1
	< -.20			0	< -12	0
IMPROVEMENT	<u>CHANGE IN LEARNING INDEX</u>			<u>RATING</u>	<u>CHANGE IN RATE</u>	<u>RATING</u>
	> .12			4	> 6	4
	.051 to .12			3	3.01 to 6	3
	-.05 to .05			2	-3 to 3	2
	-.051 to -.12			1	-3.01 to -6	1
	< -.12			0	< -6	0

3. Change from 4 initial tiers to 5 initial tiers (before deeper analysis identifies those that should enter the Priority tier) and adjust the tier ranges accordingly.

Rationale: These changes provide greater differentiation and is more aligned with the revised 7-point rating scale.

Effect of the Change: One more tier was added. “Very Good” was placed between “Exemplary” and “Good” tiers. The percentage of schools and districts in most of the tiers remained about the same (Very Good and Good schools represent roughly the same percentage of Good schools in the original rating system).

Tables A3 and A4 show the revised and original set of tiers, their range, and the school results for 2007. Tables A5 and A6 show the revised and original results for districts for 2007. (The N is lower in the revised results because no schools and districts with less than 10 assessed students received an index score.)

Table A3: Revised Tier Ranges and 2007 School Results (N=2,011)

Tier	Index Range	Number of Schools	Percent of Schools
Exemplary	5.50 – 7.00	81	4.0%
Very Good	5.00 – 5.49	131	6.5%
Good	4.00 – 4.99	591	29.4%
Acceptable	2.50 – 3.99	980	48.7%
Struggling	1.00 – 2.49	228	11.3%

Note: Schools averaged 3.71, with 4.00 being the mid-point on a 7-point scale.

Table A4: Original Tier Ranges and 2007 School Results (N=2,046)

Tier	Index Range	Number of Schools	Percent of Schools
Exemplary	3.00 – 4.00	72	3.5%
Good	2.00 – 2.99	664	32.5%
Acceptable	1.00 – 1.99	1,043	51.0%
Struggling	0.00 – 0.99	267	13.0%

Table A5: Revised Tier Ranges and 2007 District Results (N=291)

Tier	Index Range	Number of Districts	Percent of Districts
Exemplary	5.50 – 7.00	1	.3%
Very Good	5.00 – 5.49	9	3.1%
Good	4.00 – 4.99	87	29.9%
Acceptable	3.00 – 3.99	177	60.8%
Struggling	1.00 – 2.49	17	5.8%

Table A6: Original Tier Ranges and 2007 District Results (N=296)

Tier	Index Range	Number of Districts	Percent of Districts
Exemplary	3.00 – 4.00	3	1.0%
Good	2.00 – 2.99	102	34.5%
Acceptable	1.00 – 1.99	175	59.1%
Struggling	0.00 – 0.99	16	5.4%

4. Change the recognition criteria to align with the 7-point rating scale.

Rationale: This change adjusts the recognition criteria to be in alignment with the revised 7-point rating scale. The criteria were adjusted to produce similar results as the original criteria based on the 5-point scale (in both cases, ratings in the top 20-21% of the scale would receive recognition).

Effect of the Change: The number of schools and districts being recognized stays about the same.

Tables A7 and A8 show the revised criteria and the original criteria for recognition. Figure A1 shows the number of recognitions a school would receive using the revised and original criteria and rating scales.

Table A7: Revised Minimum Requirements for Recognition (1-7 Scale)

Indicator	Reading	Writing	Math	Science	Ext. Grad. Rate	Average
Non-low inc. ach.	5.50	5.50	5.50	5.50	5.50	5.25
Low-inc. ach.	5.50	5.50	5.50	5.50	5.50	5.25
Ach. vs. peers	5.50	5.50	5.50	5.50	5.50	5.25
Improvement	5.50	5.50	5.50	5.50	5.50	5.25
Average	5.25	5.25	5.25	5.25	5.25	5.25

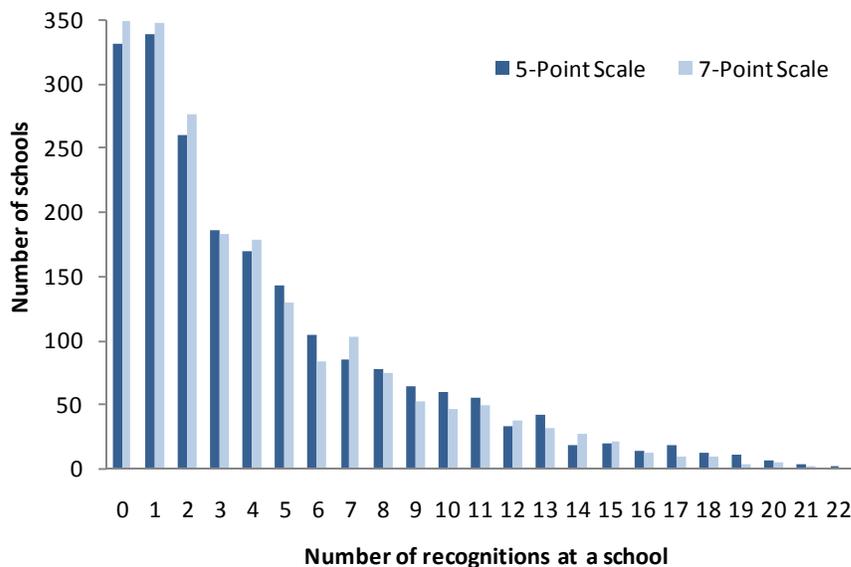
Any cell of the matrix with a 2-year average rating of 6.00 or above would be recognized “with honor.”

Table A8: Original Minimum Requirements for Recognition (0-4 Scale)

Indicator	Reading	Writing	Math	Science	Ext. Grad. Rate	Average
Non-low inc. ach.	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
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
Average	2.75	2.75	2.75	2.75	2.75	2.75

Any cell of the matrix with a 2-year average rating of 3.50 or above would be recognized “with honor.”

Figure A1: Number of Schools Receiving Recognition, 5-Point and 7-Point Scales



5. Propose exempting ELL results in the first 3 years of enrollment or until acquiring intermediate proficiency in English, whichever comes first.

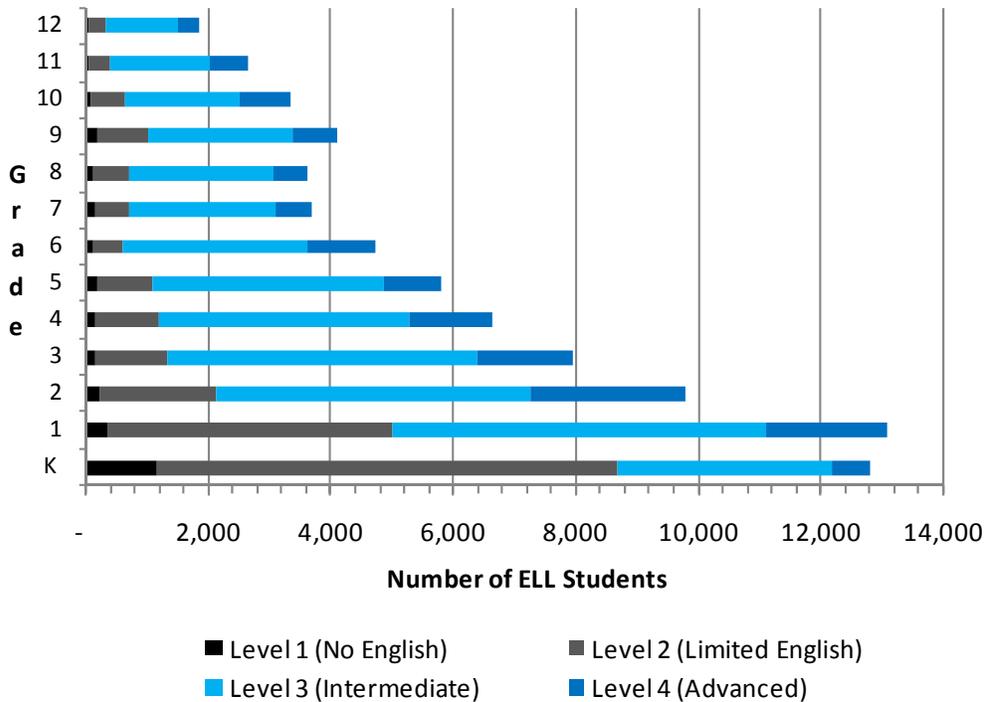
Rationale: Results for ELL students who are in their second year of enrollment in a U.S. public school are currently included in AYP calculations. OSPI has requested that ELL results be counted using the proposed policy. This request is based on research that shows it takes many years for an ELL student to acquire “academic” proficiency in English, the state assessments are given entirely in English, no translated versions are administered, and the students must be able to read and write English in order to understand and respond to the test items. Moreover, testing these students in English violates widely-adopted testing standards and ethics because of threats to validity and mistreatment of human subjects. However, the U.S. Education Department has denied OSPI’s request to change the way ELL students are included in AYP calculations.

Several stakeholders voiced strong concern about including ELL results in the index calculations using the current federal requirements. They believe the current policy supports inhumane assessment practices, produces invalid results, and has a negative impact on the acquisition of English language proficiency. Computations in the state accountability system can exclude the results for ELL students who had not achieved intermediate English proficiency (Level 3 composite) on the WLPT or for three years in a U.S. public school, whichever comes first, whenever a test requires reading and writing in only English.⁵ Although research has shown it takes longer than three years to acquire proficiency in English in an *academic* setting, this time period reflects OSPI’s position in its response to the federal regulations, and it provides motivation to help ELL students acquire English language skills. WASL testing would still take place after the ELL’s first year of enrollment, but the results would not count for accountability purposes until the student met one of the two criteria. WLPT results would be made public on the OSPI Report Card (this is not current OSPI practice) to provide more accountability for progress among ELL students.

Expected Effect of the Change: The effect of this policy will be rather small, and it will increase the validity of the index results while proposing sound assessment practices. The results of a very large percentage of ELL students will still be included because about 70% of all ELL students statewide enter school in kindergarten, and they will have attended school for three years before taking the state assessment for the first time in grade 3. Of the ELL students who were enrolled in the grades assessed by the WASL/WAAS (grades 3-8 and 10), more than 81% had reached the intermediate level of the WLPT in 2008 and would have their scores included in the accountability calculations (see Figure A2). Sensitivity analysis using data from a large district with a high percentage of ELL students found that this policy created little change (less than .2) in the district’s accountability index. So while very few ELL students would have their results excluded, this policy would increase both the actual and perceived fairness of the accountability results.

⁵ The Washington Language Proficiency Test (WLPT-II) is currently the only assessment Washington State uses to assess English language comprehension for English language learners (ELLs). This holistic test is used to determine one composite English language proficiency score. The composite score from the annual WLPT-II reflects proficiency in reading, writing, speaking, and listening.

Figure A2: WLPT Results in 2008



6. Propose using other means for holding alternative schools accountable.

Rationale: Many different types of alternative schools exist throughout the state. More than half the “schools” with this designation serve at-risk students in grades 9-12. Some schools are considered “alternative” because they do not use a normal school approach. 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 schools with this designation target special student populations (e.g., special education, gifted, ELL), and some are held on college campuses or at night. Given this diversity, no “peer” indicator is computed for these schools.

While most of these schools are relatively small (their total 2007 enrollment was less than 4% of enrollment statewide), many of them purposely serve student populations facing significant challenges. As a result, some alternative schools have a very low index and are more likely to fall in the “struggling” tiers. Consequently, alternative schools may need to be held accountable through more than just an index score.

- *Option 1:* Use the regular process for computing the index, then use the in-depth examination of data for schools in the Struggling tier. OSPI would be responsible for determining if an alternative school was following best practices and showing progress and therefore not be placed in the Priority tier. The role, status, and available resources of alternative programs within the district are important factors to be examined during this process.
- *Option 2:* Allow the schools that serve high-risk and special populations to use additional measures when determining their tier. This approach is similar to what is used in California. Alternative schools in California that serve highly mobile and at-risk students (including those

in correctional facilities) can volunteer to be evaluated using three other outcomes over at least a 3-year period. For example, a school could choose to be evaluated by the number of credits earned, attendance rates, and gains on pre-post tests given during the year. This system has become very complex, but it has been approved for federal accountability purposes and is viewed as being a more valid system for holding these types of schools accountable. OSPI could use concepts in this approach in its analysis of alternative schools that fall in the Struggling tier.

Expected Effects: Adopting additional steps when assessing alternative schools would provide greater validity and accuracy when making accountability decisions. Option 1 would be simpler and easier to implement than Option 2. Also, Option 2 gives schools more chance to pick a narrow set of outcomes that could be more easily attained.

7. Propose giving schools and districts the option to exclude the improvement indicator when they are performing at the highest achievement levels.

Rationale: Schools and districts that perform at very high levels are not able to improve much more. The ratings for this indicator would not be included in the index calculations to avoid “penalizing” these schools for a lack of improvement when they cannot achieve the maximum improvement rating. Specifically, the improvement indicator would not be used to compute the accountability index when 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 the Learning Index is at or greater than 3.85.) The first year the Index falls above 3.85, the school/district would get a rating based on their improvement. If the Index stays at or above 3.85, the maximum rating is not possible, so the indicator would not be calculated if desired. For the extended graduation rate outcome, the improvement indicator would not be used to compute the accountability index when the rate reaches or exceeds 94% and remains at that level for two consecutive years (the graduation rate must improve by more than 6 percentage points to earn a rating of 7).

Expected Effects: This policy would affect very few schools and districts. Of the schools and districts with reportable data (at least 10 students assessed), two schools reached 3.85 on the Learning Index in 2006 (one in both reading and writing and the other in writing), no district reached this level in 2006, and no school or district reached this level in any subject in 2007. So none would have qualified for excluding assessment results in 2007. Of the schools with graduation data, 11% had a rate that was at least 94% in two consecutive years and could have chosen to not have this indicator counted. Some schools and districts may decide to include the results in the second year, even if the maximum rating is not possible, if the improvement helps their accountability index results. Providing this option allows schools and districts performing at very high levels to maintain very high index scores.

8. Propose counting the highest grade 10 results through August of grade 10.

Rationale: Schools and districts have the option to allow high school students to take the state assessments in grade 9 and to retake the assessments in the spring and summer of grade 10. In some cases, 10th graders miss the spring exam, and they usually retake the exam(s) in August. The original policy was to count only the high school assessment results that were generated in the spring of

grade 10 (or earlier if a student passed the test as a 9th grader). However, AYP results are now calculated using the results through August of grade 10. This policy would align the accountability system with current practice, give credit for a student's highest score, and ensure that students who did not take the test(s) in the spring would have their August results counted.

Expected Effects: This policy would have little effect on the accountability index. Relatively few students take the exam(s) in August of grade 10. Results during grades 11 and 12 will still be considered when looking at those in the Struggling tier to recognize the effort that some districts undergo to help students who are in danger of not graduating unless they pass the required assessments.