Continuous Improvement

A Report on “Year One” of the revised DPAS-II Educator Evaluation System

November 2013

Prepared By:
Teacher and Leader Effectiveness Unit (TLEU)
Delaware Department of Education
Table of Contents

EXECUTIVE SUMMARY .................................................................4
BACKGROUND & INTRODUCTION ......................................................8
OVERVIEW OF DPAS-II & “COMPONENT V” ..................................11
DISTRIBUTION OF COMPONENT I-IV RATINGS .............................16
DISTRIBUTION OF “COMPONENT V” RATINGS .............................19
2012-2013 SUMMATIVE RATINGS .................................................26
IMPROVING DPAS-II: FEEDBACK FROM EDUCATORS ...............29
IMPROVING DPAS-II: 2013-14 & BEYOND ....................................34
Closer Look: Charter School DPAS-II Evaluation Distribution ......36
Closer Look: Statewide Student Achievement in 2012-2013 ..........37
November 6, 2013

There is nothing we do that is more important than educating our children. They have boundless hope, and represent the future of our communities and our economy. Over the last few years, Delaware’s education system has made great progress, and we have a long way to go. As we prepare our children for the next chapter of their lives, we know that less than half graduate from high school ready for college or a career. We must do better.

Teachers are the most important school-related factor in the academic success of our students. And our educational system is only as strong as our educators who work with their students every day. That’s why it is essential that we have an evaluation system that leads to an accurate assessment of the quality of practice and results our educators demonstrate; an assessment that will lead us to better educator support and accountability, while helping drive policy decisions that will best support the needs of our schools.

Although Delaware has had an educator evaluation system for more than twenty years, we know that it has often fallen short in providing adequate and authentic feedback, support, and accountability for our educators. As part of our Race to the Top plan, we committed to linking robust measures of student achievement to our educator evaluation processes. Over the past three years, hundreds of educators have come together to help us develop a revised system, which evaluates every educator in our state on multiple measures of student growth while maintaining a heavy focus on classroom observations. The revised system was implemented statewide in 2012-13, and the results are set forth in this report.

The report shows that the improved evaluation system and process has placed an increased focus on individual student achievement. Specifically, educators in Delaware met with their school leaders, set goals for student growth, and evaluated that growth at the end of the school year. Not surprisingly, the growth teachers made with their students varied, as you will read in this report.

The report also shows there is much more work to be done. In this first year, we saw almost no differences among educators on the observational evaluation components, and we often saw school administrators using discretion to set less rigorous goals and upgrading educators to higher ratings within the student growth component. The result – showing there was very little variation in overall educator performance ratings in 2012-2013 – was troubling, and should prompt greater urgency in implementation.

Our results are similar to other states that are implementing rigorous educator evaluation systems, but we should not take comfort in the fact that we are not alone. Going forward, we need to ensure that school leaders and educators implement the system well, so that our ratings reflect the reality of teaching happening in our classrooms. The system will work best when we have educators and evaluators who understand the value of the system, and feel empowered to use it to provide a valid assessment of teaching practices and to drive improvement in their classrooms. As our educators and administrators work to improve their practice, the Department will simultaneously work to improve the overall system. As the title of this report suggests, we are all committed to continuous improvement.

Sincerely,

Mark Murphy, Secretary of Education
EXECUTIVE SUMMARY

Delaware has invested heavily in revising its educator evaluation system – the Delaware Performance Appraisal System (DPAS-II) – as an essential component of the state’s efforts to offer all of its students a quality education. Research has found that the most important school-related factor in students’ academic success is teacher quality. Although Delaware has had an educator evaluation system since the 1980s, it did not have a consistent or sophisticated measure of each educator’s impact on student growth. Thus, the state was unable to identify those educators and administrators who were struggling to serve students, and who needed extra support. The reforms set in place over the last four years, and implemented statewide in the 2012-13 school year, were designed to change that. The most important and significant change – the revised measure of an educator’s impact on student achievement (“Component V”) – was designed to give more detailed and objective information about the performance of Delaware students and educators, to help understand where they are struggling and why, and to allow school leaders and decision-makers to help all educators improve their performance in the classroom.

The increased focus on student growth was particularly crucial given the persistent socio-economic and racial achievement gaps among Delaware students and growing evidence that many students are not “college and career-ready” upon graduation.

One of the most important roles of this system is to ensure educators and their evaluators have ongoing, meaningful dialogue about classroom performance that results in concrete feedback on ways for educators and students to improve, with these conversations grounded in rich data. The engagement of educators and their evaluators has been a crucial element in driving change for Delaware students. During the 2012-13 school year, for the first time, all educators and school leaders in Delaware’s schools had conversations with their evaluators about the type of growth they wanted to see their students make and how they would hold themselves accountable for helping students reach these goals. All educators had access to pre- and post-assessments created by Delaware educators they could use to measure their students’ growth. Teachers of math and English (in grades 3-10) received individualized targets for each of their students that accounted for where the student started and indicated the type of growth similar students made in previous years. This provided a nuanced view of student achievement that was not possible before in Delaware. For instance, students deemed proficient based upon the state assessment, yet not growing to expectations, and those not yet reaching proficiency, but growing beyond expectations, could be identified. These changes were the result of deliberate system enhancements and the inclusion of multiple measures of student growth in the appraisal of an educator’s practice.

The efforts to reform DPAS-II also sought, for the first time, to show meaningful variation in educator performance. Over the years, Delaware’s evaluation system has shown virtually no such variation, making it difficult to pinpoint which educators are having the most success and to help all teachers improve. This lack of meaningful data made it difficult to identify weaknesses,
acknowledge strengths, and recognize distinguished teaching practice. Meanwhile, this information is essential for an educator’s professional growth and for state, district, and school leaders making decisions about professional development resources, recognition, promotion, student assignments, and dismissal.

This report—Continuous Improvement—provides an analysis of the first year of statewide implementation of the revised DPAS-II system using student achievement data, educators’ evaluation ratings, and educators’ responses to an annual survey about the DPAS-II process. The results show a system that has made significant progress in implementing and institutionalizing appraisals of professional practice through student growth. At the same time, the results show that there is more work to be done, as a meaningful level of differentiation was not evident in the traditional teacher evaluation metrics (Components I-IV), which resulted in overall ratings that showed little variation in teaching quality. Nearly all educators receive positive ratings on classroom observations and where evaluators have the opportunity to exercise professional judgment (i.e. student growth goal-setting by educators or administrator discretion to upgrade ratings) it generally leads to higher ratings and less variation, which makes it harder to help educators improve. And while Component V data is showing some meaningful differentiation in student growth measures defined by the state (“Measure A”), the report outlines how these potential opportunities for improvement are ultimately rated as “Effective.”

The key findings about the DPAS-II system discussed in the report include:

There was more variation among educators in the state-defined metric for student growth than in measures of student growth where targets were set by educators and their evaluators:

- Using only the state-defined measure of student growth based on the Delaware Comprehensive Assessment System (Measure A) for educators in math and English (grades 3-10), 54 percent of educators received an “Exceeds” rating, 29 percent were rated “Satisfactory,” 12 percent “Unsatisfactory with administrator discretion” and 5 percent “Unsatisfactory.”

- For the other student growth measures, defined using targets set by an educator and their evaluator, the vast majority of educators received “Satisfactory” or “Exceeds” ratings: 94 percent for Measure B and 99 percent for Measure C.

- When the state set student growth targets (Measure A), more educators were rated “Unsatisfactory” than when educators and evaluators set growth targets (Measures B). Specifically, only 13 percent of educators who were rated “Unsatisfactory” on Measure A were also rated “Unsatisfactory” on Measure B. Half of educators who were rated “Unsatisfactory” on Measure A were rated “Exceeds” on Measure B. 37 percent of that same group were “Satisfactory” on Measure B when Measure A was “Unsatisfactory”.

- Ninety-nine percent of Delaware educators received “Satisfactory” ratings on each of the four components based upon classroom observations and professional responsibilities
(Components I – IV). There is almost no variation across districts in the percent of educators with four “Satisfactory” ratings on the qualitative components of the evaluation system (as assessed by school and district leadership).

- Overall, 51 percent of educators received the newly-created “Highly-effective” rating and 48% received an “Effective” rating. Only 1% of educators received a rating of “Needs Improvement.”

**Where the DPAS-II system allowed for evaluator discretion in an educator’s rating, there was a tendency to upgrade to higher ratings.**

- Evaluators opted to upgrade an educator’s rating to “Satisfactory” in 87 percent of the instances where a “Satisfactory” rating on one measure of student growth and “Unsatisfactory” on the other needed to be combined for a final Component V rating.

- Across the state, 56 percent of educators whose Measure A rating were initially “Unsatisfactory with administrator discretion” were upgraded to “Satisfactory” ratings by their evaluators. In some districts, ratings were upgraded in 60 to 90 percent of cases.

- Taking these decisions into account, educators’ overall ratings on the student growth component (Component V) showed that 53 percent were rated “Exceeds,” 46% were rated “Satisfactory”, and 1 percent were rated “Unsatisfactory.”

**While the majority of educators find value in the DPAS-II system and the student growth Component V for improving their teaching, they overwhelmingly believe the system needs improvement.**

- In a survey of educator perceptions, 64 percent believe the overall evaluation system has “some” or “major” impact on improving their teaching; 64 percent also believe the student improvement component (Component V) has “some” or “major” impact on improving their teaching.

- Eight-five percent of teacher respondents agreed that the oral feedback they receive through the DPAS-II process is “useful and applicable” and 84 percent find their DPAS-II conferences to be valuable. Seventy-two percent of teacher respondents believe DPAS-II is at least one of the top five efforts to drive student achievement gains.

- While 81 percent of educator respondents believe the evaluation process overall is of “high” or “medium” quality, the DPAS-II system was given a grade of “C” by the majority of teachers, specialists, and administrators.

These results highlight the importance of the work, and the need for continuing improvement of the evaluation system and building of capacity in Delaware schools and districts to implement the system well. Although the system provides some meaningful data about student growth in
educators’ classrooms, that data is not translating into meaningfully differentiated results in the overall rating system.

The Department continues to solicit feedback—the state’s educators express different perspectives, but the consistent theme is that the system needs improvement (to allow for more focus, differentiation based on roles and experience, peer observation, and emphasis on educator development). To address these concerns, the report concludes with some of the efforts underway or on the horizon to strengthen the DPAS-II system. These efforts include continued funding for coaches to support school leaders with the evaluation process, specific training in student growth goal-setting, and regulatory changes that allow for more flexibility in the use of unannounced observations and lessen the observation requirement for novice specialists.

The efforts also include the strengthening of alternative assessments used to measure student growth, an approval process for districts to use alternative evaluation systems to meet the needs of their educators, and recent amendments to regulations to allow districts to credential peer observers. Thus, the work of improving educator evaluation in Delaware continues at the school, district, and state levels, making the necessary adjustments to drive stronger implementation in the years ahead.
BACKGROUND & INTRODUCTION

Delaware students are achieving greater levels of academic success. Over the past two years, 7 out of 10 Delaware students have scored at the “proficient” level in math and reading on state tests. A larger share of students reached their individual growth targets in reading and mathematics in 2012-13 than the previous year, and more students reached advanced levels of proficiency in middle and high school. These rates represent significant progress when considering that only about 60 percent of Delaware students were considered proficient in the 2010-2011 school year. In fact, a recent study ranked Delaware as the 3rd highest-growth state in the nation when looking at student achievement gains over the last 20 years.¹ Such gains are the result of the hard work of Delaware educators and the collective focus of all concerned parties—parents, community representatives, and district and state policy leaders—on the growth and development of Delaware’s students.

With students’ long-term success at the forefront of the statewide discussion about Delaware’s educational improvement, any appreciation of recent gains is tempered by evidence that many Delaware students are not ready for success in college and careers. For example, only 30 percent of Delaware high school freshmen make it to their sophomore year of college and only 17 percent of low-income ninth-graders persist to their second year of college. About 35 percent of white ninth-graders remain in college by their second year, compared to 14 percent of Hispanic and 22 percent of black ninth-graders.² Further, recent evidence suggests that proficiency on the Delaware state assessment is not a strong indicator of readiness for college or career. Only 8 percent of students who scored a “3” (“Proficient”) on the Delaware state reading exam in 10th grade, met the “college-ready” benchmark on the corresponding SAT college readiness exam. Only half of the students scoring a “4” (“Advanced”) in 10th grade reading met the corresponding SAT benchmark. As the state, districts, and educators review these indicators, it is clear that there is substantial work to be done to raise standards and prepare Delaware students for academic, career, and life success.

To continue to raise standards and ensure our students are prepared for college and careers, an emphasis on educator effectiveness and performance is imperative. Research indicates that teacher effectiveness has a larger impact on student achievement than other factors within the direct control of school systems (e.g. class size, the school a student attends, etc.).³ Delaware has been a leading state for many years in acknowledging this critical role educators play by recognizing that their professional growth and continuous learning is essential for successful schools. Over the past three decades Delaware has implemented an annual statewide educator evaluation system, Delaware Performance Appraisal System (DPAS), which is based upon the principles of the Charlotte Danielson framework. This system was designed to help educators receive the feedback and coaching needed to be effective in their work. DPAS was revised to form “DPAS-II” and was piloted in 2005-2006 based on feedback from educators. It was ultimately launched statewide in 2008.

¹ "Is the U.S, Catching Up?” by Eric A. Hanushek, Paul E. Peterson, and Ludger Woessmann
² Delaware College Going Diagnostic, Harvard Strategic Data Project (2013)
As educators, administrators, policymakers and community members worked together on what would become the state’s top-ranked Race to the Top (RTTT) grant application, the commitment to improving educator effectiveness became a defining feature of Delaware’s plan. The plan established statewide professional learning communities (PLCs) to facilitate collaboration among educators and strengthen their ability to use data to inform instruction. The RTTT application also called for significant investments in coaching for both teachers and school leaders. Professional development for teachers and leaders would be provided to help educators raise standards (through Common Core State Standards) and expose students to more academic rigor (e.g. through Advanced Placement courses). The plan also committed to conducting a statewide teaching and learning survey (“TELL Delaware”) to better understand educators’ access to the conditions that research has shown to be important for effective teaching and learning.

The RTTT plan also called for large investments in the improvement of the DPAS-II system given its direct link to all of the other efforts underway to improve practice. An improved appraisal system would allow educators and school leaders to receive ongoing feedback and coaching tailored to their needs. It would also help gauge the extent to which the many efforts underway (ongoing professional development, involvement in PLCs, etc.) were translating to educator learning and professional growth. A revised DPAS-II system would identify strengths and weaknesses in educator practice, triangulate this information with student growth/performance data to tailor professional development for educators and to make decisions regarding promotion, advancement, retention and removal.

The imperative for improving the DPAS-II system was prompted by evidence that it was not matching its promise as a driver of professional growth and ultimately student achievement. Over the last five years, 99 percent of educators were rated “Effective” using the DPAS-II system. The nearly uniform “Effective” ratings received by Delaware teachers each year were a stark contrast to the rates of proficiency and college readiness seen among Delaware students. In an attempt to address this disconnect and place a stronger focus on student learning in the DPAS-II process, Senate Bill 263 was passed in 2010, thus refining the Student Improvement component of DPAS-II evaluations—also referred to as Component Five (Component V).

DPAS-II evaluations are comprised of five components based on Charlotte Danielson’s “Framework for Effective Teaching,” with components I-IV ratings determined by observations by an educator’s evaluator. Prior to changes to Delaware state code and regulation in 2010, Component V (Student Improvement) was implemented with limited rigor and comparability at the local level, only sometimes determined based upon goals for student learning set by a teacher and approved by his or her evaluator. Unlike Components I-IV, where evaluations were designed to be determined based upon an educator performance rubric, there was little guidance as to how to evaluate student improvement for Component V. As a result, there was significant variability in how Component V was implemented across schools and classrooms and the types of evidence collected to support an educator’s rating. This meant that some educators across the state were engaging in rich conversations about their impact on student improvement during the DPAS-II appraisal process while others were not receiving critical or constructive feedback grounded in student performance. Student growth was not a crucial factor in determining educator
performance—and in some cases it was a non-factor. This meant that in many cases educators received “Effective” ratings with limited to no evidence of student improvement.

Component V was therefore given significant attention as the state reformed the DPAS-II process. Under new regulations, no educator could receive an “Effective” or the newly-created “Highly-effective” rating without “Satisfactory” levels of student growth. The student improvement component would now also require a measurement of student growth using the statewide assessment for administrators and teachers in tested subjects (mathematics and English language arts) and other rigorous measures of student learning for non-tested subjects and grades. The revised system would take effect in the 2012-2013 school year and require that every educator, specialist and administrator have “multiple measures” of student growth included in his or her performance appraisal.

These changes to the state’s educator evaluation system were a major undertaking and would require the support of the many stakeholders. Even in a small state such as Delaware—with a little more than 10,000 educators and 220 schools—the implementation process would be complex. Assessments to measure student growth needed to be developed and disseminated, advisory committees representing myriad stakeholder groups needed to be convened, trainings needed to be designed and conducted, and districts and schools needed to communicate and implement the changes to the system. To Delaware’s advantage, the state's strong record of stakeholder engagement and the ongoing support of all districts, charters, the Delaware State Education Association (DSEA) and Delaware Association of School Administrators (DASA) and state policymakers were signaled by their commitment and signature on the state’s RTTT plan.

This report, Continuous Improvement, presents an account of last year’s (2012-2013) implementation of DPAS-II and an analysis of statewide data collected on the system to assess its effects and impact. These data include 2012-2013 student growth data from the Delaware Comprehensive Assessment System (DCAS) exam, DPAS-II evaluation ratings collected in the state’s Evaluation Reporting System (ERS) and information collected from the annual DPAS-II Process Evaluation survey (educator perception/input).

To contextualize the data presented, the first section of this report provides an overview of the DPAS-II system and the changes to Component V for the 2012-2013 school year. It provides details on the goals of the DPAS-II system and the policy guiding statewide implementation. The second section provides data on the distribution of educator ratings for Components I-V and compares these ratings with student achievement data. The subsequent sections include an analysis of data on the student growth component (Component V) and the distribution of overall summative ratings. The final two sections begin to look ahead at how the DPAS-II system can be improved. The first of these sections considers improvements suggested through data from the annual DPAS-II educator survey and the concluding section outlines policy changes and educator supports already being executed in 2013-2014.

---

4 This measure included individual student growth targets for each student in grades 3-10 for math and English language arts. These targets were calculated based on a student’s past performance as well as the performance of similar students over the prior two years. See page 15 for more information about how these targets were set.
A. OVERVIEW OF DPAS-II & “COMPONENT V”

Delaware’s performance appraisal system has been in place in some form since the 1980s with a goal to assure educators’ professional growth and quality educators in every classroom. In addition to naming the standards for effective teaching and school leadership, DPAS-II provides a forum for collection of evidence needed to make important decisions such as:

- Recognizing and praising outstanding practice
- Recommending continued employment and/or career growth opportunities
- Recommending strategies and/or activities that will enhance effectiveness
- Beginning educator termination proceedings
- Determining the type of professional development needed at a school/district

The DPAS-II system for teachers and specialists is largely based on the work of Charlotte Danielson and the DPAS-II system for administrators is grounded in the Interstate School Leaders Licensure Consortium (ISLLC). For each group, DPAS-II has five appraisal components. The first four components of DPAS-II are grounded in the Danielson Framework and the ISLLC standards, respectively; the fifth is Student Improvement. The DPAS-II Appraisal Components for teachers, specialists, and administrators are listed below:

<table>
<thead>
<tr>
<th>Component I</th>
<th>Teacher</th>
<th>Specialist</th>
<th>Administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning and Preparation</td>
<td>Planning and Preparation</td>
<td>Vision and Goals</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component II</th>
<th>Teacher</th>
<th>Specialist</th>
<th>Administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom Environment</td>
<td>The Environment</td>
<td>Culture of Learning</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component III</th>
<th>Teacher</th>
<th>Specialist</th>
<th>Administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction</td>
<td>Delivery of Services</td>
<td>Management</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component IV</th>
<th>Teacher</th>
<th>Specialist</th>
<th>Administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional Responsibilities</td>
<td>Professional Responsibilities</td>
<td>Professional Responsibilities</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Component V</th>
<th>Teacher</th>
<th>Specialist</th>
<th>Administrator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Improvement</td>
<td>Student Improvement</td>
<td>Student Improvement</td>
<td></td>
</tr>
</tbody>
</table>

Classroom observations\(^5\) form the basis of teachers’ ratings on the first four DPAS-II components\(^6\) on which teachers can receive either a “Satisfactory” or “Unsatisfactory” rating. A

---

\(^5\) Evidence for performance on Components I-III for teachers and specialists is gathered through observation by administrators (novice teachers receive three observations and experienced teachers receive two), which usually follows a pre-observation form, pre-conference and a formative feedback conference. Evidence for performance on Components I-III for administrators is gathered through a survey completed by professional staff, the administrator’s self-assessment on the ISLLC standards and the evaluator’s survey data. For Component IV, all educators complete a professional responsibilities form, which details their professional growth, communication with students, parents, and school colleagues, and their contributions to the professional community during the review period.
teacher or specialist’s rating (either “Exceeds,” “Satisfactory,” or “Unsatisfactory”) on Component V is based on multiple measures of student growth during the year. These measures will be discussed in further detail. A teacher or specialist can receive any of the summative ratings (highly-effective, effective, needs improvement or ineffective) during their “Summative Evaluation”, which occurs every other year for “Experienced” educators and every year for “Novice” educators. The five DPAS-II components combined form the basis of a summative rating, as outlined in the chart below:

<table>
<thead>
<tr>
<th>Total # of Satisfactory ratings in Components I-IV</th>
<th>Component V</th>
<th>Summative Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>4/4</td>
<td>Exceeds</td>
<td>Highly Effective</td>
</tr>
<tr>
<td>4/4</td>
<td>Satisfactory</td>
<td>Effective</td>
</tr>
<tr>
<td>4/4</td>
<td>Unsatisfactory</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>3/4</td>
<td>Exceeds</td>
<td>Highly Effective</td>
</tr>
<tr>
<td>3/4</td>
<td>Satisfactory</td>
<td>Effective</td>
</tr>
<tr>
<td>3/4</td>
<td>Unsatisfactory</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>2/4</td>
<td>Exceeds</td>
<td>Effective</td>
</tr>
<tr>
<td>2/4</td>
<td>Satisfactory</td>
<td>Effective</td>
</tr>
<tr>
<td>2/4</td>
<td>Unsatisfactory</td>
<td>Ineffective</td>
</tr>
<tr>
<td>1/4</td>
<td>Exceeds</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>1/4</td>
<td>Satisfactory</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>1/4</td>
<td>Unsatisfactory</td>
<td>Ineffective</td>
</tr>
<tr>
<td>0/4</td>
<td>Exceeds</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>0/4</td>
<td>Satisfactory</td>
<td>Needs Improvement</td>
</tr>
<tr>
<td>0/4</td>
<td>Unsatisfactory</td>
<td>Ineffective</td>
</tr>
</tbody>
</table>

The newly-created rating of “Highly-effective” was implemented for the first time in the 2012-2013 school year. Another important change in 2012-2013 was that an educator could not earn a summative rating of “Effective” if they had an “Unsatisfactory” rating in Component V. The opposite also was true. An educator could not be rated “ineffective” if his or her Component V rating was “Satisfactory.”

---

6 Under each component are criteria observed by a teacher’s evaluator using a rubric defining “Unsatisfactory,” “Basic,” “Proficient” and “Distinguished” performance

7 Administrators’ Component V ratings are also based on student growth, though administrators have the option of using multiple measures or a single measure (DCAS results).
Overview of Component V

The most significant changes in 2012-2013 occurred within Component V (Student Improvement) of the DPAS-II system. In 2012-2013, all educators were required to show evidence of student growth as a part of their DPAS-II evaluation. All educators would receive Component V ratings annually, even during an “off-year” for a summative evaluation. As discussed, state regulations required that measures of student growth be included in the DPAS-II evaluation process. However, defining fair and appropriate measures and targets for student growth for all educators is complex. The current framework, which requires multiple measures of student growth as a basis for Component V ratings, was the result of significant collaboration among educators (in all grades and subjects), policymakers, and content facilitators at the DDOE.

The current framework for Component V places educators into three distinct groups with three different measures available for Component V implementation. Policy implementation relies upon the professionalism of educators in setting ambitious student growth and/or professional growth targets. The groups are outlined as follows:

- **Group I Educators** include anyone who instructs reading and/or mathematics in grades 3 through 10 and is the educator-of-record for at least 10 students. These grade levels are tested by the Delaware Comprehensive Assessment System (DCAS). The state test is one of their measures.

- **Group II Educators** include anyone who teaches in any grade or subject other than ELA and/or mathematics (DCAS-tested), teaches formal “courses” to those students, and for whom an approved assessment (“Measure B”) is available. Examples include science and social studies teachers.

- **Group III Educators** include anyone who does not meet the criteria for Group I or Group II educators. Examples include school nurses, psychologists, and guidance counselors.

The types and combinations of measures of student growth used for Component V are determined based upon an educator’s group classification. There are three types of measures that can be used depending upon an educator’s group:

- **Measure A** is based upon DCAS instructional scale scores for reading and/or mathematics in grades 3 through 10.8

---

8 See page 15 for more information on how Measure A targets are determined.
• **Measure B** is comprised of two types of assessment options: external and internal assessments. External assessments are recognized and identified by Delaware educator groups but generally created by outside vendors, while internal assessments were developed for and by groups of Delaware educators, coordinated by a DDOE facilitator. Both types of assessments are reviewed by an outside vendor for validity prior to approval by the state for use. An educator, with administrator approval, may choose to use either an external measure or an internal measure\(^9\). Educators/evaluators are not required to use both.

• **Measure C** “growth goals” were educator-developed, reviewed by an outside vendor and approved by the state. These growth goals are content-specific and based upon professional standards and position responsibilities. Therefore, these goals address student growth and professional growth outcomes based upon an educator’s respective field of expertise.

The measures an educator utilized and how these measures were calculated to determine final Component V ratings were based on an educator’s group and evaluator’s professional judgment:

<table>
<thead>
<tr>
<th>Measure A</th>
<th>Measure B</th>
<th>Measure C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group I</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Group II</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>Group III</td>
<td>50%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Although the types and weights of measures used varied by educator group, the overarching focus on student growth and the guiding principles Delaware Secretary of Education Mark Murphy emphasized in the design of Component V were consistent for all educators. As noted in his communications to all educators over the past year these principles are:

- **Fairness:** Student growth would be a component of all educators’ evaluations, not just mathematics and reading teachers. Each educator would have the opportunity to demonstrate results against more than one measure.

- **Transparency and easy to understand:** Policies and processes would minimize complexity while ensuring as much transparency as possible for all those involved.

- **Respectful of the profession:** The system was designed to provide many opportunities for discretion and professional judgment. Therefore, in setting Component V performance targets, all educators were able to: 1) select their own measures from approved lists; 2) submit their own measures for approval; and 3) use additional measures if desired. Group

\(^9\) See page 18 for more information about how internal Measure B Assessments were created
Educators (mathematics and reading) also had the opportunity to vet and verify the list of students included in their final Measure A rating. Administrators had the opportunity to approve all measures and goals selected by educators in their buildings as well as any amendments to rosters for a teacher’s Measure A rating.

In the next sections we turn our focus to the impact and implementation of the DPAS-II system through a review of the distribution of educator ratings and data from the annual DPAS-II evaluation survey. The next section focuses on the four appraisal components that are based upon classroom observations and qualitative assessments of professional practice. The distribution of Component I-IV ratings is presented in the following section as well as a comparison of these ratings with some of the student achievement results noted above.

**How “Measure A” Student Growth Targets were Created**

In 2010–2011, the Delaware Department of Education (DDOE) began collecting data to measure student academic growth based on the DCAS. This measure (Measure A) makes up one portion of Component V of DPAS-II and is based on the change in performance of students in grades 3 through 10 on DCAS reading and/or mathematics assessments from fall to spring.

The 2012–2013 school year was the first year in which student achievement results using a growth model was fully integrated into educator evaluations. Ratings for educators who instruct students in grades 3 through 10 reading and/or mathematics are based on the percentage of their students’ instructional scores that meet individual student growth targets.

Student growth targets were determined based on the average growth made by “similar” students over the first two years of DCAS (2010–2011 and 2011–2012). “Similar” students are defined as those who had:

- The same fall DCAS instructional score; and
- The same grade and subject; and
- The same Student With Disability (SWD), English Language Learner (ELL), SWD/ELL designation

For the 2011–2012 school year, students had two opportunities in the spring to take the DCAS. The higher of the two scores was utilized in developing the student growth model.

The statistical model that was used to develop expected scores (growth targets) in spring was created by pooling two years of data, i.e., combining all student records for both the 2010–2011 and 2011–2012 school years and creating a regression model for each grade and subject. Using the statistical method of regression analysis, a straight line was drawn that best fits the data. The line represents the average (or statistically expected) spring scores for all students in the state. Using this line, each fall score was associated with a predicted spring growth target score. Students who scored at the highest performance level (PL4) in the fall were expected to remain at this level in the spring. Any student who was at PL4 in the spring was considered to have met his or her growth target for the year.
B. DISTRIBUTION OF COMPONENT I-IV RATINGS

School leaders play a critical role in the DPAS-II system. Through observations of teaching practice and professional conversations with educators, school leaders provide the feedback and coaching that can lead to improved instruction. These interactions and observations allow school leaders to have a deep understanding of educators’ strengths and areas in need of professional growth. With this understanding school leaders also are able to connect educators to the right professional development opportunities and make decisions about personnel (promotion, retention, class assignments, etc.) in the best interest of students.

Recognizing the important role school leaders play in evaluating the effectiveness of teaching and helping teachers improve in their practice, the DPAS-II system places considerable emphasis on classroom observations in an educator’s summative rating. Each of the five components of the DPAS-II system amounts to 20 percent of an educator’s evaluation such that a large proportion of the evaluation is based on classroom observations. In many other states that are implementing evaluation systems with a student growth component, classroom observations comprise 50 percent of an educator’s summative rating—less than in Delaware.

However, the results below outlining the distribution of Components I-IV ratings in the 2012-13 school year suggest there is more work for school and teacher leadership in ensuring that these ratings and accompanying feedback are credible and meaningful.

<table>
<thead>
<tr>
<th>DPAS-II Components 1-4 Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Component I-IV Ratings</strong></td>
</tr>
<tr>
<td>Percent with 0 “Satisfactory” Ratings</td>
</tr>
<tr>
<td>Percent of with 1 “Satisfactory” Rating</td>
</tr>
<tr>
<td>Percent with 2 “Satisfactory” Ratings</td>
</tr>
<tr>
<td>Percent with 3 “Satisfactory” Ratings</td>
</tr>
<tr>
<td>Percent with 4 “Satisfactory” Ratings</td>
</tr>
</tbody>
</table>

Includes 5,075 educators with Components 1-4 data entered into ERS

In the table above, 99 percent of Delaware educators received “Satisfactory” ratings on each of the four components based upon classroom observations and professional responsibilities. In the 2011-12 school year, 99 percent of educators also received “Satisfactory” ratings on Components I-IV. There is very little variation when looking at the four components individually as well. Only 1 percent of educators received an “Unsatisfactory” rating in Component I (Planning and Preparation), Component II (Classroom Environment) and Component III (Instruction). Less
than half of a percent of educators received an “Unsatisfactory” rating on Component IV (Professional Responsibilities).

Overall, 75 percent of schools had all of their teachers receive “Satisfactory” ratings in all four of the classroom observation components. While it is entirely plausible that all of a given school’s teachers are “Satisfactory” in the components measured through classroom observation, such lack of differentiation statewide will continue to make it difficult to acknowledge and reward excellent teaching. Such a system might also complicate efforts to identify a teacher’s areas in need of improvement and allocate resources to support them. Additionally, the credibility of an appraisal process might be called into question where nearly all educators in an entire state are rated “Satisfactory” on each of the observational components, despite significant variation in the performance of students across the system. This is illustrated in the chart above. While the percent of students meeting their growth targets in DCAS mathematics in 2012-13 ranged from 49 percent in Seaford School District to 80 percent in Sussex Tech School District, there is almost no variation (from 96 percent in Laurel to 100 percent in Sussex Tech) in the percent of teachers with four “Satisfactory” ratings on Components I-IV.

The Component I-IV ratings distribution data also do not correspond with findings about teaching effectiveness in the recently released “Delaware Educator Diagnostic: An Analysis for the First State’s Workforce.” The study included an analysis of five years of Delaware teacher and student achievement data and found that teacher impact on student achievement varies

---

11 Only DCAS math data are included here given that English scores provide a similar result.  
12 The study was released in April 2013 and can be found at the following link: http://www.doe.k12.de.us/tleu_files/Delaware_Educator_Diagnostic.pdf
widely across the state. In some cases, the difference in student growth amounted to a difference in a year of learning.

This discordant information is likely to raise questions about why 99 percent of educators receive “Satisfactory” ratings each year on their classroom observation components in the DPAS-II system. Some would contend that it is the result of school leaders’ lack of capacity to distinguish or acknowledge differing quality in teaching practice. Studies have found that school leaders vary in their ability to effectively observe and evaluate teaching practice. Additionally, studies have found that professional relationships with staff can lead internal observers to provide higher ratings than an outside observer would provide. It is also possible that the nearly unanimous “Satisfactory” ratings are the result of the limitations of the DPAS-II rubric (i.e., design flaw). Evaluator assess teaching using a set of criteria under each of the four observation components along a scale of unsatisfactory, basic, proficient and distinguished. However, the options for their final rating on Components I-IV include only “Unsatisfactory” or “Satisfactory.” Thus, one educator may be “Distinguished” on each criterion within a component and another “Proficient” yet both will receive a final rating of “Satisfactory” on that component.

Whatever the reasons are for the lack of differentiation in Components I-IV in the DPAS-II system, it is clear there is room for continuous improvement. Through the state’s Development Coach Project the state has made major investments in building administrator capacity to conduct the DPAS-II process with success. Improving Component V may also serve to bolster the type of feedback educators receive through the DPAS-II process and ensure what school leaders communicate through appraisals of the classroom is ultimately aligned to educators’ impact on students. The next section focuses on the data collected from the state’s first year of implementation of a student growth measure through Component V.

---

**Investing in School Leadership and DPAS-II: “Development Coaches”**

Building the capacity of school leaders was a major strategy outlined in Delaware’s 2010 Race to the Top application. School leaders were provided with coaches to strengthen their skills as both managers and instructional leaders in their buildings. Understanding the design of the DPAS-II system was only as good as how well it was implemented in schools the DDOE sought a vendor to provide coaching for school leaders focused on the DPAS-II process. The DDOE ultimately contracted with the Delaware Academy of School Leadership (DASL) at the University of Delaware College of Education and Human Development to provide “development coaches” for school leaders. The goal of the coaching was to increase school leaders’ proficiency with the DPAS-II process and improve the accuracy of appraisals.

In the first year of the project (SY 2011-12), nine development coaches provided almost 5,000 hours of coaching services (12 hours per month per evaluator) to 140 evaluators (principals, assistant principals and district expert evaluators) at 63 schools in all 19 districts and six charter schools in Delaware. In Year 2 of the program (2012-13) coaches worked with school leaders to move beyond compliance to a deeper level of implementation to improve instruction. Participating evaluators continued to receive DPASII coaching services at a “maintenance” level of six hours per month, with the exception of new principals, principals in high-need schools, and principals/assistant principals who would benefit from more extensive coaching or who requested additional coaching. These participants received 12 hours of development coach services per month.

---

13 Rethinking Teacher Evaluation: Findings from the First year of the Excellence in Teaching Project in Chicago Public Schools (2010); The Reliability of Classroom Observations by School Personnel (2013);
C. DISTRIBUTION OF “COMPONENT V” RATINGS

In the previous section we presented DPAS-II data collected through classroom observations. A comparison of student achievement results at the district level with the distribution of classroom observation ratings demonstrated that the two indicators of performance are often not strongly correlated. The data from Component V in the next section provides further evidence of some of the misalignment between qualitative appraisals of professional practice and measures of student growth from last school year.

The revised Component V was implemented statewide for the first time in the 2012-13 school year. Every Delaware educator and administrator received a Component V rating based upon evidence of student growth and state code requires that educators and administrators receive Component V ratings annually. The system was designed such that all educators had multiple measures of student growth. Thus, some combination of Measure A, Measure B or Measure C options were used to determine an educator’s Component V score. Below we present and compare the distribution of ratings within each of these student growth measures.

Measure A Educator Ratings Distribution

In 2012-13 every “educator of record” for at least 10 students in reading and/or mathematics in grades 3 through 10 received a “Measure A” rating based on the statewide DCAS exam as one of his or her measures of student growth. “Measure A” ratings were determined based upon the percent of an educator’s students that met their individualized growth targets in reading and/or math. These targets were projections of expected growth based upon two years of data on similar students and a student’s score on the fall administration of the DCAS exam. In keeping with one of the core principles of the system outlined by Secretary Murphy, the evaluation system using Measure A is transparent to the extent that every Group 1 educator knew each of his or her students’ targets at the beginning of the year. To ensure fairness, another core principle of the system, all Group 1 educators were able to use the state’s online Roster Verification System (RVS) in Spring 2013 to confirm their students, with administrator approval, and their role as the “educator of record” for any student who they did/did not instruct for at least 85% of the course. After rosters were fully-verified, Group 1 educator ratings were determined using the following:

<table>
<thead>
<tr>
<th>Exceeds</th>
<th>Satisfactory</th>
<th>Unsatisfactory (discretion)</th>
<th>Unsatisfactory</th>
</tr>
</thead>
<tbody>
<tr>
<td>65% or more of a teacher’s students meet their DCAS student growth targets.</td>
<td>50-64% of a teacher’s students meet their DCAS student growth targets.</td>
<td>35-49% of a teacher’s students meet their DCAS student growth targets. Administrator has discretion to upgrade to a “Satisfactory” rating.</td>
<td>Less than 35% of a teacher’s students meet their DCAS student growth targets.</td>
</tr>
</tbody>
</table>

14 Even when an educator is in an “off-year” in their summative evaluation cycle, Component V ratings are still required.
Overall, 2,897 educators were classified as Group 1 and the distribution of Measure A ratings was as follows:

**2011-2012 Component V Measure A Ratings**
- Exceeds: 16%
- Satisfactory: 31%
- Unsatisfactory: 41%
- Unsatisfactory Discretion: 12%

**2012-2013 Component V Measure A Ratings**
- Exceeds: 12%
- Satisfactory: 54%
- Unsatisfactory: 5%
- Unsatisfactory with Discretion: 29%

In the 2012-2013 school year, 54 percent of Group 1 educators received an “Exceeds” rating for Measure A, 29 percent were rated “Satisfactory,” 12 percent “Unsatisfactory with discretion” and 5 percent “Unsatisfactory.” A greater share of teachers were rated “Exceeds” in Measure A in 2012-2013 (54 percent as compared to 41 percent) than during the pilot year in 2011-2012. Additionally, a smaller share of teachers was rated “Unsatisfactory” in Measure A in 2012-2013 (5 percent as compared to 12 percent in 2011-2012). It is important to note, however, teachers were not provided with their students’ growth targets in advance during the pilot year or the opportunity to verify their roster of students.

The distribution of Measure A ratings varied widely across the 19 Delaware school districts. The share of Group 1 educators rated “Exceeds” ranged from 96 percent to 22 percent. Of Delaware’s 19 school districts, 13 had at least 50 percent of their Group 1 educators receive “Exceeds” ratings on their Measure A. In all of Delaware’s school districts, the majority of Group 1 educators received either “Satisfactory” or “Exceeds” ratings on Measure A.

The percentage of Group 1 educators receiving “Satisfactory” ratings was even higher when administrator decisions—including the ability to upgrade a teacher’s “Unsatisfactory” rating—were considered. In cases in which only 35 percent to 49 percent of a teacher’s students met their growth targets, administrators had the option to upgrade “Unsatisfactory” Measure A ratings to “Satisfactory.” Across the state, 56 percent of educators whose Measure A rating were initially “Unsatisfactory with administrator discretion” were upgraded to “Satisfactory” ratings by their evaluators. After accounting for administrator discretion, 54 percent of Group 1 teachers received an “Exceeds” rating, 36 percent a “Satisfactory” rating, and 10 percent an “Unsatisfactory” rating.
The share of educators whose ratings were upgraded to “Satisfactory” varied across districts. In some districts only a third of educators rated “Unsatisfactory with administrator discretion” were upgraded to “Satisfactory” ratings. However, in other districts administrator discretion to upgrade a teacher’s rating was used in more than two-thirds of the cases. In one district, teachers were upgraded to “Satisfactory” 90 percent of the time. For Group 1 educators, however, Measure A ratings formed only one of the student growth measures that would factor into their final Component V rating.

**Measure B and C Ratings Distribution**

All Group 1 and 2 educators received Measure B ratings as a portion of their Component V rating. Group II educators teach grades and/or subjects not covered by the DCAS reading or math exam. Unlike the Measure A ratings, comprised of targets set for students taking the DCAS exam, an educator’s Measure B can include either external or internal assessments administered to the educator’s students. External assessments are recognized and identified by Delaware educator groups but generally created by outside vendors, while internal assessments were developed for and by groups of Delaware educators. An educator, with administrator approval, may choose to use either an external measure or an internal measure. Whereas Measure A included predefined targets, Measure B targets are set by an educator along with their school leader after administering a pre-assessment to students in the fall of the school year. A Measure B post-assessment is administered later in the school year and an educator receives an “Unsatisfactory,” “Satisfactory” or “Exceeds” rating based upon whether students met the growth targets agreed upon by an educator and school leader in the fall. Likewise, Measure C “growth goals” are set by an educator and approved by a school leader at the beginning of the school year with ratings determined based upon which goals are met by the end of the year. The
key difference is that Measure C includes content-specific growth goals based upon professional standards rather than an assessment.

Measure B ratings were entered into ERS for 6,794 Group 1 and 2 educators statewide. Measure C ratings were entered for 6,282 Group 2 and Group 3 educators. The distribution of ratings was as follows:

**2012-2013 Component V Measures and Overall Rating Distribution**

- **Measure A**: 10 Unsatisfactory, 36 Satisfactory, 54 Exceeds
- **Measure B**: 6 Unsatisfactory, 30 Satisfactory, 64 Exceeds
- **Measure C**: 1 Unsatisfactory, 30 Satisfactory, 69 Exceeds
- **Overall**: 1 Unsatisfactory, 46 Satisfactory, 53 Exceeds

Measure C had the greatest share of teachers rated “Exceeds” (69 percent) and nearly two-thirds of Group 1 and 2 educators received “Exceeds” ratings on their Measure B. While 10 percent of educators received “Unsatisfactory” ratings on Measure A, 6 percent and 1 percent received “Unsatisfactory” ratings on Measure B and C, respectively. Similar to Measure A ratings, the share of educators receiving “Exceeds” ratings on Measure B differed across Delaware districts. While in Smyrna School District 94 percent of Group 1 and 2 educators received an “Exceeds” rating.
rating on Measure B, 51 percent of educators in Caesar Rodney School District received the same rating. One in 10 educators received “Unsatisfactory” ratings in Measure B in Indian River, Brandywine and POLYTECH School Districts. In contrast, in six districts less than 5 percent of educators received “Unsatisfactory” ratings in Measure B.

Engaging Delaware Educators in the Design of “Internal Assessments”

At the beginning of the 2011–2012 school year, the DDOE engaged 600 Delaware educators in the construction of assessments from the ground up, for non-tested grades and subjects. This was not a new challenge: The DDOE had attempted to engage 400 educators in assessment identification and construction in January 2010 but the effort floundered, in part because of the absence of needed support from assessment specialists. Missteps in this first attempt, however, became an opportunity for learning and redesign.

In its new initiative, DDOE provided much more structure, carefully designed to provide teachers with preparatory professional development, engage the union leadership and define the assessment deliverables that the educators would be asked to create.

To get started, DDOE engaged a vendor to create a series of professional development modules for training that would total five days, spread over several months. The modules would train teachers to create standards-based assessments for those grades and courses whose students do not take the DCAS. The DDOE leadership team organized the teachers according to grade and subject into working groups of six to twelve—mathematics grades 9-12, for example. Facilitated by DDOE staff, the training began with instruction in the basics of assessment—distinctions between summative and formative assessment, for example—and moved on to the complexity of writing growth goals and test items. The modules covered the fundamentals of what constitutes a good rubric, explaining the eight strands in their model, which would pertain both to evaluating external measures and building internal measures.

Supported by professional development, the educator groups worked on the assessments for their assigned subject and grade range the remainder of the 2011–2012 school year. Each group was charged with producing a package of six deliverables that would constitute the complete set of materials necessary to provide data points for Component V of DPAS-II: the purpose statement, test blueprint, test specifications, pre-test/post-test, administrative guidelines and scoring guidelines. The package was designed like this to ensure that the assessments met the expectations for rigor and comparability.

As the new school year approached, in July 2012, DDOE neared completion of an analysis of every item on every assessment, reconvening teachers to create an additional item bank in anticipation that some items would fail to perform. By August 2012, 43 educator groups had produced assessment packages with growth goals specifying targets and evidence.

Source: “View from the States” a Brief on Non-Tested Grades and Subjects, Reform Support Network
For a list of internal assessments see: http://www2.ed.gov/programs/racetothetop/communities/tle2-de-measure-b.pdf
Although an educator’s Measure A and B were both used to evaluate student growth, the ratings a Group 1 educator received on the measures did not necessarily correspond. When an educator received an “Exceeds” rating on Measure A they also received an “Exceeds” rating on Measure B 76 percent of the time. However, the majority (55 percent) of Group 1 educators who received a “Satisfactory” rating on Measure A also received an “Exceeds” rating on their Measure B. Likewise, half of the educators rated “Unsatisfactory” on Measure A were rated “Exceeds” on Measure B and 37 percent were rated “Satisfactory.”

With the exception of Group 2 educators who received an “Unsatisfactory” rating on Measure B, the ratings received on Measure B and C were generally the same: 85 percent of educators who received an “Exceeds” rating on Measure B also were rated “Exceeds” on Measure C. Similarly, 61 percent of educators rated “Satisfactory” on Measure B received the same rating on their Measure C. Conversely, 70 percent of educators rated “Unsatisfactory” in Measure B received an “Exceeds” rating on Measure C.
In cases where a Group 1 or 2 educator was rated differently on their two measures (A/B or B/C), the DDOE provided guidance for what an educator’s overall Component V rating would be. If an educator received a combination of “Exceeds” and “Satisfactory” the final Component V rating would be “Satisfactory.” For a combination of “Exceeds” and “Unsatisfactory” ratings the final Component V rating would be “Satisfactory.” Finally, where the combination of ratings was “Satisfactory” and “Unsatisfactory,” an administrator had the discretion to upgrade to a “Satisfactory” rating. In 87 percent of these instances administrators opted to upgrade an educator’s rating to “Satisfactory.”

After administrator discretion was taken into consideration and Measure A, B, or C ratings were combined, of the nearly 9,000 educators with Component V ratings more than half (53 percent) received “Exceeds” ratings, 46 percent were rated “Satisfactory,” and 1 percent were rated “Unsatisfactory.” These Component V ratings mirrored Component I-IV ratings discussed earlier in that 99 percent of educators received the highest categories of ratings and only 1 percent were rated “Unsatisfactory.” When Component V ratings were compared for educators who also received Components I-IV ratings in 2012-13, the relationship between the classroom observation-based components and the student growth component was weak. Educators who received an “Exceeds” rating on Component V were most likely to receive “Satisfactory” ratings on Components I-IV. And 100 percent of educators rated “Exceeds” on Component V received 4 out of 4 “Satisfactory” ratings in Component I-IV. However, 98 percent of educators rated “Satisfactory” on Component V had the same ratings. While educators rated “Unsatisfactory” on Component V were significantly more likely to receive at least one “Unsatisfactory” rating in Components I-IV (27 percent received at least one “Unsatisfactory” rating), nearly three-quarters of these educators received “Satisfactory” ratings based upon classroom observations.

### DPAS-II Components 1-4 vs. Component V*

<table>
<thead>
<tr>
<th>Component V Rating</th>
<th>Less than 4/4 Satisfactory ratings in Comps I-IV</th>
<th>4/4 Satisfactory ratings in Comps I-IV</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unsatisfactory</td>
<td>27%</td>
<td>73%</td>
<td>100%</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>2%</td>
<td>98%</td>
<td>100%</td>
</tr>
<tr>
<td>Exceeds</td>
<td>0%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

*Includes 4,920 educators “on-cycle” with Components 1-5 data entered into ERS
D. 2012-2013 SUMMATIVE RATINGS

For all educators “on-cycle” in a given school year, Component V ratings and Components I-IV ratings are eventually combined using a rubric provided by the Delaware Department of Education to form those educators’ summative ratings. Unlike the formative feedback an educator receives at various junctures of the year—designed to identify strengths and weaknesses and avenues for professional growth—the summative rating is designed as an indicator of the educator’s effectiveness in his or her role. Whereas a formative appraisal is used by an evaluator to monitor and coach an educator’s professional growth, the summative appraisal is an evaluation that can be used to make personnel decisions (i.e. compensation, promotion, student assignments, dismissal, etc.). For example, in Delaware two consecutive “ineffective” summative ratings are used to establish a “pattern of ineffective teaching” which is grounds for dismissal proceedings. This does not suggest that summative ratings cannot be used formatively. Both educators and school leaders can use these ratings to guide their actions in future school years. However, the stakes are necessarily high with summative evaluations as the decisions have clear implications for the educators, students, and schools.

In 2012-13 5,275 educators received summative evaluations in Delaware. Of these educators, half (51%) received the newly-created “Highly-effective” rating and 48% received a rating of “effective.” Only 1% of educators received a rating of “Needs Improvement.”

Of the three educator groups, Group 3 educators were the most likely to receive “Highly-effective” ratings followed by Group 2 educators. Nearly two-thirds of Group 3 educators were rated “Highly-effective” as compared to only 52% of Group 2 educators and 38% of Group 1 educators.

This pattern—where Group 1 educators were least likely to receive “Highly-effective” summative ratings and Group 3 educators were most likely—was consistent across nearly all Delaware school districts. However, in some districts Group 3 educators were substantially more likely to be rated “Highly-effective” than

<table>
<thead>
<tr>
<th>Distribution of 2012-13 DPASII Teacher Summative Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>ineffective</td>
</tr>
<tr>
<td>0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Share of Educators in 2012-13 with &quot;Highly Effective&quot; or &quot;Effective&quot; Summative Ratings by Educator Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>Effective</td>
</tr>
<tr>
<td>Highly Effective</td>
</tr>
</tbody>
</table>
Group 1 educators. In Capital School District, 66% of Group 3 educators were rated “Highly-effective” as compared to only 19% of Group 1 educators (a 47 point difference). Brandywine, Colonial, Laurel, Red Clay, and Seaforth school districts also had differences around 40 percentage points for Group 1 versus Group 3 educators rated “Highly-effective.” By contrast, in Appoquinimink and Indian River school districts, the share of educators rated “Highly-effective” were similar for each educator group. Group 2 educators had the highest percentage of “Highly-effective” educators (49%) in Appoquinimink but 47% and 44% of Group 2 and Group 1 educators had the same rating, respectively. In Indian River school district, 53% of Group 3 educators were rated “Highly-effective” as compared to 52% of Group 2 educators and 45% of Group 1 educators.

These patterns likely reflect the fact that while DPAS-II improvements strengthened the focus on student growth for all educators across the system, many of the student growth goals set by educators and their evaluators were less rigorous than predicted DCAS growth targets for Group 1 educators. This was evident in the earlier comparison of Group 1 educators’ Measure A and B ratings. Half of the educators with less than 35% of their students meeting DCAS growth targets (“Unsatisfactory” on Measure A) were rated “Exceeds” on their Measure B assessment ratings (where educators collaboratively set targets and evaluators assess the results).

Overall, at the summative rating level, DPAS-II continues to present a more optimistic picture of educator effectiveness than the data available on the proficiency and college-readiness of Delaware students. At the same time, there is evidence of movement toward a system where students’ growth and educators’ performance appraisals are not contradictory. For example, unlike previous years, in 2012-13 there was some difference between summative ratings of educators in schools where students are not meeting their growth goals at high rates and those where students are making significant gains. In top quartile schools, where 68% to 95% of

---

15 The chart above suppresses data for any district or group with less than 10 educators in a given cell. Districts without a bar for a particular group did not have sufficient sample size in that cell. POLYTECH is not included because it only has more than 10 educators in Group 2 (where 59% were rated “Highly-effective”).
students met their growth targets in DCAS math in 2012-13, 100% of Group 1 educators were rated “Effective” or better. In comparison, in schools where less than 51% of students were meeting growth targets, 93% of Group 1 educators were rated “Effective” or better. Whereas 61% of Group 1 educators were rated “Highly-effective” in schools where students were meeting growth expectations at high levels, only 6% were rated “Highly-effective” in schools in the bottom quartile. While the 7 percentage point difference in educator summative ratings in schools in the top quartile of DCAS math growth and in the bottom quartile is not substantial it represents a change from last year’s distribution of ratings. In these same schools during the 2011-12 school year, close to 99% of Group 1 educators were rated “Effective” across each of the quartiles of student growth performance.

When we include the summative ratings of all educators (Group 1, 2, and 3 vs. Group 1 only) the already minor differences in the distribution of educator ratings across schools with different levels of student growth are diminished. While only 35% of educators are rated “Highly-effective” in schools in the bottom quartile of student growth as compared to 61% in schools in the top quartile of student growth, close to 99% of educators are rated “Effective” or better across schools with very different levels of student growth.

In the previous sections we have reviewed the changes to the system and the data on the distribution of evaluation ratings across the state and districts. In the next section we consider how the DPAS-II system and changes to the system were experienced by educators and school leaders across the state. We outline the various avenues that were available for educator involvement and feedback on the system—namely the DPAS-II Annual Process Evaluation Survey—and the areas of improvement recommended by educators.
E. IMPROVING DPAS-II: INCORPORATING FEEDBACK FROM EDUCATORS

The data presented in the previous sections demonstrates there is a continued need for improvement of the DPAS-II system. The quantitative information provided above, while revealing, tells only a part of the story. The evaluation system—and any changes made to it—are experienced first by the educators and school leaders implementing the system. It is for this reason that the feedback and involvement of educators is essential for improving DPAS-II and assessing its strengths and weaknesses.

Since the DPAS-II system’s inception in the 1980s educators, school leaders, and other key stakeholders have played a role in making it responsive to the current needs of the educator workforce. The 2012-13 school year was not different. There were a number of opportunities for educators, school leaders, district leaders, and other stakeholders to shape the design and implementation of the DPAS:

- **DPAS-II Review Committee**: The Department of Education is required by regulation to organize a committee for the purpose of developing and revising the DPAS-II guides and considering additional changes to the system. This committee is comprised of DSEA, DASA, superintendents, principals, district leaders, teachers, and representatives from higher education. The committee advises the DDOE on matters related to the DPAS-II evaluation system and over the past three years has made recommendations related to (but not limited to) regulatory changes, DPAS-II guide revisions, DPAS-II implementation, and training/professional development.

- **Delaware Principals Advisory Group (D-PAG)**: All principals and assistant principals were invited to join the monthly meetings of the D-PAG to advise the department on the DPAS-II system and other matters pertaining to teacher and leader effectiveness.

- **Delaware Technical Advisory Group (DE-TAG)**: The Delaware Technical Advisory Group (DE-TAG) met throughout the 2011-12 and 2012-13 school years and was comprised of local and national experts with knowledge and skill in assessment development, accountability, and teacher evaluation systems for the purposes of supporting the State with the ongoing development and refinement of DPAS-II and Component V.

- **Teacher Advisory Councils (TACs)/ Teacher of the Year Committee**: TACs were launched in the fall of 2012 as an opportunity for groups of teachers to meet regularly with the Delaware Department of Education to weigh in on the most pressing educational issues for students and schools. All educators were eligible to participate. The Teacher of the Year Committee, comprised of Teacher of the Year representatives from each district, also convened regularly to advise the Delaware Department of Education.
• **Mid-Year Conversations with Secretary Murphy:** All educators were invited to participate in “mid-year conversations” with Secretary Murphy and leaders of the DDOE to reflect on the DPAS-II system during a meeting held in each county.

• **Ongoing meetings with DSEA and DASA:** Secretary Murphy’s monthly meetings with Chief School Officers and their district leadership teams; quarterly meetings with district personnel directors were also regularly used to collect feedback about the DPAS-II system.

In addition to these opportunities, the annual DPAS-II process evaluation study provided one of the most wide-reaching efforts to collect feedback from Delaware educators and school leaders on the DPAS-II system. The annual process evaluation has been conducted by an external vendor for the last six years and includes a survey, focus groups, and interviews. In 2012-13 46 percent of teachers, 47 percent of specialists and 44 percent of administrators responded to the survey. 444 interviews were conducted with teachers, specialists, and administrators across the state. In addition, 45 educators and administrators participated in one six focus groups held in various parts of the state. While the full report is available online\(^{16}\) below we provide select findings from the study that identify some of the strengths and areas of improvement needed for the DPAS-II system:

• **The process is viewed as useful and impactful:** 85 percent of teacher respondents agreed that the oral feedback they receive through the DPAS-II process is “useful and applicable” and 80 percent feel the same way about the written feedback they receive. 84 percent find their DPAS-II conferences to be valuable in general and 64 percent believe the overall system has “some” or “major” impact on improving their teaching. 64 percent also believe the student improvement component (Component V) has “some” or “major” impact on improving their teaching. This percentage is in the 70s for Components I-III (Planning and Preparation, Classroom Environment, and Instruction).

• **The DPAS-II system is viewed as one of the top efforts to improve student achievement:** 47 percent of teacher respondents and 45 percent of specialists believe DPAS-II is “one of the top three efforts to improve student achievement” or is the “most significant driver of student achievement gains.” For the teacher respondents who believe the “DPAS-II evaluation system is being implemented appropriately in their school” (75 percent of the total respondents), 52 percent believe DPAS-II is one of the top three efforts or the most significant driver of student achievement gains. Among teachers who do not believe the system is being implemented appropriately in their school, only 32% view DPAS-II as at least a top three driver of student achievement gains.

---

\(^{16}\) Reports for the past six DPAS-II process evaluation studies at the following link:

[http://www.doe.k12.de.us/csa/dpasii/default.shtml](http://www.doe.k12.de.us/csa/dpasii/default.shtml)
The DPAS-II system was given a grade of “C” by the majority of teachers, specialists, and administrators for the first time: While the system is viewed as a driver of student achievement, and the process as one that impacts teaching, educators expressed much discontent with the implementation of the reformed system via the 2012-13 process evaluation. Many of the problems expressed in interviews and focus groups centered on problems in the rollout and implementation of the revised Component V. This included concerns about the consistency and timeliness of communication, ease of use of DPAS-II related technology, and the need for more training and support in the process. Only 42 percent of teacher respondents felt that DPAS-II related communications from the DDOE have been clear and only 45 percent agreed Performance Plus (one of the systems used for Component V) was “easy to utilize for the DPAS-II process.” A little more than half of teacher respondents felt that training in the process was adequate and only 47 percent of administrators felt that training was adequate.

Teachers who believe educators have been adequately involved in improving the DPAS-II system were more likely to give the system a higher grade. 62 percent of the educators who agreed with the statement “I believe all educators have been adequately involved in improving the DPAS-II system” gave the system an A or B grade. In contrast, only 17% of the teachers who disagreed with the same statement gave the system an A or a B grade.

The majority of educators and administrators believe the DPAS-II system needs improvement: While 81 percent of teacher respondents believe the evaluation process overall is of “high” or “medium” quality, only 25 percent of teacher respondents agreed that “the current DPAS-II evaluation system should be continued in its current form.” Only 21 percent of specialists and 18 percent of administrators believe the system should be continued in its current form.
Through the DPAS-II process evaluation (in 2012-13 and in previous years) and feedback collected through the various forums for educator engagement described above over several years, some common themes have emerged as areas most worthy of ongoing attention:

- **Allow for Focus:** According to educators, the system is time-consuming and keeps the evaluator from honing-in on key feedback. There are two reasons as to why this might be: (1) the burden of paperwork associated with the system and (2) the lack of flexibility in the use of announced and unannounced observations. 71 percent of administrator respondents, 58 percent of teacher respondents and 64 percent of specialist respondents disagreed with the statement “the time it takes to complete the paperwork is reasonable.”

- **Allow for Differentiation:** Evaluators should spend the most time with those who warrant the most time. The evaluation system should be differentiated based on an educator’s years of experience or role. 80 percent of teacher and 87 percent of specialist respondents agree it should be differentiated based on an educator’s role.

- **Allow for Peer Observation:** Some educators have named that a “second set of eyes” affirms/pushes an evaluator’s judgment and feedback. 78 percent of educators feel that peer observations are either “valuable” or would provide “some value.”

- **Place emphasis on Educator Development:** The authenticity/usefulness of the process is often questioned when the evaluation becomes more of an exercise in compliance.

Addressing these areas, in addition to the technical and implementation complications that were prevalent in the 2012-13 school year, is crucial for creating a system that is more responsive to educators and impactful for teaching practice.

At the same time, in order to be meaningful the system and policy-level changes need to be accompanied by increased capacity at the school-level to implement the system with fidelity. The results of the DPAS-II survey indicate that educators who believe the system is not being implemented appropriately in their work location are much less likely to perceive the system to be fair and equitable. While 60 percent of the educators who “believe the DPAS-II system is being implemented appropriately in their work location” also believe the system to be “fair and equitable,” only 19 percent of those who do not believe the system is implemented appropriately in their schools see it as fair. Where the system is not believed to be implemented appropriately, educators also see less value in the feedback received and report the process has less impact on improving their teaching. This underscores the importance of effectively monitoring implementation of the system and providing sufficient training and support. In the next section we consider both potential changes and those changes taking effect in the 2013-14 school year to improve the DPAS-II system overall.
State Monitoring of LEA DPAS-II Implementation

Delaware state regulations require the DDOE to engage in ongoing monitoring of DPAS-II implementation at the local level. As the DPAS-II system underwent significant revisions in the 2012-13 school year the DDOE increased its supports (Development Coaches, trainings, DPAS-II hotline, etc.) for district and school staff while also expanding its monitoring efforts. The Department’s Teacher and Leader Effectiveness Unit led the cross-departmental teams visiting Delaware districts and schools to assess the fidelity of the DPAS-II process and the quality of documentation. Overall the DDOE conducted monitoring visits in 32 schools and each of the 19 school districts in 2012-13. The state’s monitoring activities included:

1. A review of data completion and quality through the state’s Evaluation Reporting System (ERS)
2. A review of quality of documentation and overall implementation of the DPAS-II process.

In order to monitor data completion and quality, the DDOE reviewed reports generated monthly and shared (through ERS) with districts to capture evaluators’ progress with the evaluation process at various points of the school year. The DDOE also reviewed ERS to identify schools where required data were not entered into the system or were entered with errors (e.g. evaluation ratings did not match the DPAS-II rubric). The DDOE followed up with schools and districts with missing data or reporting errors.

Quality of documentation and overall implementation of the DPAS-II process was monitored through onsite visits to districts and schools to assess processes, procedures, and forms. Visits to districts included a district presentation on their approach to DPAS-II implementation from an operational, support, and compliance perspective. Formative feedback documentation as well as Component V documentation was also reviewed during the visits to schools and districts. Schools and districts received feedback/ratings indicating the degree to which they were on-track with implementation.
F. IMPROVING DPAS-II: 2013-14 & BEYOND

This report demonstrates the need for continuing improvement of the evaluation system and, perhaps most importantly, building capacity in Delaware’s schools and districts to implement the system well. Although the system provides some meaningful data about which educators are excelling, and which educators are struggling to help their students grow, that data is not translating into meaningfully differentiated results in Delaware’s overall rating system. Educators have expressed their views about how the system can be improved through advisory committees, town-hall meetings, focus groups, and surveys. In response to all of these observations and suggestions—from state, district, and school leaders and educators across Delaware—the DDOE has taken the following actions to improve the system in the 2013-14 school year:

- **Continuation of Development Coaches:** Nearly 70 schools signed up to use a Development Coach to assist with educator evaluation this year. Each school receives coaching every week focused on helping school leaders: improve their observations of teachers, strengthen the process through which they set expectations for their educators, and identify opportunities for teacher and student improvement.

- **Specific Training in Student Growth Goal-setting:** Over 600 evaluators participated in the DDOE’s DPAS-II “Base Camp” trainings in the summer of 2013. The trainings involved evaluators in a review of actual goals/targets that were collected throughout the year from each educator group and reflection about ways to improve them. The trainings also discussed best practices for goal/target setting and the use of discretion.

- **More Flexibility in Classroom Observations:** The Delaware State Board of Education passed amendments to Regulations 106A/107A, which took effect in July 2013. The amendments allow for more flexibility in the use of announced versus unannounced observations. The revised regulations also allow State or Local Education Agencies (SEAs or LEAs) to credential educators’ peers to serve as classroom observers.

- **Strengthening Alternative Assessments:** Recognizing that a significant portion of educators are evaluated using Measure B assessments, DOE has refined existing assessments, and invited districts and schools to submit their best assessments for state approval. The DDOE will keep the window open for the submission of assessments to be used in the 2014-15 school year.

- **Alternative Educator Evaluation System Application:** Recognizing that each district and school brings its own unique strengths to the development of an educator evaluation process, the DDOE has developed a process by which LEAs can create their own educator evaluation system in an effort to increase educator support and accountability and, most importantly, student achievement.
These efforts are based upon feedback from stakeholders and lessons learned from previous years, and are designed to improve the system in a manner that leads to learning and professional growth for all educators and leaders in Delaware schools, resulting in better educational outcomes for students. Delaware has invested heavily in this system because it is essential to be able to identify the best educators, and those who are struggling to serve our students, so that decision-makers can replicate best practices and support those who need extra help.
G. CLOSER LOOK: CHARTER SCHOOL DPAS-II EVALUATION DISTRIBUTION

Evaluation data were available for 352 educators in 17 Delaware charter schools in 2012-13. Of these 352 educators 248 received a summative rating during the 2012-13 school year. The distribution of summative ratings was as follows:

32 percent of educators receiving summative evaluations in charter schools were rated “Highly-effective” as compared to 51 percent in traditional public schools. While a smaller share were rated “highly-effective” in Delaware charter schools the share of educators rated “Effective” or better is closer to the figure for traditional public schools. 93 percent of charter school educators received a summative rating of “Effective” or better as compared to 99 percent in traditional public schools.

The distribution of Component I-IV ratings was similar in charter schools and traditional public schools. 96 percent of charter school teachers received a “Satisfactory” rating on each of the four qualitative appraisal components in the DPAS-II system. The distribution of Component V ratings differed, however. While 53 percent of educators in traditional public schools received an “Exceeds” rating 40 percent of charter school educators received an “Exceeds” rating. Overall, 96 percent of educators in charter schools received a Component V rating of ‘Satisfactory’ or better.

The largest share of educators received ‘Unsatisfactory’ ratings in charter schools on Measure A. Before “administrator discretion” was taken into account, 17 percent were rated “Unsatisfactory with administrator discretion,” 13 percent “Unsatisfactory,” 24 percent “Satisfactory,” and 47 percent “Exceeds.” After discretion is accounted for 20 percent receive “Unsatisfactory” ratings in Measure A. In contrast, 10 percent received “Unsatisfactory” ratings on Measure B and 2 percent on Measure C. Educators in charter schools were also less likely to receive “Exceeds” ratings than their counterparts in traditional public schools on Measure B and C. 49 percent received an “Exceeds” rating on Measure C (compared to 70 percent in traditional public schools) and 43 percent on Measure B (compared to 64 percent).
H. CLOSER LOOK: STATEWIDE STUDENT ACHIEVEMENT IN 2012-13

In the 2012-2013 school year a little more than 77,000 Delaware students in grades 3 through 10 took the Delaware Comprehensive Assessment System (DCAS) exams. Substantial gains in rates of proficiency were seen at the end of the 2011-2012 school year, and there were notable successes to build upon: 72 percent of students scored proficient as compared to 61 percent in 2011. Math proficiency jumped significantly higher in 2012, when 73 percent of students were proficient up from 63 percent in 2011. This declined slightly (by three percentage points) to 70 percent in 2012-13. The percentage of students who scored “Advanced” in reading increased, and the 2011-12 gains in the percentage of students who scored advanced in math were sustained.

In addition to rates of proficiency, students were measured, for the first time in 2012-2013, against individualized student growth targets. These growth targets provide a progress benchmark for all students — including students who start the year significantly below the proficiency mark and students who begin the year as proficient or better—by setting an expectation for growth based upon the results of similar students in previous years. The targets —used for Component V “Measure A” ratings—were piloted in 2011-12. (Results were shared with teachers for informational purposes). The 2012-13 academic year was the first year where student growth was reported publicly and the results factored into Component V ratings. Group 1 Educators (mathematics and reading) were provided growth targets for each student at the beginning of the school year using a growth table provided by the Delaware Department of Education.

More than two-thirds (67.8 percent) of students met their growth targets in reading and 62 percent met student targets in mathematics in 2012-2013. This represented a 3 percentage point gain in reading and 2 point gain in mathematics compared to 2011-2012 (the pilot year). A little more than 4 out of 10 Delaware students met their mathematics growth targets in both the 2011-12 and 2012-13 school years; 20 percent of students met the target after missing the growth target last year and 18 percent of students missed their growth targets in both years. Students who met their growth targets in 2011-12 were more likely to meet their targets in the following year. While 69 percent of students who met their DCAS math target in 2011-12 also met their math target in 2012-13, only half of the students who did not meet their target in the 2011-12 met their math target in 2012-13.