

Reviewing Changes in School and District Performance Across Years

A Brief on School and District Performance Framework Outcomes in the
2009-2010, 2010-2011, and 2011-2012 School Years

Elena Diaz-Bilello, Center for Assessment

Introduction

This brief is the last in a three-part series examining three issues related to outcomes from the school and district performance frameworks. These issues include:

- The relationship between a school and district's size to their median growth percentiles (MGPs), growth ratings, and overall ratings achieved on the framework each year;
- The relationship between poverty, English language learner status, minority status, and a school or district's ratings by performance indicator¹ and overall rating; and,
- The stability of performance achieved by schools and districts over time (2010-2012) on status, growth, and overall ratings on the framework.

This brief presents results from undertaking exploratory analysis to examine the degree to which school and district performance ratings change across years on the performance frameworks. Two pieces of information are reviewed: the year-to-year correlations for each performance indicator in the framework and overall points earned, and the year-to-year percentage changes in framework points earned by schools and districts by each key indicator and overall points earned relative to school and district size. The year-to-year correlations may serve as reliability estimates that can inform stakeholders whether the performance signals in an accountability system are fairly consistent across time (Kane & Staiger, 2002). That is, if the signals tend to move in different directions for each year reviewed, this finding would raise questions about the quality of the information being provided by the frameworks to drive and inform decision-making across schools and districts in the state. In the analysis presented, the framework points for the 2010 and 2011 years have been adjusted based on the method used to assign points for each key performance indicator and the overall rating in 2012. This adjustment was made to ensure comparability of points earned by schools and districts across years reviewed. Additionally, the n size reported for schools and districts throughout this paper are based on the population of math test takers at a given school or district. The proxy n size reported in this brief for a school or district was based on Math since this subject typically reports larger numbers of test takers relative to the language arts subjects. In this brief, findings from the school performance frameworks (SPFs) will first be reviewed and followed by findings from the district performance frameworks (DPFs).

School Performance Framework Results

To understand the degree to which performance ratings for schools and districts fluctuate over time on the frameworks, the correlations for the year-to-year percentage of points earned on each key indicator and the overall framework were reviewed in three areas: the 1-year framework, the 3-year framework, and the set of framework points from either the 1-year or the 3-year frameworks used to make accountability decisions in each year. The correlations based on the underlying metrics for each key indicator are not reviewed here since the points earned at the indicator level are designed to account for a range in performance and are used to determine ratings for schools and districts. All descriptive information presented in this paper reflects results from traditional schools since alternative education campus schools are evaluated using a separate framework. Table 1 presents findings from reviewing the year-to-year correlations across all schools in the state for the 1-year SPFs by indicator, overall rating and year. Overall, the year-to-year correlations for each pair of years reviewed in the table are similar. The correlations in Table 1 indicate that the reliability of the status indicators (i.e., academic achievement and

¹ The four performance indicators are: Academic Achievement, Academic Growth, Academic Growth Gaps and Post-Secondary and Workforce Readiness (PWR).

post-secondary and workforce readiness or PWR) is very strong, and that the reliability of the signals for the growth indicators in the framework is moderate. Overall, Table 1 suggests that most schools perform fairly consistently, but results are not as strong as the overall rating results in Tables 2 and 3.

Table 1: Correlations for 1-Year Framework

Indicator	2011 to 2012	2010 to 2011
Growth Overall	.60	.62
Academic Growth Gaps	.55	.57
PWR	.85	.86
Academic Achievement	.93	.93
Overall Rating	.73	.74

Table 2: Correlations for 3-Year Framework

Indicator	2011-2012	2010-2011
Growth Overall	.87	.87
Academic Growth Gaps	.88	.88
PWR	.93	.94
Academic Achievement	.97	.97
Overall Rating	.93	.93

Table 3: Year-to-year Correlations for Framework Used to Rate Schools

Indicator	2011-2012	2010-2011
Growth Overall	.77	.77
Academic Growth Gaps	.70	.69
PWR	.90	.92
Academic Achievement	.96	.95
Overall Rating	.85	.85

In contrast to the results presented in Table 1, the correlations presented in Table 2 for the 3-year framework are strong for the growth indicators and high for all other areas in the framework. However, this finding is expected especially since two out of three years of data overlap across frameworks being compared. However, the high correlations do suggest that pooling data across years helps ensure that year-to-year performance changes in growth are less pronounced than the degree of changes found for growth reported on the 1-year framework.

Table 3 presents the year-to-year correlations based on outcomes from a mix of results from the 1-year or the 3-year frameworks used for accountability². Overall, the correlations are reasonably high for growth and strong for all other areas in the framework. Considering that the associations found for the growth indicators for Table 3 are considerably stronger than those presented in Table 1 and slightly weaker than those presented in Table 2, this would suggest that many schools are being evaluated by the 3-year frameworks, but that there are several schools where only one year of data are being used for accountability. It is worth noting here that CDE’s policy of using the higher rating from the 1- or 3-year framework (i.e., if both frameworks have the same number of data points available) means, for some schools, that the 1-year framework is more likely to be used for accountability reporting than the 3-year frameworks, despite the fact that the 3-year framework may provide more consistent signals for a school. This policy addresses a desired outcome from district stakeholders who would prefer to highlight strong performance demonstrated in one year over more “consistent” results from the 3-year framework. Although this approach provides one avenue for the state and districts to recognize higher performance achieved by a school in a single year, this also means that state and districts favor consistent results over lower performance achieved by a school in a single year. In both cases, schools are given the benefit of the doubt since higher performance is always favored. Yet, despite this policy, the correlations from

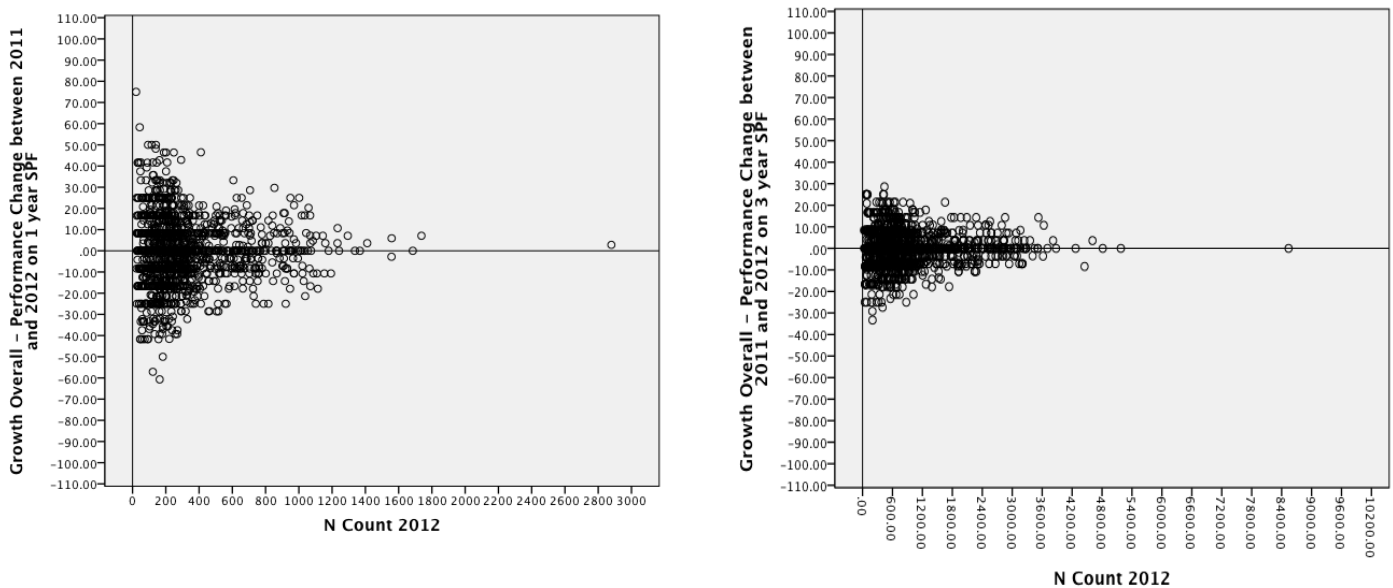
² Schools are evaluated by the state using either the 1 or 3 year SPFs depending upon: 1) which framework yields the most data about the school; or, 2) if both frameworks yield the same amount of data (i.e., have the same number of indicators), then the higher rating is selected.

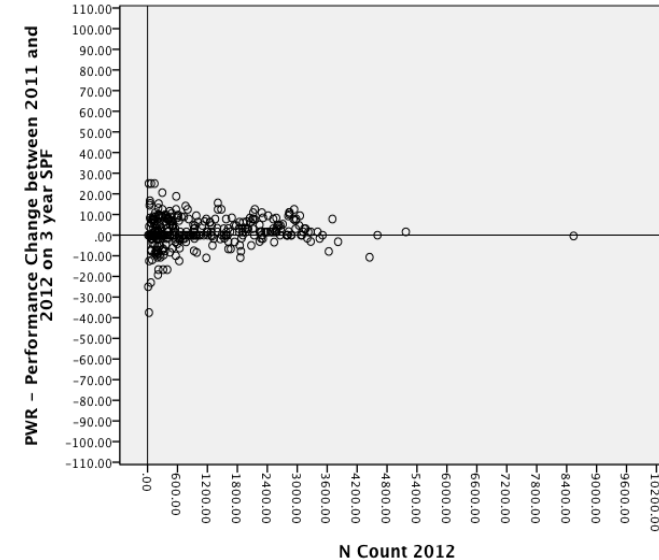
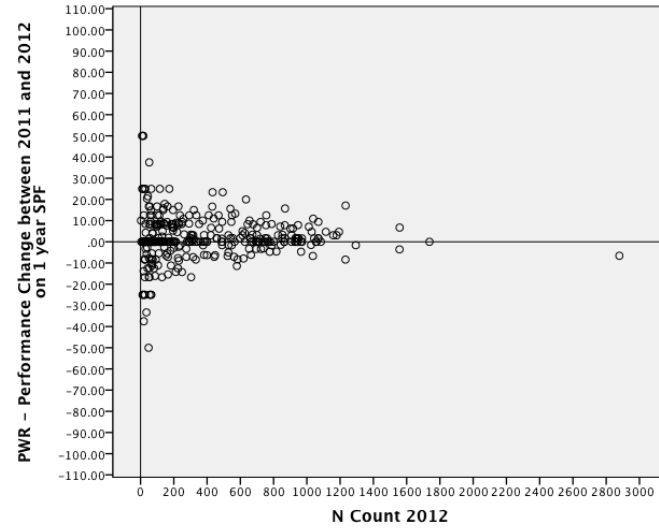
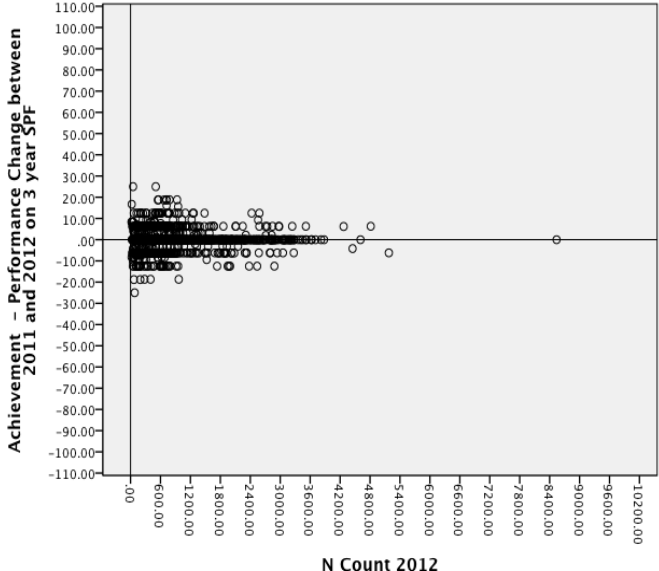
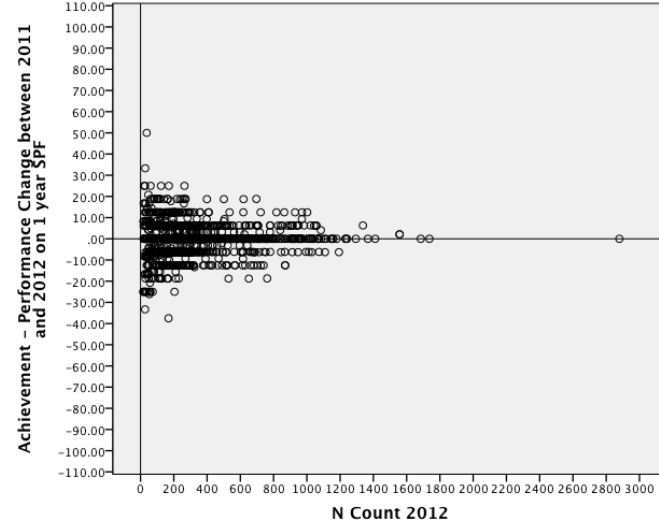
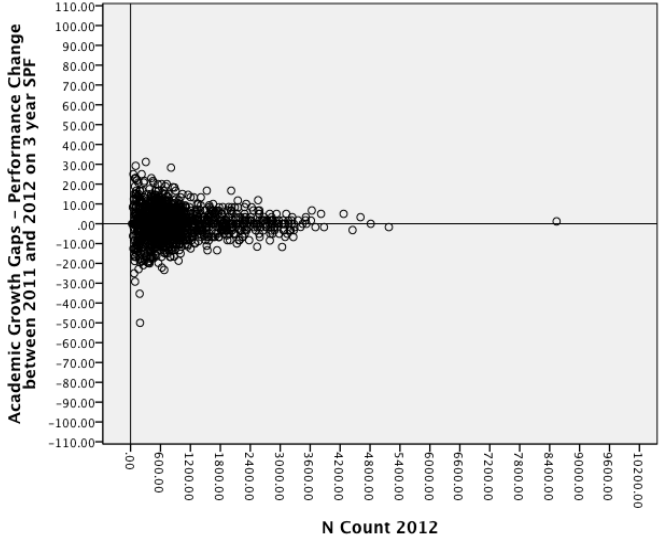
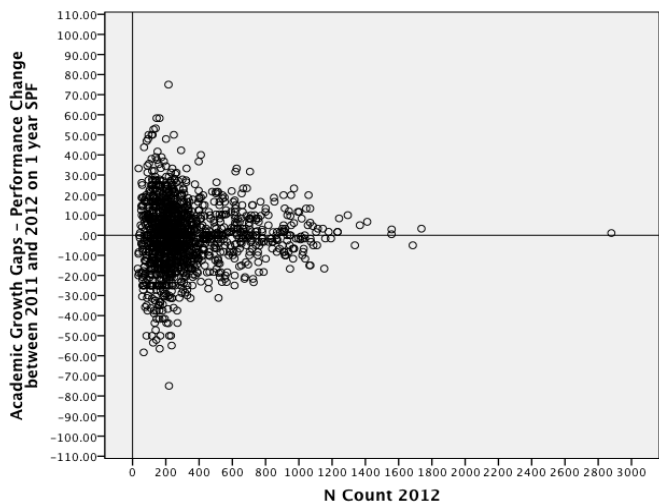
Table 3 still indicate that these shifts are not remarkably differently from one year to the next since the correlations are fairly strong when considering the overall performance across schools.

As indicated in the first paper of this series, the 3-year pooled data will present higher levels of reliability in the accountability system than the information yielded by one year of results. The findings in Tables 1-3 appear to support this concept that pooling data across years provides more reliable performance signals than outcomes reported using results from the one year framework. Although the state's approach of using pooled data helps ensure that performance signals are less susceptible to large year-to-year performance bounces, the state has also signaled to the field that they value efforts by schools to achieve higher performance in the most recent year by favoring results of the 1-year framework over the 3-year framework when the same number of data points are available on both frameworks.

The first paper in this series also highlighted the issue that many of the smaller schools and districts will likely exhibit more year-to-year variation in performance than larger entities. To evaluate the extent to which instability emerges more prominently with smaller n sizes, the year-to-year changes in performance were reviewed relative to school size. The following set of graphs present the year-to-year change in the percentage of total points earned across all schools from 2011-2012 on each of the four key indicators. In each graph, the y-axis represents the change in the percentage of points earned on each indicator and overall points and the x-axis represents the n size of schools. The horizontal line at 0 on the y-axis indicates that no changes have taken place from year-to-year. All of the graphs on the left hand side reflect the year-to-year change in performance by indicator between the 2011 and the 2012 school years for the 1-year SPF. The graphs on the right hand side reflect the year-to-year change in performance by indicator for the 3-year SPF. The n-sizes reflected on the x-axis reflect the total number of students with a math score in the 2011-2012 school year on either the one or the three-year framework. The same set of graphs generated for the 2010-2011 school year is located in Attachment A.

Figure 1. Performance Change on Key Indicators in 2012





To put the performance changes exhibited in the graphs into context, Table 4 presents the average change, standard deviations, and the minimum and maximum associated with performance changes made by all schools between the 2011 and 2012 school years by each key indicator and total scores.

Table 4: Summary of change by indicator and overall performance on the SPFs

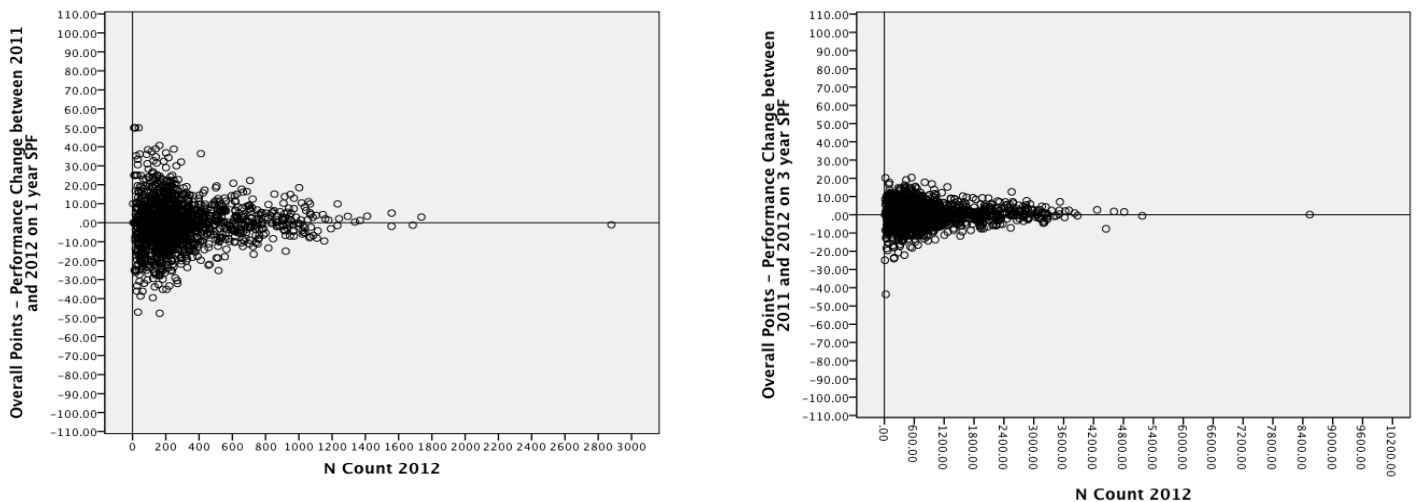
Change in Performance between 2011 and 2012 by Indicator and Overall Points Earned										
	Growth 1 Year	Growth 3 Years	Achievement 1 Year	Achievement 3 Years	PWR 1 Year	PWR 3 Years	Growth Gaps 1 Year	Growth Gaps 3 Years	Overall Pts 1 Year	Overall Pts 3 Year
N	1505	1577	1549	1590	345	347	1390	1538	1574	1594
Mean	-0.6	0.0	-0.1	0.1	1.8	1.8	-0.6	0.1	-0.5	0.1
Median	0.0	0.0	0.0	0.0	0.0	1.5	0.0	0.0	0.0	0.2
Std. Deviation	15.5	8.0	7.5	4.7	10.8	7.0	16.3	7.5	11.9	5.8
Minimum	-60.7	-33.3	-37.5	-25	-50	-37.5	-75	-50	-47.7	-43.6
Maximum	75	28.6	50	25	50	25	75	31.2	50	20.4

The information in Table 4 indicates that with the exception of PWR, the average gains or losses were approximately less than 1 point on the framework. Additionally, the standard deviations in the table reveals that the majority of schools experienced gains or losses of approximately 16 or fewer percentage points on a given indicator on the 1-year framework and approximately 8 or fewer percentage points on a given indicator on the 3-year framework.

In each of the graphs, the smaller schools show larger performance shifts on each indicator reviewed on both sets of frameworks relative to larger schools. For example, for schools with fewer than 100 students on the 1-year framework, approximately 30 percent of 193 schools with growth scores experienced performance changes exceeding one standard deviation (a 15.5 point gain or loss on the framework). In comparison, approximately 19 percent of 203 schools with more than 600 students in the population experienced performance changes exceeding one standard deviation.

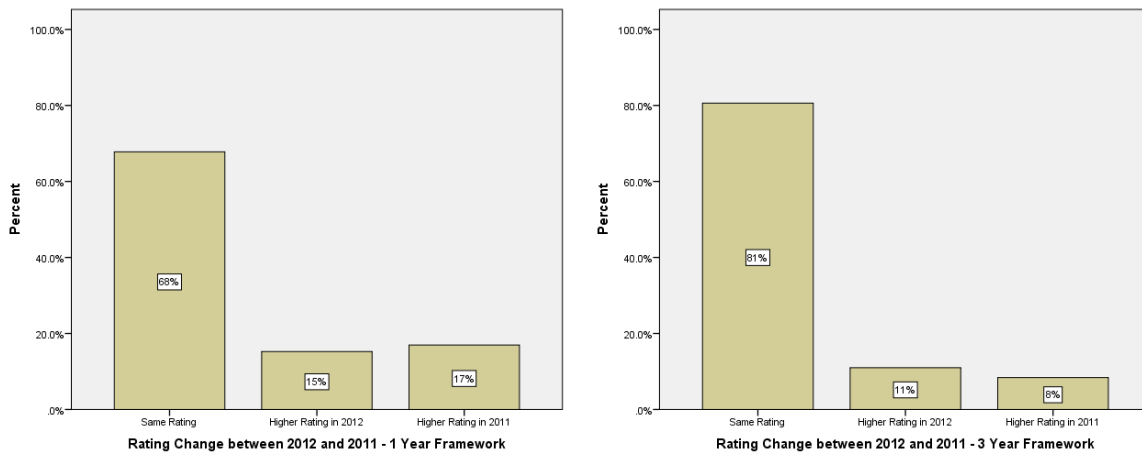
However, the smaller schools show considerably lower levels of year-to-year shifts in performance when the data are pooled across years. That is, when comparing the range of gains or losses made by schools across frameworks, the range for each indicator is more narrow and compressed on the 3-year framework relative to the 1-year framework. Additionally, many of the extreme outliers found on the 1-year framework (e.g., above the plus or minus 20 mark on the y-axis) are eliminated as a result of pooling. The set of graphs presented in Figure 2 shows the overall performance change on the SPFs relative to n size.

Figure 2. Overall SPF Framework Points Change in 2012



Similar to the patterns found for the key indicators in Figure 1, the results from the 3-year framework in Figure 2 are more compressed and clustered around the horizontal line and less extreme changes in performance on the overall framework are found for smaller schools. Table 4 also highlights key differences in the changes captured on the 1-year framework relative to the 3-year framework. As indicated by the table, the standard deviations and changes on the 1-year framework are larger than those reported on the 3-year framework. The results presented in this section appear to support the use of pooling data across years as a viable strategy for reducing the amount of year-to-year volatility found on performance for each indicator and the overall performance across schools. This pooling strategy is particularly useful for stabilizing the year-to-year volatility in performance typically found at smaller schools. To understand the implications of using pooled or un-pooled data to classify schools for accountability purposes, the set of graphs in Figure 3 reflect the change in performance ratings assigned to schools between the 2011 and 2012 school years. The bar chart on the left-hand side presents the change in performance ratings for the 1-year SPF, the bar chart on the right presents the change in performance ratings for the 3-year SPF.

Figure 3. Performance Ratings Change between 2011 and 2012



The correlations reviewed earlier and the differences in year-to-year performance highlighted between the 1- and 3-year frameworks support the outcomes presented in the above bar graphs where the 3-year framework yields the most consistent year-to-year information. In sharp contrast to the 1-year SPF results where 1,067 or 68 percent of schools earn the same ratings for both years, approximately 1,285 or 81 percent of schools retain the same ratings for both years on the 3-year SPF.

District Performance Framework Results

The same set of approaches used to review the SPF results was also used to review the year-to-year changes in performance for the DPFs. In this section, the year-to-year correlations for the DPFs are presented, and are followed by a review of the performance gains made on each key indicator in the framework relative to the n size of districts. Tables 5-7 present the year-to-year correlations found for the DPFs by indicator and for each set of years reviewed.

Table 5: Correlations for 1-Year Framework

Indicator	2011 to 2012	2010 to 2011
Growth Overall	.52	.57
Academic Growth Gaps	.57	.39
PWR	.88	.86
Academic Achievement	.90	.91
Overall Rating	.84	.8

Table 6: Correlations for 3-Year Framework

Indicator	2011-2012	2010-2011
Growth Overall	.88	.87
Academic Growth Gaps	.81	.81
PWR	.95	.93
Academic Achievement	.98	.98
Overall Rating	.96	.96

Table 7: Year-to-year Correlations for Framework Used to Rate Districts

Indicator	2011-2012	2010-2011
Growth Overall	.72	.77
Academic Growth Gaps	.59	.60
PWR	.91	.89
Academic Achievement	.97	.96
Overall Rating	.92	.91

The correlations reported in Tables 5-7 present slightly stronger but similar correlations to those found for the SPFs. In Table 5, the correlations for the growth indicators on the 1-year framework are moderate for each pair of years reviewed and this would suggest that district performance on these two indicators tend to exhibit larger shifts between the two years. The lower growth gap correlation reflected between the 2010 and 2011 school years is due to large shifts (e.g., moving from “Exceeds” to “Does not meet”) in performance exhibited by a few small school districts with fewer than 150 students. However, the correlations for the status based indicators (i.e., PWR and achievement) for each pair of years reviewed are high. These status correlations suggest that many districts tend to exhibit similar year-to-year performance on these two indicators. The correlations for the overall rating on the framework are also high, and this would also suggest that many districts earn similar percentages of framework points on a year-to-year basis.

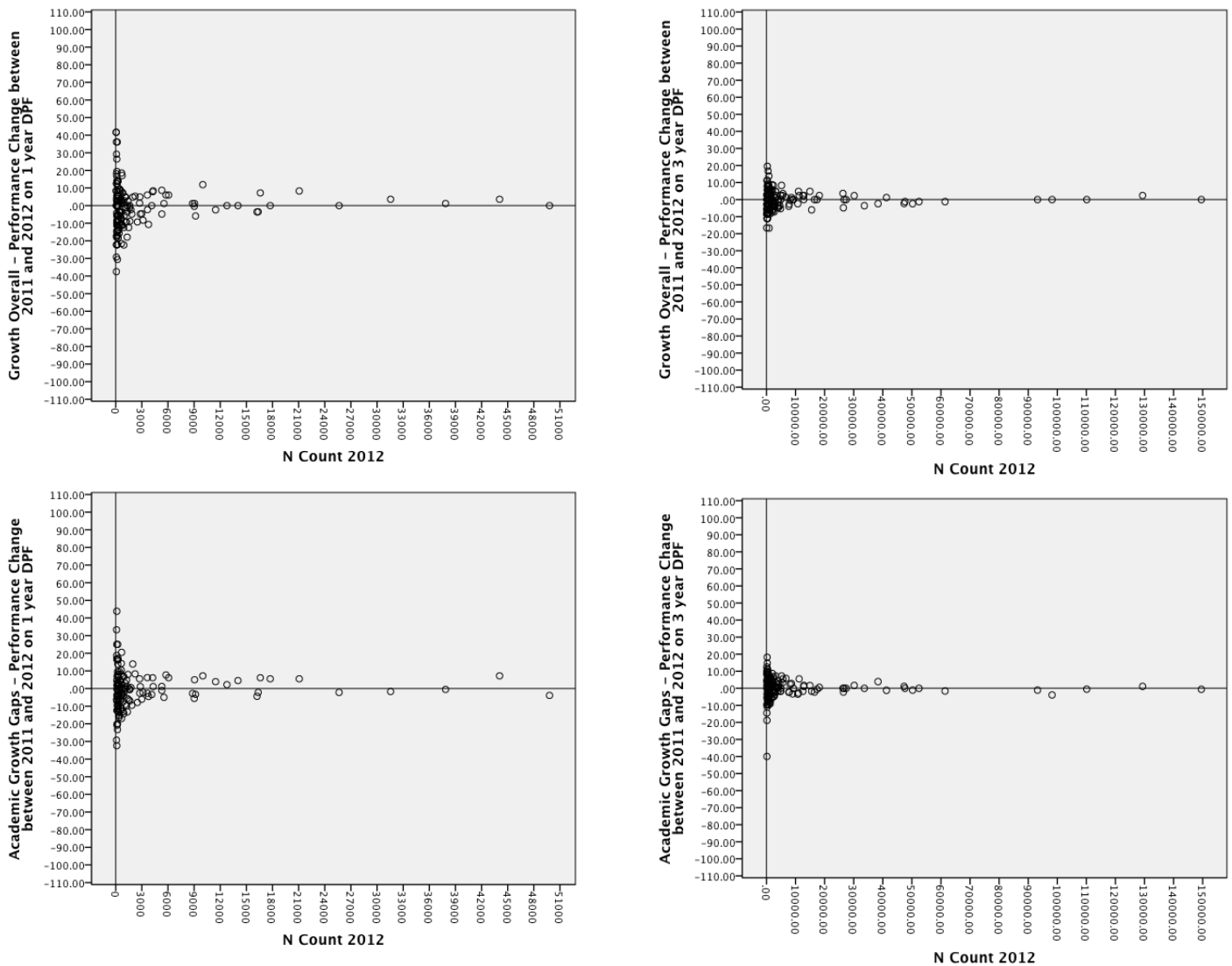
In Table 6, the correlations for the 3-year framework appear considerably stronger than the correlations found in the one year framework for all indicators and the overall rating. The strong correlations for the two growth indicators suggest that there are smaller performance shifts found in the year-to-year growth results reported for districts when using pooled data. The high correlation of .96 achieved for the overall rating when utilizing the 3-year frameworks indicates that the overall percentage of points achieved on the DPFs tend to stay fairly similar from year-to-year across the majority of districts.

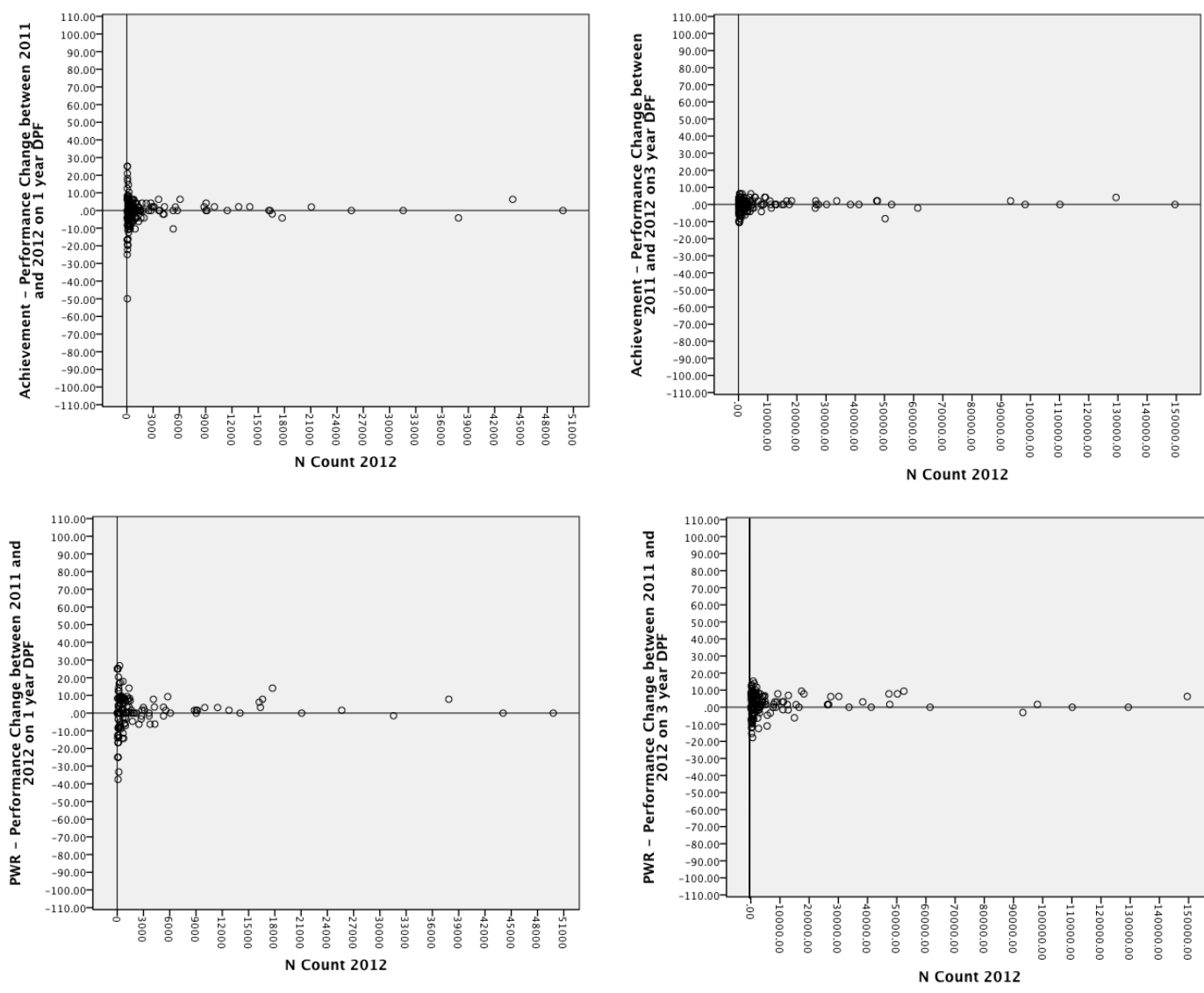
The correlations presented in the Table 7 are based on the mix of 1- and 3-year framework results reported for accountability. Although the year-to-year correlation for the overall rating is still strong and indicates similar levels of performance achieved by districts each year, the two growth indicators exhibit lower but moderate correlations. Further, the correlations found for the two growth indicators are unexpectedly lower than the correlations found for the SPFs, and this discrepancy is due largely to the fact that the growth data from small school districts (e.g., Creede and Pawnee) experienced unusually large shifts in performance on growth and growth gaps. These districts did not meet the minimum n for reporting at the school level, but met the minimum n for district-level reporting therefore growth scores associated with these two districts only appear in the DPFs. As noted earlier for the SPFs, these moderate correlations are also attributed to the state’s policy of favoring higher 1 year results over 3 year results if a district has the same number of data points available in both frameworks. Again, although consistency is

valued for ensuring that year to year performance signals do not “bounce” each year, the state has also highlighted the importance of valuing higher performance over consistency in order to acknowledge performance gains achieved by a district in the most recent year.

The next set of graphs in Figure 4 present the year-to-year changes in the percentage of points earned by each indicator relative to the n size of the district. The x-axis reflects the n size of the district based on the total number of students with math scores during the 2012 school year. The y-axis represents the year-to-year performance change in the percentage of points earned by indicator between the 2011 and 2012 school years. Graphs located on the left-hand side reflect performance from the 1-year DPF and graphs on the right-hand side are based on the 3-year DPF. Attachment B presents a similar set of graphs for the year-to-year performance change between the 2010 and 2011 school years.

Figure 4. Performance Change on Key Indicators in 2012





Similar patterns are found in the graphs under Figure 2 as those found in Figure 1 where higher year-to-year shifts in performance on each key indicator are reduced considerably when using pooled data. To provide more context for the changes reflected in each graph, Table 8 presents the mean, maximum, minimum, and standard deviations associated with changes on each key indicator and the overall percentage of points earned.

Table 8: Summary of change by indicator and overall performance on the DPFs

	Change in Performance between 2011 and 2012 by Indicator and Overall Points Earned									
	Growth 1 Year	Growth 3 Years	Achievement 1 Year	Achievement 3 Years	PWR 1 Year	PWR 3 Years	Growth Gaps 1 Year	Growth Gaps 3 Years	Overall Pts 1 Year	Overall Pts 3 Year
N	167	180	177	180	181	181	142	178	181	181
Mean	-0.8	-0.2	-0.8	-0.6	1.1	1.4	-0.7	-0.2	-0.2	0.4
Median	0.0	0.0	0.0	0.0	0.0	1.5	-0.9	0.0	0.2	0.8
Std. Deviation	12.2	5.4	8.1	3.3	9.7	5.6	10.7	6.0	7.8	3.6
Minimum	-37.5	-16.7	-50.0	-10.4	-37.5	-17.8	-32.4	-40.0	-29.7	-8.7
Maximum	41.7	19.5	25.0	6.3	26.8	15.3	43.8	18.2	31.8	19.8

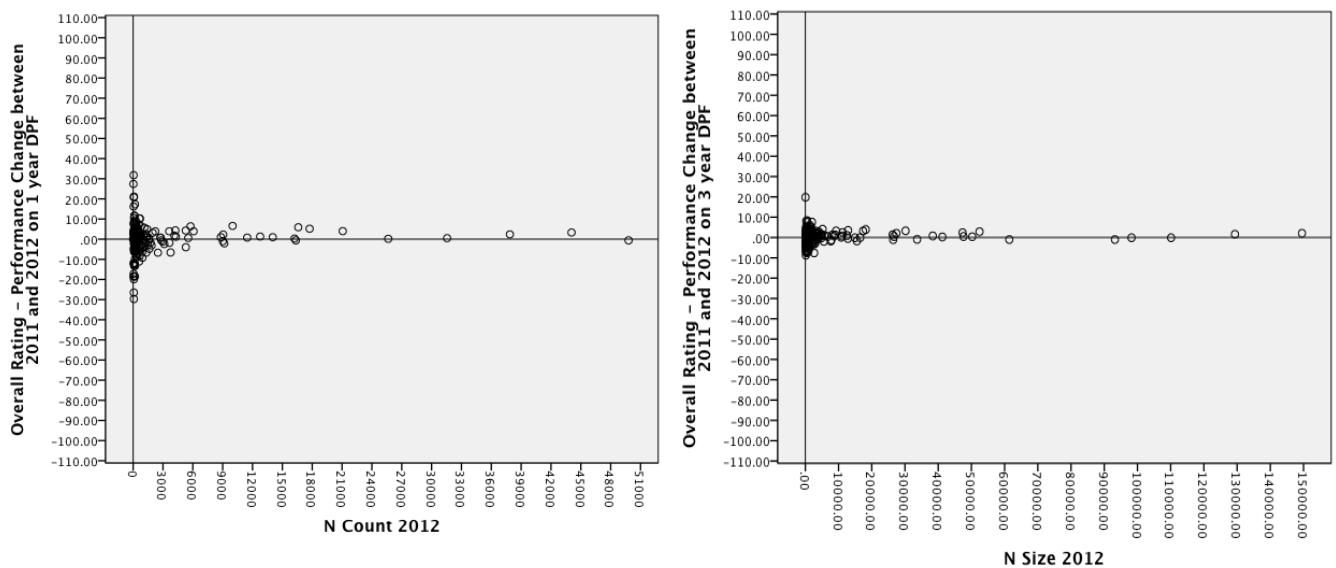
In Table 8, the average change scores exhibited on the 3-year framework are generally smaller than the changes reported for the 1-year framework and fall, on average, within two points on any given indicator.

Additionally, the standard deviations are considerably smaller on the 3-year framework than the 1-year framework. The standard deviations in the table indicate that the majority of districts experience a performance change of plus or minus 12 or fewer percentage points on any given indicator on the 1-year framework and plus or minus 6 or fewer percentage points on any given indicator on the 3-year framework.

As indicated by each of the 1-year framework graphs on the left hand side of Figure 4, the smaller districts (e.g., below 2000 students in the entire district) show greater performance shifts for each indicator relative to districts with larger n sizes (e.g., with more than 4,000 or more students). For example, on growth overall, 31 percent of 131 districts with fewer than 2,000 students experienced performance changes that exceeded one standard deviation or 12 percentage points. In contrast, out of the 26 school districts with 4,000 or more students, none of those districts experienced changes exceeding one standard deviation. On the 3-year frameworks, pooling has the effect of eliminating most outliers which consist only of smaller districts.

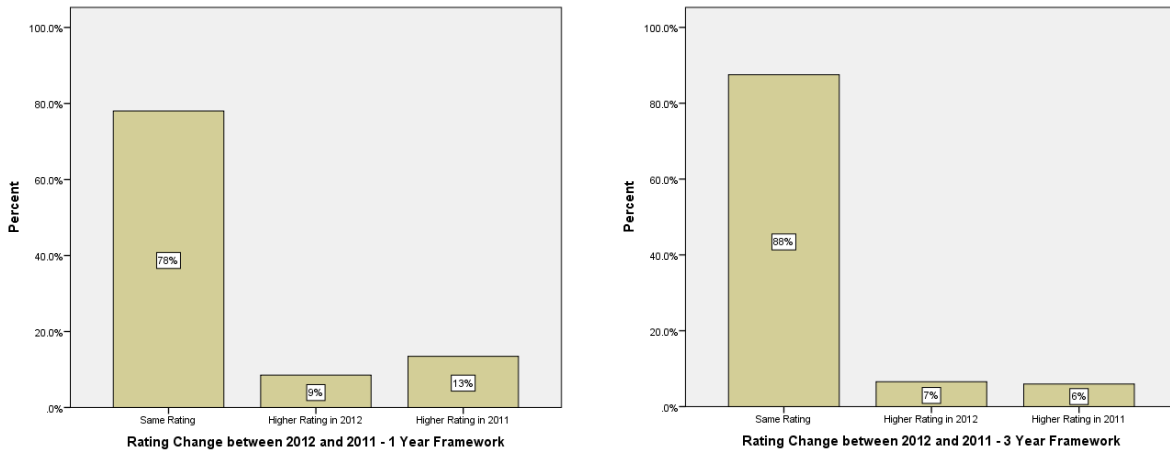
The same pattern holds in Figure 5 when reviewing the performance change on overall points earned on the DPFs.

Figure 5. Overall DPF Framework Points Change in 2012



As indicated by the graphs in Figure 5 and supported by the high correlations in Table 6 for each pair of years reviewed, the 3-year framework shows smaller overall performance shifts for most districts across the state. The impact of using either a 1-year or 3-year framework on the year-to-year ratings assigned to districts can be seen in the bar graphs in Figure 6. In Figure 6, the two bar graphs present the percentage of districts that either changed or retained accreditation ratings between the 2011 and 2012 school years.

Figure 6. Performance Ratings Change between 2011 and 2012



For the 3-year framework, 147 or 88% of districts maintain the same rating assigned in the previous year, compared to 110 or 78% of districts for the 1-year framework. Similar to the results reported for the SPFs, the overall performance ratings signal emitted by the 3-year DPF is more consistent than the signal conveyed by the 1-year DPF.

Conclusion

Since the year-to-year correlations for overall points earned by schools and districts in the frameworks and reported for accountability are strong (see Tables 3 and 6), these correlations suggest that the overall performance signal for most schools and districts appears similar across years. While the consistency of the performance signal is not sufficient to demonstrate validity, this finding would suggest that policy decisions and inferences about school performance informed by the frameworks are not based on volatile or short-term performance fluctuations.

As suggested by the correlations and graphs presented throughout this document, the reliability of the performance signals on each indicator and overall points provided by the frameworks are improved greatly when using pooled data from the 3-year frameworks. However, for larger districts and schools, the state has also provided the opportunity to report and use data from the 1-year framework in the event that higher performance is achieved by these entities in the most recent year. As indicated earlier, this policy serves to privilege higher performance over consistency when schools and districts show gains or improvements that surface on a year to year basis. This finding of consistent signals from pooled data is more relevant to the smaller schools and districts that may exhibit performance swings on each key indicator and the overall points earned from year-to-year compared to larger schools, and districts that exhibit considerably smaller shifts in performance using either pooled or un-pooled data. For this reason, the following considerations were forwarded in the first brief as strategies for either improving the signals from the frameworks or improving the assessment of a school and district's performance: increasing the minimum n size required by the growth indicators that typically convey less reliable performance information than the other indicators; evaluating all entities using pooled data; considering other academic and non-academic indicators of performance outside of those reflected within the framework; and, encouraging specific entities to undergo the request for reconsideration process to ensure that more academic evidence is used to better assess performance taking place at these schools and districts.

In 2012, for the 74 districts in the state with fewer than 300 students, 12 or 16 percent of those districts experienced a shift in performance ratings in 2012 compared to the 2011 school year. In contrast for the same year, for the 66 larger school districts with over 600 students, only 5 or 8 percent of districts switched performance ratings from the prior year. Similarly for schools, most of the reported year-to-year performance rating shifts occur with the smaller schools. In 2012, over half of the schools in the state (915 out of a total of 1,594 rated schools) had fewer than 250 tested students. For the same set of 915 schools, 253 or approximately 28 percent of those schools shifted performance rating categories by one performance level. In contrast to the smaller schools, a total of 36 out of 271 larger schools or approximately 13 percent of schools with an n size of over 500 students switched performance rating categories between 2011 and 2012.

As indicated by the first brief provided earlier in the year to the S.B. 163 taskforce, the performance shifts found in smaller schools and districts may be attributed to several factors. For example, a small change in demographic composition or student enrollments could result in a larger impact on a smaller school or district's performance compared to larger entities where considerably more students need to move in performance before seeing a discernible impact on overall results. In some cases, these performance rating shifts are entirely reasonable as a result of a school or district being located close to a rating cut-point or due to known contextual factors. However, in some cases where the year-to-year shifts have occurred for two consecutive years and these shifts move in different directions each year (i.e., no consistent trend), these schools and districts should undertake a careful review of their performance data and identify whether these shifts seem consistent in comparison to other data points collected. In 2012, out of the 164 schools showing inconsistent performance trends across 3 years, approximately 122 or 74 percent of those schools had fewer than 250 students. In the same year, 7 out of the 10 districts showing inconsistent and large shifts in year-to-year performance represented smaller districts with fewer than 300 students. The Colorado Department of Education may want to encourage that smaller districts or schools³ with inconsistent performance patterns or with an incomplete set of indicators review their data carefully and determine whether they may want to submit more evidence to re-evaluate their performance during the request for reconsideration process.

Despite this brief's focus on evaluating the stability of performance results on key indicators and overall ratings reported on the SPFs/DPFs, district and school stakeholder reviews of the underlying metrics that support the overall results are vital for advancing further conversations about student performance. Additionally, CDE may want to consider a future study to document how the SPFs and the DPFs are being used during the UIP process to inform efforts to improve student performance. That is, although the three briefs provided to CDE are useful for informing broader stakeholder understanding of performance outcomes relative to the design of the frameworks and identifying areas for improvement, a critical area that has yet to be evaluated is whether the data from the frameworks are being used in conjunction with other data points collected to help inform decisions to improve student performance. This area will be important to evaluate in order to determine whether the state is meeting a key goal to strengthen local capacity on using data from the frameworks to monitor student performance over time.

³ As also noted in the first brief, this consideration is focused on those schools and districts in the "priority improvement" or "turnaround" accreditation categories since CDE already grants districts the right to lower the ratings of a school if a district deems that the school requires targeted assistance and support.

References

Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4), 91-114.