RECOMMENDATIONS

EMPHASIZE GROWTH, MEASURED OVER MULTIPLE YEARS

SPREAD GRADE RANGES MORE EVENLY

LIMIT FUTURE CHANGES TO THE FORMULA

SEPARATE THE REPORTING AND USE OF PROFICIENCY AND GROWTH

BROADEN AND ALIGN COMPONENTS OF THE ACCOUNTABILITY SYSTEM

EXECUTIVE SUMMARY

School grades were introduced in Florida in 1999, the first such A – F model for reporting on school accountability in the nation. The purpose of school grades was to make it easy for parents and citizens to understand and compare how schools were performing academically, and to compel low-performing schools to improve.

Since then, the A – F grades have been the consistent standard system for measuring school performance results in Florida. However numerous changes in the standards, tests, and even the formula used to calculate school grades over this time has caused much confusion — and occasionally mistrust — over what these grades actually mean.

In the coming year, Florida will be implementing new standards and adopting new assessments that will require elements of the school grades formula to be changed once again. Rather than continuing to make piecemeal, disconnected adjustments whenever circumstances like this arise, Florida should take this opportunity to conduct a top-to-bottom review of the school grading system.

By identifying what has worked well in the past, how students' needs may change in the future, and what else should be updated, state decision-makers can make sure that the system is more stable, consistent and meaningful moving forward.

In this brief, we take a closer look at how the school grades calculation formula works, identify the sources of some of the most prominent issues of confusion in recent years and offer suggestions for how to address these issues now to avoid further problems down the road.

THE CURRENT FORMULA

School grades are an important way that parents and community members evaluate public schools. A fall 2013 poll conducted by the Jacksonville Public Education Fund found that test scores and school grades are by far the primary factor citizens in Duval County use when evaluating a school.[1]

Besides being a source of public scrutiny, school grades lead to state-mandated rewards and consequences for schools. Earning an A or improving one letter grade or more leads to financial rewards through the Florida School Recognition Program.[2] A school earning an F and then one more F or multiple D's in the following few years can mean it must become a district-managed turnaround school; close and reassign students to other schools; close and re-open as one or more charter schools; contract with an outside management group to take over; or implement an approved combination of these.[3]

That's why it is so important to make sure school grades are accurate and meaningful. Florida is now in a critical transition that provides an invaluable opportunity to evaluate what has worked well in the past and what could be improved, to modernize the way school grades are calculated and ensure they are used to best benefit students moving forward.

So how are school grades currently determined? **Table 1** summarizes all of the components that go into the school grade calculation formula as it stands now.[4] The components and scales vary somewhat between elementary, middle, and high schools. For example, graduation and college readiness rates are important for evaluating high school performance but not applicable to elementary or middle schools. A school's final grade is calculated primarily based on the sum total of points earned in each category, as shown in **Table 2**.[5] The current school grading system measures performance predominantly in two ways: proficiency and learning gains (often known as "growth"). Proficiency and growth are very different types of measurements, but both are important in measuring school performance.

Table 1: Current Florida school grades calculation components

|	Elementary 800 pts total	Middle 900 pts total	High 1600 pts total
Proficiency			
Reading	100 pts	100 pts	100 pts
Math	100 pts	100 pts	100 pts
Writing	100 pts	100 pts	100 pts
Science	100 pts	100 pts	100 pts
Growth			
Reading, All students	100 pts	100 pts	100 pts
Reading, Lowest 25%	100 pts	100 pts	100 pts
Math, All students	100 pts	100 pts	100 pts
Math, Lowest 25%	100 pts	100 pts	100 pts
Other			
Accelerated Coursework, Participation		50 pts	150 pts
Accelerated Coursework, Performance		50 pts	150 pts
Graduation Rate, Overall			200 pts
Graduation Rate, At Risk			100 pts
College Readiness, Reading			100 pts
College Readiness, Math			100 pts

Source: Jacksonville Public Education Fund

Table 2: Current Florida school grades thresholds

	Elementary	Middle	High
A	525 pts. +	590 pts. +	1,050 pts. +
B	495 - 524 pts	560 - 589 pts.	990 - 1,049 pts.
C	435 - 494 pts.	490 - 559 pts.	870 - 989 pts.
D	395 - 434 pts.	445 - 489 pts.	790 - 869 pts.
F	< 395 pts.	< 445 pts.	< 790 pts.

Source: Jacksonville Public Education Fund

Measuring Proficiency

PROFICIENCY

Proficiency is primarily a measure of student performance, showing whether students are at, above, or below defined grade-level proficiency targets. Proficiency may also be influenced to varying degrees by other factors (such as health, poverty and varying levels of academic readiness that students enter school with) that are sometimes beyond the school's direct control.

Portion of a school's letter grade determined by proficiency:

50%
Elementary

44%
Middle

25%
High
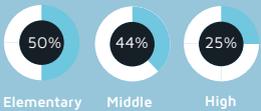
Measuring Growth

GROWTH

Growth is primarily a measure of school performance in terms of how well the school is moving students forward in their learning. It looks at whether students are advancing at or above an appropriate rate each year, regardless of where they began — making it less predicated on outside influences than proficiency alone.

Portion of a school's letter grade determined by growth:

50%
Elementary

44%
Middle

25%
High
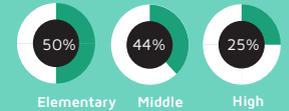
Source: Jacksonville Public Education Fund

The connection between proficiency and growth in the current school grading formula

In the current school grading formula, learning gains are determined by the percentage of students making one year's worth of growth or more from the prior year to the current one (i.e., 1 percent of students making at least one year's growth = 1 point). There are few specific criteria for what constitutes "one year's growth," but they generally reflect either moving up a proficiency level from one year to the next, or maintaining performance at a high proficiency level.[6]

While growth is an important measure of how schools are moving students forward, how it is measured can often cause schools to vary significantly from one year to the next in terms of their growth scores. Student growth should be a primary indicator of school performance, and can be emphasized by maintaining or increasing the value of growth measures in the school grade formula. However, using a single year of growth can cause abnormal spikes or dips for a school in any given year.

Adopting a more rigorous way to measure growth would tell us more of what we want to know about how our schools are moving students forward.

RECOMMENDATION = EMPHASIZE GROWTH, MEASURED OVER MULTIPLE YEARS

Schools can often vary significantly from one year to the next in terms of growth captured by the test. In **Figure 1**, we looked at the total learning gains point totals earned by elementary schools in Duval County in 2012 and 2013.

The black squares show the total growth scores that each individual school earned in 2012. The green dots directly above or below the black squares show the total growth scores earned by the same school in 2013. The bars between the black squares and corresponding green dots show the change in growth between 2012 and 2013. As can be seen, there are often wide differences from year to year.

The wide differences seen at many schools are likely due to a number of factors, including but not limited to different students moving in and out of a school each year, year-to-year changes in test content or proficiency thresholds, or just random error you would expect from administering any given test on a single occasion (for example, some students not paying attention or making mistakes due to test anxiety).

Measuring growth is extremely important in the school grading calculation, and should remain so, but there are issues with using only a single year's change in growth as the current system does. By using just a single snapshot of growth each year, it's possible to miss a number of important contextual factors in the different student populations being compared. There are a number of other options for measuring growth that could help minimize these dramatic single-year differences at any given school.

One possibility would be to calculate the average percent of students making learning gains at a school over multiple years. This could be done by either incorporating a multi-year average of the same type of single year-to-year comparisons currently used, or by using a **cohort model** that tracks only the growth of students who entered and remained in that school across multiple years.

Figure 1: Duval County Public Schools elementary school learning gains points earned in 2012 vs. 2013

Growth scores can vary widely from year to year.
This graph compares 2012 and 2013 growth scores.

2013 Growth Scores
2012 Growth Scores
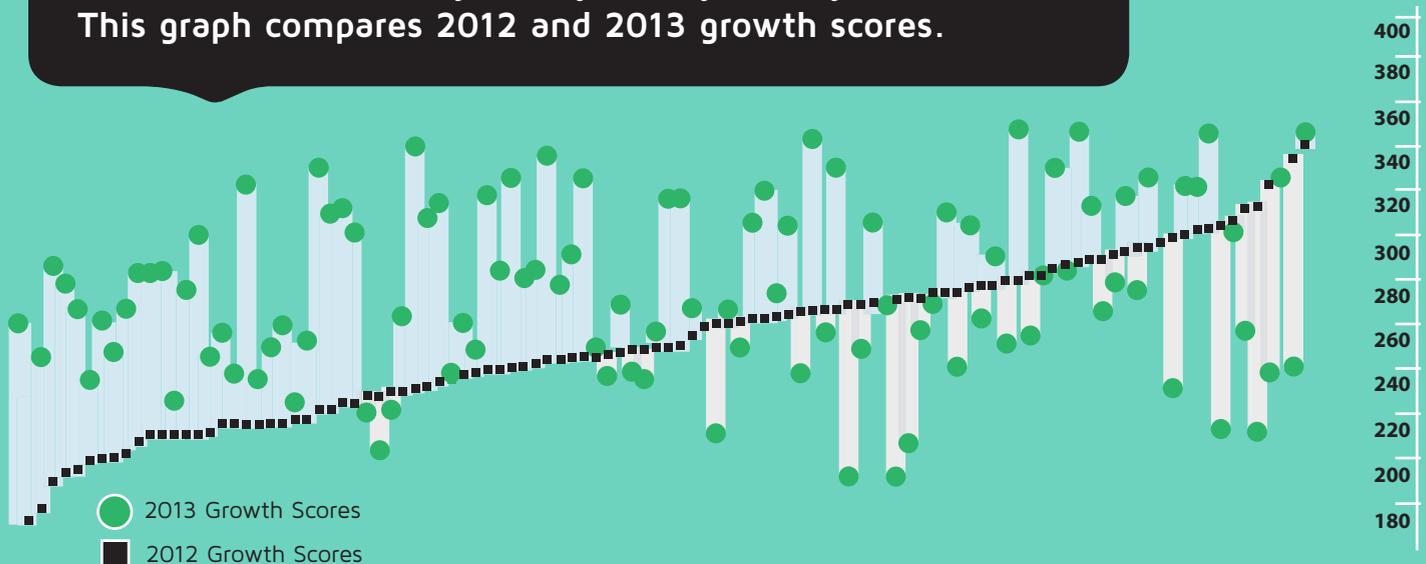
The main difference between these two is that in the **multi-year averages model** (similar to the single-year method currently used), schools would still be held accountable for the growth of students who were enrolled a minimum required amount of time that year, whereas in the cohort model they would only be held responsible for the growth of those students who have been continuously enrolled at that school over multiple years. In either type of model, there would still be consideration of numerous technical rules regarding which students' scores (and what percentage of them) should count towards a school's accountability grade.

Another possibility would be to measure growth using a **value-added model** aligned with the formula used for teacher evaluations.[7] In this type of model, rather than set predetermined growth standards at the same level for all schools statewide, predicted growth levels would be calculated for each school individually based on their own students' prior performance, and schools would be judged by their ability to exceed those predictions (i.e., their overall "value-added" to student learning).

It is worth noting that developing reliable value-added prediction models for individual teachers at the classroom level has been a complicated and often controversial ongoing task; further refining the precision of school-level effects in the model would be just as complex. But it is an idea worth at least considering, and now is the time to put all ideas on the table for open review. The benefit would be a clearer alignment between teacher and school evaluations and potentially, depending on how closely aligned the teacher and school formulas were, an opportunity to track and report both teacher and school effectiveness in the same system without duplicating efforts.

Evaluating how we measure growth

Student growth should be a primary indicator of school performance, and can be emphasized by maintaining or increasing the value of growth measures in the school grade formula. However, using a single year of growth can cause dramatic single-year differences at any given school. Adopting a more rigorous way to measure growth, such as one of the models below, would tell us more of what we want to know about how our schools are moving students forward.

GROWTH MEASUREMENT MODELS

CURRENT MODEL

The current model for measuring growth simply looks at how a student performed in the current year compared to the previous year. By using just a single snapshot of growth each year, it's possible to miss a number of important contextual factors in the different student populations being compared.

MULTI-YEAR AVERAGES MODEL

This option would calculate growth for any given year the same way the current model does, but would incorporate averages for multiple recent years worth of these measurements in determining a school's grade.

COHORT MODEL

In this option, growth would be tracked longitudinally for the same students over multiple years at a school (as opposed to averaging multiple "cross-sectional" measures which may include different students). In other words, schools would only be evaluated on the growth of students who were continuously enrolled there across multiple years.
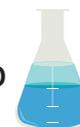
VALUE ADDED MODEL

A value-added model is a statistical formula used to predict student performance levels in a given year, based on prior performance and other factors, and then measure the difference between how the student was predicted to perform and how they actually did. The difference between how a student was predicted to perform and actually performs is considered the "value-added" to the student's learning by the teacher or school.

RECOMMENDATION = SPREAD GRADE RANGES MORE EVENLY

Table 2 (on page 2) lists the current point total thresholds required for earning each A – F accountability grade. When we look at these thresholds laid out along their scales, as in **Figure 2** below, a key issue becomes clear.[8] From this perspective, we see wide bands of possible point totals in which a school could fall into A or F categories, but closely bunched ranges in the middle, with little room separating what constitutes a B, C, or D school. In these ranges outlined in the blue bar, a change as small as 30-60 points in the overall 800-point scale can mean the difference of an entire letter grade.[9]

Figure 2: Comparison of school grade performance thresholds in current scale and possible equidistant scale

Current Grade Model

F		D	C	B		A
395 point span		40	60	30		275 point span
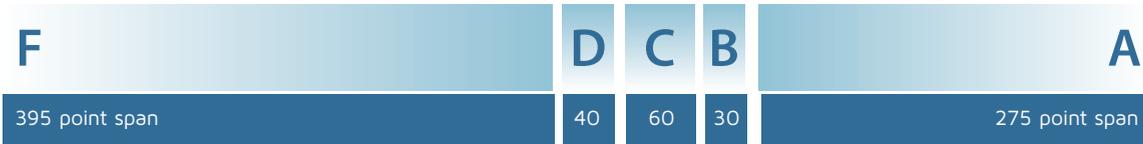
Possible Equidistant Grade Model[11]

F	D	C	B	A
320 point span	120	120	120	120 point span

Figure 3: Distribution of Florida elementary schools by total school grade points earned in 2012 on current scale

Figure 4: Distribution of Florida elementary schools by total school grade points earned in 2012 on possible equidistant scale

CURRENT MODEL

F   D C B   A

EQUIDISTANT MODEL

F   D   C   B   A
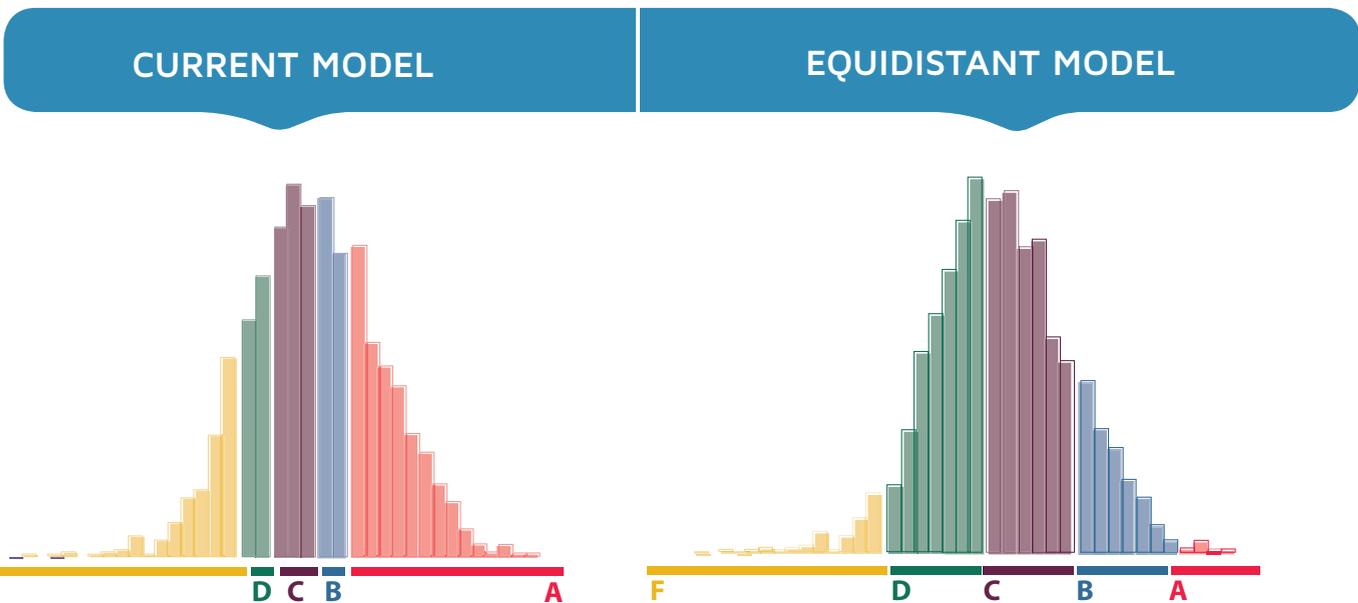
Source: Jacksonville Public Education Fund

To look at it another way, **Figure 3** shows the distribution of all elementary schools in the state based on the total points earned in their school grade calculations last year, with an overlay of the current thresholds. The schools themselves generally form a normal distribution of points earned (most in the middle, fewer on either ends). But the threshold for an A school includes everything from nearly the middle of the pack forward. Behind that the B, C, and D ranges fall much closer together and the F range covers a wide range at the low end.

This is one of the major reasons school grades are vulnerable to volatile changes every time tests or proficiency standards change for any one component area. Most recently, it was the cause of much concern about potential impact of the Writing score projections in 2013.[10]

What this means for an individual school is that a change of just 50 – 100 points in any one component area could potentially swing a school's overall grade by up to two whole letter grades from year to year.

For the broader community, when changes in the formula cause these types of grade swings among schools state-wide, widespread confusion about — and even mistrust in — the grading system often follows.

One way these issues might be addressed is by spreading out the grade ranges more evenly.

The red scale in **Figure 2** shows one possible model, which we will call the "equidistant" scale, in which all grade level cutoffs are more equally spaced out across the scale.[11] Here each letter grade now has an equal range of points needed between cutoffs that makes them easier to understand and explain and more meaningful. It also makes it less likely that changes in any single component area would cause large numbers of schools to move one or two whole letter grades.

Figure 4 shows the same distributions of schools by total points earned as in **Figure 3**, this time using the proposed "equidistant" thresholds. Again, we see the new thresholds are much more evenly spaced out and less vulnerable to small point changes

that can massively shift schools across grade levels. We also see this type of scale would produce fewer A and F schools on either end, and many more B, C and D schools, at least to begin with.

This last point is an important, real-world implication to consider for accountability consequences attached to grades in the short-term if such a model were adopted. The long-term benefit is that moving forward, all grades would be more meaningful and reflective of actual performance in terms of points earned.

Another idea would be to get rid of set point scales altogether and judge all schools by the percentage of possible points earned. That would ensure that the scales for elementary, middle and high schools were the same. And it could create a system that's more relatable to the types of grading scales that people associate with A through F grades, such as 90 percent and above is an A, 80-89 percent a B, and so on.

How evenly spaced point ranges will stabilize school grades

Right now, the thresholds for school grades are not evenly distributed — instead they are bunched narrowly in the point thresholds for B, C, and D grades. That creates a system where even a small change can cause school grades to swing dramatically. That would be similar to trying to drive a car using a speedometer that looks like the one on the left below. Spacing the grade ranges more evenly will make the system less vulnerable to sudden shifts and more easily understandable — like driving with the speedometer on the right.
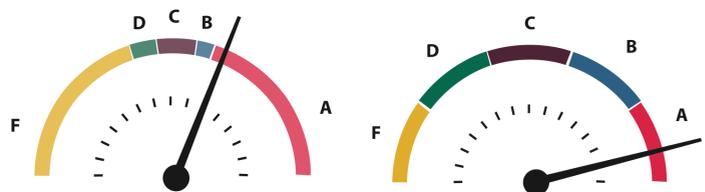
RECOMMENDATION = LIMIT FUTURE CHANGES TO THE FORMULA

School grades were first introduced in Florida in 1999, when the state adopted what was then called the A+ Plan for Education. Since then, the A – F grades have been the consistent standard reference system for reporting and comparing school performance results among all traditional and charter public schools. Some evidence suggests that the introduction of the accountability system itself, along with consistently increasing standards, has compelled performance upward through the combination of rewards, consequences, and increased public scrutiny.[12]

Changes in the formula for calculating school grades, however, have also played a role in changing what constitutes an A (or other grade) school across these years. Perhaps most significantly, the decision to factor in a learning gains component in 2002 has been associated by some with the tremendous spike in A schools around that time.[13] Now, instead of solely being judged on whether or not students were performing "on grade level" each year, schools were also able to earn points toward their grade for significantly moving students forward in their development — regardless of proficiency levels.

Figure 5 shows the percentages of schools achieving each grade level statewide between 1999 and 2012. The most pronounced trend that jumps out is an ongoing rise in A schools between 1999 and 2009, including a dramatic leap between 2001 and 2003. This trend appears to coincide with a decrease in C and D schools around the same time. The addition of learning gains was an important and necessary change, for the reasons discussed on page 3.

But as the system has aged, the formula and components used to calculate what school grades mean has changed dozens of times, including 16 changes since 2010 alone.[14] That doesn't include additional changes to standards, tests and educational policies that could all impact grades.

Changes to the actual standards and tests used to evaluate students, changes in education policies (such as the class-size amendment and expansion of school choice vouchers) and demographic changes in the student population all likely played a role in the fluctuations seen in school grades over this time. To specifically disentangle the individual impact of every factor is beyond the scope of this report. It is only highlighted here to keep in mind that the potential effect of all of these elements must be considered in trying to develop and evaluate a stable and meaningful system moving forward.

In the 2014-2015 school year, Florida is scheduled to fully incorporate the **Common Core State Standards**, and accompanying new assessments, for English/Language Arts and Mathematics. This transition will require a number of changes to the way school grades are measured. At a minimum, it will

PROFICIENCY THRESHOLDS

Proficiency thresholds define the minimum level at which a student must achieve on a test to be considered performing "at grade level" or above. Proficiency thresholds are also sometimes referred to as cut scores.

COMMON CORE STATE STANDARDS

Standards are state-adopted guidelines that specifically define what a student should know or be able to do in each subject area by the end of each grade level. Standards do not define how a teacher must teach the content, or what assessment must be used to measure student achievement of the standards. The Common Core State Standards are a new, multi-state set of English/Language Arts and Mathematics standards designed to promote deeper understanding and make sure all students have the knowledge and skills to be college or career ready in the new economy by graduation. They were developed by a multi-state coalition of educators, education researchers, and policy leaders, and adopted by Florida in 2010.

require the identification and inclusion of new grade-level proficiency thresholds appropriate to the new assessments.

But the transition also provides an invaluable opportunity to review the system as a whole. This is a time to identify critical areas of concern and address them now so that the system moving forward can remain stable, consistent and meaningful for all schools, parents, policymakers and community members.

In considering improvements to the school grades system, a key element will be to acknowledge what sources of instability can and cannot be controlled. Then, safeguards can be put in place so that changes to the areas controlled by the state are as limited and infrequent as possible.

One possibility would be to schedule regular adjustment windows, during which any elements of the content, formula or assessments can be reviewed and updated. Between those windows however, no changes can be made to those elements of system. For example, if adjustment windows were set for every five years, during those windows the state could review and adopt changes to standards, content, assessments, and calculations as seen fit, but during the five year periods between windows the system would remain unchanged to provide consistency and stability for students, schools, parents and community members.

How to minimize change without sacrificing accuracy

As Florida's school grading system — the nation's oldest — has aged, numerous and frequent changes to the grading formula have caused considerable confusion about what the grades mean. Building stability into the system would mean limiting formula adjustments to keep comparisons and expectations consistent from year to year, while leaving pre-defined opportunities for necessary future updates.

The state should plan for adjustment windows to allow for any needed changes to assessments, proficiency standards or formula calculations every five years, with a moratorium on any changes in between. That also means planning for the fair application of accountability rewards and consequences during transition years when tests, expectations or calculations change.

Figure 5: Percentage of Florida schools achieving each grade level, 1999-2012

2002 Learning Gains integrated into the school grades formula
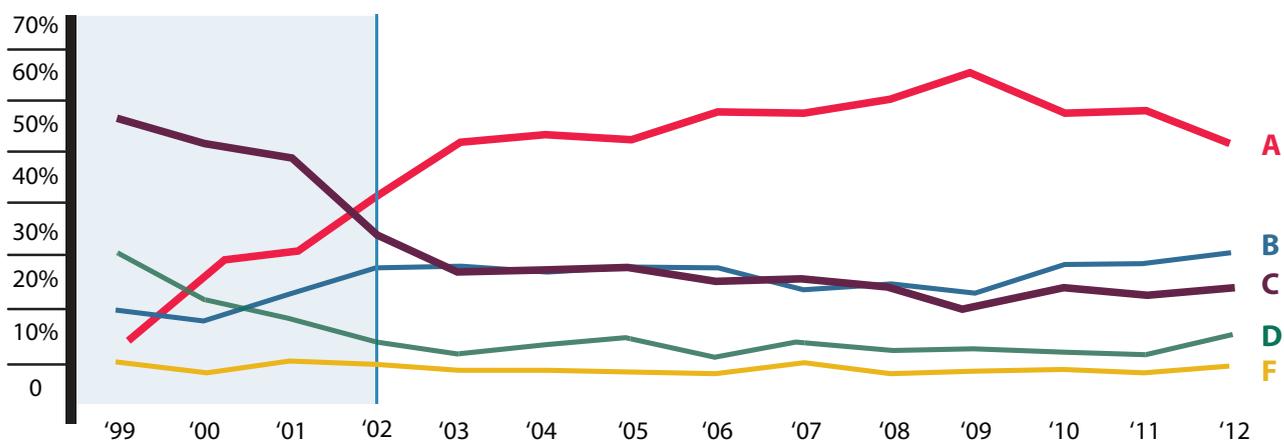
Source: Jacksonville Public Education Fund

While it is important to find the right balance in holding schools accountable for both proficiency and growth rates of their students, another source of potential confusion in the current system stems from the fact that these two very different measurements are currently combined and reported as a single score.

In **Figure 6**, we broke down the 2013 elementary school grades for Duval County into their proficiency and growth components. The size and color of the marks indicate the percent of students who are eligible for free and reduced price lunch (FRL) at that school. The shape of the marks indicate whether a school is a traditional or charter school.[15]

By separating these components in this way, much more meaningful and useful information emerges right away.

For example, a number of high FRL enrollment schools that are below average in students meeting grade-level proficiency expectations are actually performing above average in growth (upper left area). This may indicate lower proficiency levels here are more connected to other issues outside of school, such as where they were when they entered compared to students at other schools.

On the other hand, some elementary schools are behind in proficiency and also very much behind in the percent of students showing appropriate growth last year (lower left area).

Because proficiency and growth components are combined in school grades, some schools in these two different areas ended up with the same overall school grade last year. But when we look at it this way, we get much more meaningful information and can see that we would perhaps want to reward and further develop whatever those schools in the upper-left area (high growth) are doing, while taking a closer look at whether accountability actions are needed to change what is happening at those schools in the lower-left (low growth) area.

One option for addressing the problem of confounding proficiency and growth into a single grade may be to separate the two types of measures for reporting, accountability and support purposes. The growth component could be used primarily for recognition and accountability actions. Proficiency could help determine where more resources are needed. For example, a school with low proficiency levels but high growth may be identified as a model for making gains with students who need help the most, and targeted for more support to expand what they are doing even further until proficiency levels catch up.

Figure 6: DCPS elementary school performance plotted by growth vs. proficiency components
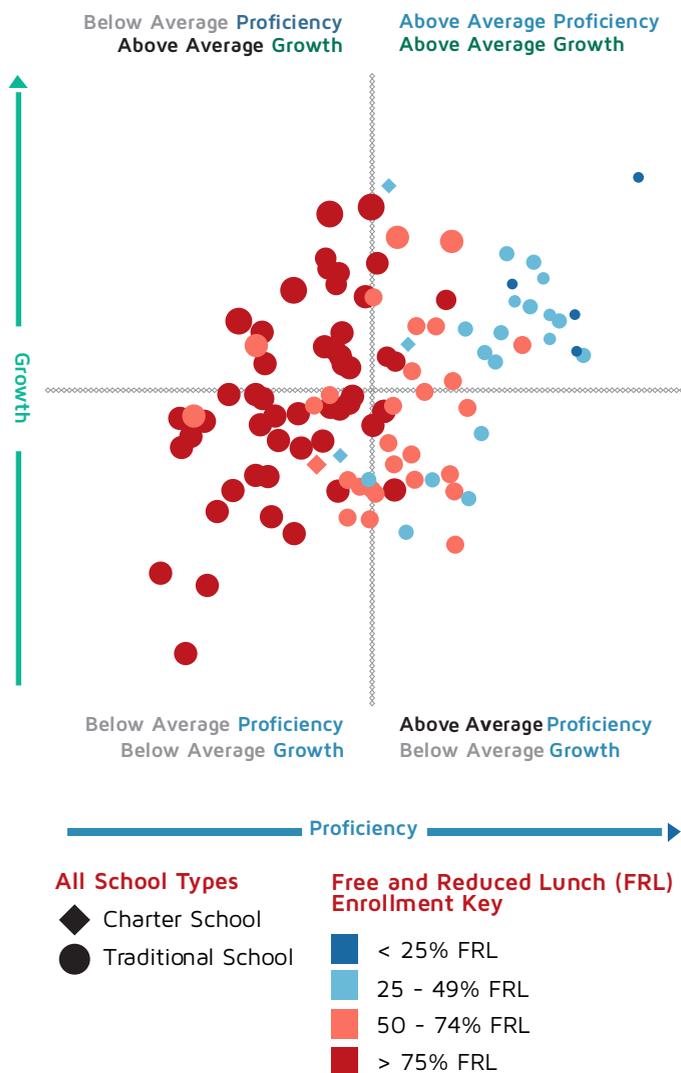
Below Average **Proficiency**
Above Average **Growth**

Above Average **Proficiency**
Above Average **Growth**

Growth

Below Average **Proficiency**
Below Average **Growth**

Above Average **Proficiency**
Below Average **Growth**

Proficiency

All School Types
◆ Charter School
● Traditional School

Free and Reduced Lunch (FRL) Enrollment Key
< 25% FRL
25 - 49% FRL
50 - 74% FRL
> 75% FRL

Source: Jacksonville Public Education Fund

In addition to the specific technical issues discussed here, now is the time to fully review whether everything we want schools to be focused on is incorporated as part of the school grade formula. Additional components that could be considered to provide a more holistic accountability system are factors such as attendance, discipline/safety measures, percentages of highly effective teachers or diversity of course options offered. In addition, this is the time to make sure the systems for evaluating schools, teachers, principals and district administrators are as aligned as possible in terms of what they measure and how they measure them.

REIMAGINING SCHOOL GRADES

The issues and opportunities explored in this report are just a few important areas among many that should be considered in a comprehensive review of what works best in the current school accountability system, and what the state has learned over the last 14 years that can help improve and solidify the system in the future.

As the state prepares for its transition to the next set of standards and assessments that we expect to better prepare students to succeed in the world beyond graduation, it is critical that we use this opportunity to make sure we have the most accurate, understandable, and stable system of accountability possible. This will ensure that all students are in schools that can deliver on that promise.

To be clear, the past 14 years of Florida's accountability system has driven many positive changes in the education of Florida's students, often compelling significant improvement simply by shining a light on school performance that did not exist before.

As with any successful system or product, it is always important to look back at regular intervals and consider what has worked well, what could be improved, and what may or may not be still aligned with our needs moving forward into a new era.

As a state, we have a pivotal opportunity to do that right now, an opportunity that should take into account the input of all stakeholders — students, parents, educators and citizens. To learn more about how you can be involved, and explore some more interactive information about the issues and possibilities explored here, visit **www.jaxpef.org**.

ENDNOTES:

1 – Jacksonville Public Education Fund (2013). Annual Survey of Education Perceptions in Duval County. Retrieved from http://www.jaxpef.org/media/1465507/jaxpef_2013_annual_education_perceptions_poll.pdf

2 – For more information about Florida's School Recognition Program, visit: http://www.fldoe.org/faq/default.asp?Dept=177&ID=613

3 – For more information, visit: http://www.fldoe.org/board/meetings/2013_06_18/TOPS.pdf

4 – Note: Combination (K-8, K-12, 6-12) use a slightly different scale not included here. For more information on all scales, see: http://schoolgrades.fldoe.org/pdf/1213/Guidesheet2013ShoolGrades.pdf

5 – Some final points adjustments may be made based on additional decision rules, for more information see: http://schoolgrades.fldoe.org/pdf/1213/Guidesheet2013ShoolGrades.pdf

6 – For more information, see http://schoolgrades.fldoe.org/pdf/1213/Guidesheet2013ShoolGrades.pdf

7 – For more information, see: http://www.fldoe.org/committees/pdf/PresentationValue-addedModel.pdf

8 – For the purpose of these figures, only the elementary school scales were featured. Middle, combination and high schools have different scales but exhibit the same issues.

9 – Ibid

10 – For more information, see http://www.jaxpef.org/news/2013/06/whats-happening-with-fcat-20-writing-standards.aspx

11 – Because approximately 99% of elementary schools statewide scored over 200 points on the 800 point (ES) scale, the equidistant ranges proposed here are spaced evenly from 200-800 (as opposed to 0-800) with anything below 200 also considered an F.

12 – West, M. R., & Peterson, P. E. (2006). The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments. The Economic Journal, 116(510), C46-C62.; Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure (No. w13681). National Bureau of Economic Research.

13 – DiCarlo, M. Why did Florida schools' grades improve dramatically between 1999 and 2005?. Retrieved August 20, 2013, from http://shankerblog.org/?p=7642

14 – O'Connor, J. Is Florida's school grading system too complicated?. Retrieved August 20, 2013, from: http://stateimpact.npr.org/florda/2013/08/13/is-floridas-school-grading-system-too-complicated/

15 – To learn more and find your school on a fully interactive version of this figure, visit: http://www.schoolfactsjax.org

SCHOOL GRADES | Brief IN

A REPORT FROM
THE JACKSONVILLE PUBLIC
EDUCATION FUND

In this report, we examine Florida's school grading system and make suggestions to modernize an aging system.

The Jacksonville Public Education Fund is an independent nonprofit organization that works to connect research with civic voice to bring about unified action in support of universally high-quality public schools for all children in Duval County. We believe that an informed and active citizenry, together with attentive and aligned community and district leadership, will move our schools forward to prepare students with the skills they need for future success.

JACKSONVILLE
Public Education Fund

245 Riverside Avenue, Suite 310
Jacksonville, FL 32202

FOR MORE INFORMATION, CALL (904)356-7757 OR VISIT WWW.JAXPEF.ORG