THE WASHINGTON STATE BOARD OF EDUCATION

Governance l Accountability l Achievement l Oversight l Career & College Readiness

Title:	**Achievement Index Update**
As Related To:	☐ Goal One: Effective and accountable P-13 governance. ☐ Goal Four: Strategic oversight of the K-12 system. ☒ Goal Two: Comprehensive statewide K-12 accountability. ☐ Goal Five: Career and college readiness for all students. ☐ Goal Three: Closing achievement gap. ☐ Other
Relevant To Board Roles:	☒ Policy Leadership ☐ Communication ☒ System Oversight ☐ Convening and Facilitating ☐ Advocacy
Policy Considerations / Key Questions:	The State Board of Education has expressed interest in creating an English Language Acquisition Award in a manner that recognizes the increased achievement of English Language Learners (ELLs). The proposed qualification criteria are presented for the SBE consideration.
Possible Board Action:	☐ Review ☐ Adopt ☒ Approve ☐ Other
Materials Included in Packet:	☒ Memo ☐ Graphs / Graphics ☐ Third-Party Materials ☐ PowerPoint
Synopsis:	The Achievement Index Update memo presents results of the descriptive analyses conducted on the Preliminary 2-Year Composite Index. The analyses address issues such as the relationship of the Revised Index to the old Index, the relationship of Revised Index rating to school characteristics (enrollment and Free and Reduced Prioce Lunch Program participants), and some preliminary information about the stability of median SGPs over time. The academic performance of ELL (current) and Former ELL students in the Index is discussed. The memo also provides the proposed qualifying criteria for a possible English Language Acquisition Award.

ACHIEVEMENT INDEX UPDATE

Policy Consideration

The State Board of Education has expressed interest in creating an English Language Acquisition Award in a manner that recognizes the increased achievement of English Language Learners (ELLs). The proposed qualification criteria are presented for the SBE consideration.

Summary

Five major findings for the Preliminary 2-Year Composite Index school ratings are summarized below. The findings include:
- The 2-Year Composite rating is statistically similar to the old Index rating.
- Both school enrollment and the percentage of Free and Reduced Priced Lunch participants at a school are weak predictors of the 2-Year Composite Index rating, indicating a lack of analytical bias.
- Median SGPs are more variable from one year to the next than are proficiency rates but the variability is smoothed through averaging. This confirms the value of the averaging three years of ratings.
- The achievement gap as measured by proficiency is large for ELL students in both reading and math but is much smaller when framed in the context of SGPs.
- The Former ELL subgroup outperforms the All Students group in both reading and math and growth and proficiency.

The vast majority of the Achievement and Accountability Workgroup (AAW) supports the idea of an English Language Acquisition Award based both the WELPA and the regular state assessments. Based on input from the AAW, five criteria are proposed to qualify for the English Language Acquisition Award. Preliminary analyses show that approximately 20 to 30 schools might qualify for the award.

Background

Approval of the Revised Index

As was reported in November, the production of cohort graduation rates and SGPs (student growth percentiles) occurs on a lagged schedule, meaning that these data become available several weeks after the publication of annual assessment results. This schedule has prevented OSPI and SBE from generating completed Index results from three complete school years of data until very recently. Previous data runs have utilized one and two years of data. Staff's initial plan was to submit simulations to the federal government based on these preliminary data. However, the Index is proposed to be a composite of three years of data, and through staff analysis we have noticed some subtle differences in the results as multiple

years of data have been incorporated into the composite score. As a result, a decision was made mutually by SBE and OSPI staff to delay submission of the results to the USDOE until all three years of completed data could be used to generate the Priority, Focus, and Emerging schools lists as part of the ESEA flexibility application. Those lists have now been generated and are going through a validity testing process with staff and external consultants. What follows is some analysis of the results that have been generated as part of this validity testing process. Under the current submission plan, USDOE approval could occur as early as mid-February 2014.

In December 2013, the OSPI and SBE received two separate data files providing information about the 2-Year Composite Ratings for the Revised Achievement Index computations for 2011 and 2012. The analyses described and discussed below are derived from the 2-Year Composite ratings. As mentioned above, similar analyses will be required when the 3-Year Composite ratings are delivered to the OSPI and the SBE.

Relationship of the Revised Index to the Old Index

We use correlation coefficients to numerically describe the relationship between two variables. A correlation is characterized as positive when high scores on one variable associate with high scores on the other variable and low scores on the first variable associate with low scores on the second variable. A negative correlation results when high scores from one variable are associated to low scores on the other variable. Correlation coefficients range from +1.00 to -1.00. Correlation coefficients near zero indicate no consistent relationship among the measured variables.

Correlation coefficients were computed comparing the old Index rating to the 2-Year Composite Revised Index rating for 1814 schools for the each of the six scenarios. The correlation coefficients are moderately strong and positive (0.639 to 0.686). This means that schools rated high on the old Index would be predicted to score generally high on the Revised Index but that some differences would be expected.

Of the six scenarios examined, the 60:40 (growth to proficiency) 2-Year Composite scenario yielded the lowest correlation (0.639), meaning that this is the scenario (of the six) most dissimilar to the old Index. This is because the scenario utilizes the highest percentage of growth and includes the Targeted Subgroups in the Priority and Focus School analysis.

These values show that the Revised Index provides school ratings similar to those of the old Index. The similarity is likely due to the reliance on proficiency measures, and yet the Revised Index ratings differ due to the inclusion of growth measures and a new Targeted Subgroup calculation.

As the State Board of Education tasked with the design of the Revised Index, you should be concerned if the school index score derived from the Revised Index is highly correlated to the old Index score. If the correlations here are too high, the Revised Index might be criticized as being too similar or essentially "the same" as the old Index. If the correlations are too low, the Revised Index might be criticized as being far too different from the old Index. A near-perfect balance appears to have been achieved here. The Revised Index scenarios are sufficiently different from the old Index, values growth, and remain credible because the Revised Index is not too different from the old Index.

Relationship to School Characteristics

<u>School Enrollment</u>
Correlation coefficients were computed comparing the 2-Year Composite Revised Index rating for 1797 schools to the 2012 school enrollment. For all six different scenarios, the correlation coefficients are weak and positive (0.144 and 0.237). This means that there is little relationship between the variables. For the 60:40 (growth to proficiency) scenario, the weak and positive correlation coefficient (0.166) shows that on average the Composite Index rating increases as school enrollment increases. The systematic relationship is poorly developed and school enrollment is not a good predictor of Composite Index rating. If this correlation was too high (greater than 0.500), the Revised Index might be labeled as "biased" in manner favoring large or small schools. The correlations show that the Revised Index is fair and unbiased with respect to school enrollment or school size.

<u>Poverty</u>
The correlation coefficients were computed comparing the 2-Year Composite Revised Index rating for 1802 schools to the percentage of students at the school who participated in the Free and Reduced Price Lunch (FRL) Program in 2012. The correlation coefficients for all six of the scenarios are moderate to moderately strong and negative (-0.448 to -0.604). The old AYP analyses that relied exclusively on proficiency rates resulted in very strong and negative correlations (-0.800 and higher). The inclusion of student academic growth as an indicator reduces the relationship between the Index and poverty, which is most desirable. The 60:40 (growth to proficiency) scenario, yielded the lowest correlation coefficient (-0.448) indicating that on average the Composite Index rating decreases as the percentage of Free and Reduced Price Lunch participants increases. The percentage of Free and Reduced Priced Lunch participants at a school is not a good predictor of the 2-Year Composite Index rating. For this relationship, a low correlation means that the analysis is unbiased with respect to school FRL percentage. The bias here is not excessive, and can be minimized through the use of the higher weighting of growth as compared to proficiency. I would certainly be concerned if this correlation were to be greater than -0.700.

Because the State Board of Education was tasked with school accountability, you should be concerned if the school index score is too closely related to school characteristics, such as school enrollment, percentage of FRL students, percentage of ELL, students, and percentage of students with a disability. A very close relationship or high correlation may imply that the Revised Index is unfair to a school for one reason or another. The analyses conducted and presented here do not indicate any serious analytical bias, meaning that the Index is fair for all schools.

There are some limitations in the data available, so to better assess the relationships of the Index to school characteristics future work should include:
- the number of students assessed at the school should be used in place of the enrollment figures because the school enrollment can differ substantially from the assessed population (for example, the tested population at a K-5 school (50 percent) would differ substantially from the tested population at a 6-8 middle school (100 percent),
- the percentage of FRL program participants assessed at the school should be used in place of the total FRL population because the school FRL percentage can differ substantially from the assessed population,
- include an analysis demonstrating the relationship between the percentage of ELL students participating in the state assessments and the 2-Year Composite Index rating, and

- include an analysis demonstrating the relationship between the percentage of students with a disability participating in the state assessments and the 2-Year Composite Index rating.

Based on the data available, it is safe to say that neither school enrollment nor percentage of FRL participants are good predictors of the 2-Year Composite Index rating.

Growth and the Revised Index

Previous paragraphs demonstrate that the school ratings computed through the 2-Year Composite Index are statistically similar to the school ratings computed through the old Index methodology. Further, that the differences are largely brought about through the use of the Targeted Subgroup and the inclusion of median SGPs in reading and math for all subgroups. The current model of the Revised Index weights growth and proficiency at a 60:40 ratio based on stakeholder input that values growth above proficiency. With such weighting, one might question the stability of median SGPs over time. The year to year comparison of median SGPs is a new topic nationally and is just now beginning to be addressed locally.

Computed medians are sensitive to n-counts and smaller n-counts tend to result in a greater variability of medians. This means the median SGPs of smaller schools would be more likely to increase or decrease dramatically from one year to the next. However, this phenomenon is not limited to growth indicators as the proficiency rate for small groups also can fluctuate substantially from year to year. Nonetheless, computed medians must be interpreted carefully and always in the context of n-counts. The scope of this discussion is limited because of the absence of n-counts in the available dataset and because the dataset is limited to only two years of data (2010-11 and 2011-12). This discussion will be expanded upon when the 3-Year Composite Index and related dataset is available.

Based on the two years of data, the discussion of changes in median SGPs can only be started. For purposes here and to address the stability over time question, the 2011 SGP is subtracted from the 2012 SGP. Three results are possible:
1. A negative number results which means the median SGP in 2012 was less than the median SGP in 2011 (median SGP went down).
2. A positive number results which means the 2012 median SGP was greater than the 2011 median SGP (median SGP went up).
3. The result is zero which means the median is unchanged from one year to the next.

The median SGP in reading declined for 867 schools by up to 35 percentile points and the average decline for a school was 7.6 points. The median SGP in reading increased for 865 schools by up to 39 percentile points and the average increase for a school was 7.8 points (Table 1). The median SGP in reading was unchanged for only 55 schools. Predictably, approximately the same number of schools yields an increase in median SGP as do schools showing a median SGP decline. The median SGPs for reading changed (increased or decreased) by five points or less for approximately 800 schools. This is important because a small (five point) change in median SGP would change the indicator rating by only one point. When averaged with other content and subgroups, the overall impact to the school rating would be minimal unless the change is similar for all subgroups and content. If a median SGP change were to be similar for all subgroups, a marked change would be evident in the overall rating, which is the intent.

The median SGP in math declined for 891 schools by up to 45 percentile points and the average decline for a school was 8.7 points. The median SGP in math increased for 851 schools by up to 57.5 percentile points and the average increase for a school was 7.8 points

(Table 1). The median SGP in reading was unchanged for 36 schools. Math median SGPs changed by five points or less for approximately 700 schools. In the manner described above for reading, the impact from isolated median SGP changes would be reduced through group averaging but systematic changes would be reflected in the rating attributed to growth as intended.

Table 1: Growth and Proficiency Rate Changes for Reading and Math (2011 to 2012).

| | Reading Change | | Math Change | |
	Proficiency (percentage points)	Growth (percentile points)	Proficiency (percentage points)	Growth (percentile points)
High	30.15	39.00	38.99	47.00*
Low	-36.18	-35.00	-33.41	-45.00

*Note: one outlier of 57.50 was removed from the analysis.

The year-to-year variability of measures occurs for small schools for all measures; the variability is not unique to SGPs as Table 1 shows substantial variability for proficiency as well. Table 1 also shows that the variability for math is greater than that for reading and that the variability for growth is slightly greater than that for proficiency.

As would be expected, a moderate correlation between median SGP change and proficiency rate change for reading is indicated ($r = 0.475$, $n = 1777$). The correlation between median math SGP change and math proficiency rate change for reading is 0.482. Generally speaking, this means that schools where a reading or math proficiency rate increase occurred in 2012 the median SGP also increased. However, the relationship is not at all well developed.

Just as expected, median SGPs fluctuate from one year to the next but to a slightly greater degree than do proficiency rates (Table 1). The changes from one year to the next will be most pronounced in schools with small n-counts. Schools with the greatest increases or decreases are characterized by small school enrollments and presumably an even smaller number of SGP records. The Composite Achievement Index mitigates the negative impacts of median SGP fluctuation by averaging the three years of median data.

The relationships between median SGPs, proficiency rates, the changes from year to year, and the school ratings are not at all simple. The interplay between growth and proficiency is complex, strewn throughout the Revised Index, and creates rating scenarios that have the appearance of impossibility but do in fact occur. As examples:
- School proficiency rate can go up but median SGP go down,
- School median SGP can go up but the school proficiency rate can go down,
- A school with relatively high proficiency rates (60 to 70 percent) can be identified in the bottom five percent of schools due to low student growth.

Regardless of the seemingly impossible results, the Revised Index appears to be working exactly as intended and as designed. There is no indication of analytical bias based on school enrollment or other school characteristics. The impact of year-to-year wobble in indicators is reduced through subgroup averaging, content averaging, and 3-Year averaging.

As the State Board of Education tasked with school accountability, you want to be certain that the inclusion of student academic growth in the Revised Index provides a higher degree of confidence in the school rating or identification. The Revised Index creates the circumstance whereby schools will be acknowledged for high growth rates, high proficiency rates, or a combination of the two. These identifications are being scrutinized for face validity by the OSPI and SBE. The OSPI identified several schools that may not pass the face-validity test

based on the 2-Year Composite Index. The face validity issue may be resolved when the 3-Year Composite Index ratings are available.

Performance of ELL and Former ELL Students

For purposes here, ELL students are those who were receiving services at the time of testing, and Former ELL students are those who had been enrolled for ELL at some prior time but were exited from ELL services prior to testing.

In the context of proficiency, the academic achievement of ELL students is amongst the lowest of all reported ESEA subgroups. The proficiency rates for ELL students at a school are approximately 20 to 23 percent for reading and for math in 2011 and 2012. Schools with high performing ELL subgroups yield proficiency rates of approximately 45 to 47 percent for reading and math.

With respect to student growth as indicated by school median SGP, a different picture emerges. The school median SGP for the ELL subgroup is 44 to 45 for reading and 46 to 48 for math in 2011 and 2012. These median SGP values are slightly below the median value of 50 for all schools. The achievement gap as measured by proficiency is large for ELL students in both reading and math but is much smaller when framed in the context of SGPs. By weighting growth more heavily than proficiency in the Revised Index, the ELL students at a school are more likely to make a positive contribution to the school rating.

Former ELL students achieve at significantly higher levels. The median proficiency rate for the Former ELL subgroups at schools is approximately 73 to 77 percent for reading and 62 to 67 percent for math. These performance levels exceed the state average for the All Students group. Schools with high performing Former ELL subgroups yield proficiency rates of approximately 92 to 93 percent for reading and math. For growth, the school median SGP for the Former ELL subgroup is approximately 53 for reading and 54 to 55 for math in 2011 and 2012. These median SGP values are slightly higher than the median value of 50 for all schools across the state. For both proficiency and growth, the Former ELL subgroup outperforms the All Students group.

Both the ELL subgroup and the Former ELL subgroup proficiency rates and median SGP factor into the Revised Index rating through the Targeted Subgroup measures for elementary and middle schools. In addition, the Revised Index rating for high schools is impacted by the graduation rates for each of the subgroups.

For all schools with reportable proficiency rates for ELL and Former ELL subgroups, the reading and math proficiency rates for Former ELL students was two to three times higher than that for the ELL students, and in nearly every case the rating for the Former ELL subgroup exceeded the average of the Targeted Subgroup. Correlation coefficients for ELL, Former ELL, and the 2-Year Composite Index rating were computed for all four proficiency content areas. In all cases, the correlation coefficients for the Former ELL group were substantially higher than the correlation coefficients for the ELL subgroup. This means that the school rating is more closely related to the achievement of the Former ELL students that the achievement of the current ELL students.

Collectively, these two facts provide evidence that the Former ELL subgroup has a greater influence on the Composite Index rating than does the ELL subgroup. From this, one could arguably conclude that the ELL performance with respect to proficiency rates has little overall

impact (positive or negative) on the 2-Year Composite Index rating when the Former ELL subgroup is also reportable, because the lower performance of the ELL subgroup is mitigated by the much stronger performance of the Former ELL subgroup.

For growth measures, the median SGP calculations for reading and math for ELL students are demonstrably lower than those for the Former ELL students but only slightly below the state average. Correlation coefficients for ELL, Former ELL, and the 2-Year Composite Index rating were computed for reading and math, and in both cases, the correlation coefficient for the Former ELL group was nearly identical to the correlation coefficient for the ELL subgroup. This would indicate that ELL and Former ELL subgroups have somewhat equal impact on the 2-Year Composite Index rating and that the two subgroups perform in the range of typical growth.

An analysis was conducted to determine whether the ELL subgroup performance in reading and math proficiency and growth was statistically different in Priority versus Non-Priority Schools. The t-tests showed that for all ELL measures (reading and math, proficiency and growth), the performance of the ELL students at Priority Schools was substantially lower than and statistically different from the ELL performance at Non-Priority Schools. However, this was true for the Former ELL, SWD, and FRL subgroups as well, indicating that this is not a phenomenon unique to the ELL subgroup. One would conclude that the weak ELL performance on state assessments contributed to the Priority School identification in combination with other low performing subgroups. It would be inappropriate to attribute Priority School identification to the weak academic performance to any single subgroup.

A very weak performance by any subgroup(s) would have the negative impact of contributing to a lower 2-Year Composite Index rating and make the school more susceptible to Priority or Focus School identification. Under the current methodology, some schools with a low performing ELL subgroup would be expected to be identified as Priority and or Focus Schools while others would not. There is no evidence to suggest that low academic measures for the ELL subgroup will result in the identification as a Priority School as causation cannot be established.

Given that the Former ELL cell is a new cell utilized in the Revised Index, the SBE has an interest in assessing validity and determining that it does not create unintended consequences for students or schools. After analyzing the datasets, it is evident that the Former ELL subgroup bolsters the school rating. The Board would also want to be sure that the lower academic performance of current ELL students is not masked or concealed by the presence of the Former ELL subgroup. The Focus School identification will be based upon a rank ordering of the lowest performing subgroups (including ELL), which means that the ELL academic performance will not be masked in any manner.

English Language Acquisition Award

Language acquisition is an indicator of school success separate but not entirely distinct from the typical indicators of school success such as reading proficiency rates and median school SGPs in reading. The Achievement and Accountability Workgroup (AAW) met on December 9 to discuss the appropriateness and possible indicators for an English Language Acquisition Award. The vast majority of the AAW membership agreed that an award recognizing language acquisition was appropriate and that the award be based upon both the WELPA and the regular state assessments. With respect to the state assessment indicators, the majority of the

AAW membership agreed that reading and math growth (SGPs) should be factors in the award criteria.

In order to be eligible for the English Language Acquisition Award, the following qualifying criteria are proposed:
1. School must meet or exceed the Title III AMAO 1 and AMAO 2 targets.
2. The ELL subgroup must earn an Index rating of six or higher (median SGP ≥ 50) for both reading and math on the current year Index and that indicator must be based upon the SGP measures for at least 50 ELL students.
3. Priority and Focus Schools currently identified through low ELL performance are excluded from award consideration.

A preliminary review of the Title III AMAO calculations and based on the 2012 median SGPs, approximately 40 schools meet criteria 1, 2, and 3. This number of schools would be reduced when the ELL n-counts are considered. After factoring in criteria 4 and 5 and adjusting for n-count thresholds, 20 to 30 schools might be deemed to have met all 5 criteria for the English Language Acquisition Award. This number approximates the top 5 percent of schools with reportable ELL populations in the 2012 Index.

Action

The Board will consider approval of qualifying criteria for the English Language Acquisition Award.