



AMERICAN UNIVERSITY
WASHINGTON, DC

SCHOOL OF EDUCATION

Perspectives on DCPS IMPACT Teacher Evaluation System:
Findings from Teachers and School Leaders

Corbin M. Campbell, Ph.D.
Carolyn Parker, Ph.D.
Robert Shand, Ph.D.
Adelaide Kelly-Massoud, Ed.D.
Toks Fashola, Ph.D.
Justin Blanc

Acknowledgements: Renee Metellus was a Graduate Research Assistant during data collection. Gabrielle Levy and Veer Sawney participated as student research team members during report development. These contributions were very significant to this study.

Executive Summary

Background

In spring of 2020, DCPS contracted with American University School of Education (SOE) to provide independent research on teacher and school leader perspectives of the DCPS teacher evaluation system: IMPACT. As independent experts on teacher evaluation, SOE did not receive funding from DCPS for the project. The AU research on IMPACT is one component of a larger review that DCPS commenced in the 10th year of the IMPACT implementation. The DCPS IMPACT Review continues to gather feedback from many stakeholder groups, including teachers, school leaders, instructional superintendents, and national experts. DCPS asked the AU research team to focus on collecting data from teachers and school leaders on their perceptions of IMPACT, contextualized by a broad swath of IMPACT data and other related data given to the AU research team. The AU research team would also provide expertise on the existing literature on teacher evaluation. In addition, members of the AU research team as well as the AU School of Education Dean participated on the national advisory group of experts for the DCPS IMPACT Review.

The AU research team collaborated with DCPS to develop the research questions that guided the AU research on IMPACT. The research questions considered the primary purposes of IMPACT as a teacher evaluation system. According to DCPS, “DCPS’s vision is that every student feels loved, challenged, and prepared to positively influence society and thrive in life.” DCPS shared with the AU team that IMPACT supports this vision through three mechanisms: “Recognize and retain our very best,” “Support growth,” and “Transition out low performers.” Based on this information about IMPACT as well as the goals of the IMPACT Review, the following research questions (Appendix A) guided the AU study of IMPACT:

1. How do DCPS teachers and school leaders perceive IMPACT as a feedback, evaluation, accountability, and incentive system? What do they perceive could be improved and how?
2. How does IMPACT facilitate DCPS teacher improvements? How can IMPACT be altered to better support teacher improvement?
3. To what extent can the validity and fairness of IMPACT be improved, and if so how?
4. To what extent does IMPACT relate to the pipeline to and through DCPS for teachers?

Methodology

Data for this report were gathered through teacher interviews (46 interviews), school leader focus groups (4 focus groups), school leader survey (63 respondents), and quantitative data given to the AU team by DCPS. Quantitative data given to the AU team included data from the DCPS Insight survey¹, IMPACT results, and other IMPACT related data. DCPS Details on the methodology, including data collection procedures, sampling procedures and participant characteristics can be found in the main report. All sampling, data collection, and data analyses contained in this report were conducted independently by the AU Research Team. Participants in the study were assured confidentiality and given an opportunity to review this final report to

¹ TNTF Instructional Culture Insight survey (TNTF 2021) (<https://tntp.org/teacher-talent-toolbox/insight-survey>)

ensure that all quotes were not identifiable. The AU teacher sample represents a broad swath of DCPS teachers across a diverse range of identities and experiences (see introduction for more information). However, the sample does have limitations, which are articulated in the full report. Beyond sample, an additional important context to the results is the timeframe. Spring 2020 was not a typical time period for DCPS teachers given the timing of the COVID-19 pandemic.

Summary of Findings

Summary of Findings for Research Question 1:

Overall perceptions of IMPACT were more negative than positive, but with a great deal of variation in perspectives across teachers as well as school leaders. Many teachers and school leaders perceived that IMPACT played an important role in expectation setting for DCPS. However, many teachers and school leaders perceived that IMPACT created an unhealthy environment of distrust, fear, and competitiveness in schools that trickles down into the classroom. These themes held true across teacher effectiveness ratings, teacher race, as well as Ward.

Summary of Findings for Research Question 2:

Findings from teacher interviews, teacher Insight survey, school leader focus groups, and school leader survey demonstrated that across stakeholders, and across teacher identities, ward, and effectiveness ratings, most participants saw a need for improvement in the alignment between IMPACT and professional growth. However, there was some variation among participations, and a smaller group of participants was able to make strong connections between IMPACT and professional growth. Both teacher and school leader participants shared that one significant threat to the potential for growth due to IMPACT is the high-stakes nature of IMPACT. *Both groups desired more formative ties to authentic professional development, including improved feedback cycles and coaching.*

Summary of Findings for Research Question 3:

Broadly, teachers and school leaders were moderately positive about whether they perceive IMPACT as valid in surveys. However, subjectivity, bias, and gaming the system were all cited as threats to validity in teacher interviews, in school leader focus groups, and in the school leader survey. Many teacher and school leader participants recommended the use of external, more objective, and subject-matter knowledgeable observers to reduce subjectivity and favoritism. Subjectivity and gaming were consistent themes for teachers across race, ward, and effectiveness rating as well as across school leaders. In terms of fairness, participants were somewhat less favorable. Some (particularly White teachers) felt that the subjectivity and favoritism were unfair. Other teacher participations (particularly Black teachers) articulated concerns for equity, particularly for under-resourced schools.

Summary of Findings for Research Question 4:

Overall, teacher and school leader participants discussed two roles that IMPACT has played for teacher recruitment, retention and attrition patterns. First, several spoke about how IMPACT has the ability to transition out low performing teachers. School leaders saw this as a particular asset to their work. Second, participants discussed that IMPACT has the possibility of retaining teachers through incentives, such as bonuses and the LIFT ladder. However, other teachers

perceived that the high-stakes and anxiety producing environment may cause them (or others) to leave DCPS.

Teacher and school leader participants shared several ideas for improvements of IMPACT that address the above areas of concern, including:

1. Improving the way that feedback is given, for example, by having pre-observation conferences, setting specific goals, and focusing feedback on specific goals tied to subject-matter.
2. Multiple low stakes observations that are more closely tied to coaching.
3. Reducing the connection between IMPACT and high-stakes monetary incentives;
4. Considering external evaluators, rotating evaluators, evaluators with subject-matter expertise, and multiple evaluators;
5. Increasing formative professional development opportunities;
6. Greater depth in the ways teachers learn about IMPACT (e.g., orientation);
7. Improved training of administrators on implementing IMPACT, including more norming
8. Greater trust and autonomy for teachers and more inclusion of teachers' voices in the evaluation process;
9. Including trauma-informed and culturally relevant teaching in IMPACT measures;
10. Giving more resources to teachers at under-resourced schools;
11. Improving the observation process by providing more flexibility and more transparency;
12. Measuring closing the equity gap;
13. Many participants, especially school leaders, recommended eliminating IVA as a means to reducing inequalities.²
14. Changing the name of IMPACT and redesigning to get away from historical inequities.

In terms of specific components of IMPACT, there was, largely, agreement by both teachers and school leaders that the Essential Practices were the most growth-oriented component of IMPACT. Likewise, there was by and large agreement by both teacher and school leader participants that they were concerned that IVA was biased and unfair. There were significant concerns with the validity of the student survey and TAS; and concerns about subjectivity and inconsistent implementation with CSC. Both teachers and (particularly) school leaders seemed to find CP to be a useful standard for performance, but not for growth. Participants across both stakeholders described appreciation for a multiple measures approach to evaluation.

In sum, perceptions of IMPACT matter. They matter to school culture and climate, to motivation and buy-in, they are related to retention and job satisfaction, and they matter because they exemplify the lived experiences of teachers and school leaders in DCPS. Teachers and school leaders expressed appreciation for their voices being heard and considered in the upcoming evolutions of IMPACT.

² Per DCPS analyses, the IVA has more equal outcomes by Ward and Title I than other components, whereas there are more disparate outcomes in the EP component.

Table of Contents

1	Introduction	2
2	Literature Review	3
3	Methodology	6
3.1	Teacher Interviews.....	7
3.1.1	Teacher Interview Sample.	7
3.2	School Leader Focus Groups and Survey.....	8
3.3	Data from DCPS used in the Report	8
4	Findings	9
4.1	Teacher Experience of IMPACT.....	9
4.1.1	RQ 1 General Perceptions of IMPACT.....	10
4.1.2	RQ 1 Theme: Understanding How IMPACT Works.....	11
4.1.3	RQ 1 Theme: Variation in Perspectives on Student Learning.....	14
4.1.4	RQ1 Theme: School Culture/Climate of Distrust, Anxiety, and Competitiveness..	16
4.1.5	Summary of Research Question 1 Findings.....	21
4.2	Professional Growth.....	21
4.2.1	RQ2 Quantitative Findings Pertaining to Professional Growth	21
4.2.2	RQ2 Theme: Alignment with Professional Development	22
4.2.3	RQ2 Theme: Consequences of High-Stakes Incentives for Teacher Growth & Innovation	28
4.2.4	RQ2 Summary of Findings.....	31
4.3	Validity and Fairness.....	31
4.3.1	Broad Perspectives on IMPACT Validity and Fairness.....	32
4.3.2	RQ3 Validity Theme: Subjectivity/Inconsistency of the Evaluation System.....	34
4.3.3	RQ3 Validity Theme: Manipulation/Gaming the System.....	36
4.3.4	RQ3 Fairness Theme: Perceived History as An Inequitable Evaluation System...40	
4.3.5	RQ3 Fairness Theme: Questioning of Equity Outcomes	41
4.3.6	RQ3 Summary of Findings.....	46
4.4	Labor Market.....	46
4.4.1	RQ4 Theme: Effects on Teacher Retention/Attrition.....	46
4.4.2	RQ4 Summary of Findings.....	48
4.5	Perspectives on Specific IMPACT Components.....	48

4.5.1	Essential Practices.....	49
4.5.2	Individual Value-Added (IVA).....	52
4.5.3	Teacher Assessed Student Achievement (TAS)	56
4.5.4	Commitment to School Community (CSC)	58
4.5.5	Student Survey	60
4.5.6	Core Professionalism (CP).....	63
5	Limitations.....	64
6	Conclusion	64

1 Introduction

In spring of 2020, DCPS contracted with American University School of Education (AU) to provide independent research on teacher and school leader perspectives of the DCPS teacher evaluation system: IMPACT. As independent experts on teacher evaluation, SOE did not receive funding from DCPS for the project. The AU research on IMPACT is one component of a larger review that DCPS conducted in the 10th year of the IMPACT implementation. According to DCPS leadership, the goals of the IMPACT Review were threefold:

- “Identify what’s working well and what might be improved in IMPACT’s 10th year of implementation, informed by feedback from DCPS teachers and school leaders;”
- “Make changes to IMPACT policy, processes, and supports will lead to improved outcomes for students;” and
- “Further ensure our teachers feel supported and valued; increase teacher satisfaction with their evaluation experience.”

The DCPS IMPACT Review continues to gather feedback from many stakeholder groups, including teachers, school leaders, instructional superintendents, and national experts. DCPS asked the AU research team to focus on collecting data from teachers and school leaders on their perceptions of IMPACT, contextualized by a broad swath of IMPACT data and other related data given to the AU research team. The AU research team would also provide expertise on the existing literature on teacher evaluation. In addition, members of the AU research team as well as the AU School of Education Dean participated on the national advisory group of experts for the DCPS IMPACT Review.

The AU research team collaborated with DCPS to develop the research questions that guided the AU research on IMPACT. The research questions considered the primary purposes of IMPACT as a teacher evaluation system. According to DCPS, “DCPS’s vision is that every student feels loved, challenged, and prepared to positively influence society and thrive in life.” DCPS shared with the AU team that IMPACT supports this vision through three mechanisms: “Recognize and retain our very best,” “Support growth,” and “Transition out low performers.” Based on this information about IMPACT as well as the goals of the IMPACT Review, the following research questions (Appendix A) guided the AU study of IMPACT:

1. How do DCPS teachers and school leaders perceive IMPACT as a feedback, evaluation, accountability, and incentive system? What do they perceive could be improved and how?
2. How does IMPACT facilitate DCPS teacher improvements? How can IMPACT be altered to better support teacher improvement?
3. To what extent can the validity and fairness of IMPACT be improved, and if so how?
4. To what extent does IMPACT relate to the pipeline to and through DCPS for teachers?

DCPS and the AU Research Team met regularly throughout the study period to ensure that the study was continually responsive to DCPS needs. However, all sampling, data collection, and data analyses contained in this report were conducted independently by the AU Research Team.

Participants in the study were assured confidentiality and given an opportunity to review this final report to ensure that all quotes were not identifiable. Additionally, the purpose of the AU research was to provide an independent analysis of the perspectives of DCPS teachers and school leaders on IMPACT as a teacher evaluation system. Any recommendations in this report are taken from teachers and school leaders as participants in the study.

2 Literature Review

American University's study of IMPACT contributes new data on teacher and school leader perspectives within a high stakes evaluation system to a broad base of existing research. Recent literature has explored numerous elements that are salient in the current study, including the relationship between teacher evaluation scores and student outcomes, reliability and potential bias of administrator evaluations of teachers, the effectiveness of merit pay, the relationship between evaluation scores and the characteristics of students and individual teachers, and the value of evaluation systems as an instrument for professional growth.

Jackson, Rockoff, and Staiger (2014) reviewed recent findings around two relevant topics: the effects of individual teachers on student outcomes and how the different features of evaluation systems might be leveraged to improve student outcomes. They concluded that teachers are by no means interchangeable with respect to student outcomes. There was a wide variation in teacher effectiveness, but they found that this variation is largely unpredictable on the basis of observable characteristics. The authors highlighted studies (Chetty et al. 2014; Jackson 2013) that support the idea that students assigned to high value-added teachers benefit from a raised human capital, as these students were more likely to have measurably higher educational and socioeconomic outcomes. Yet, they described that these effects on human capital and long-term outcomes may be due to teacher effectiveness factors other than those captured by test scores. Notably, Cohen and Goldhaber (2016) pointed out that the connection between observation ratings and long-term student outcomes has not been established as it has been with value-added ratings.

The analysis of Jackson and colleagues showed that teacher effectiveness tends to improve as teachers gain years of experience, and teachers may improve most rapidly if they "(a) have similar assignments from year to year so that they may gain mastery of a particular curriculum, (b) have opportunities to interact with high-quality colleagues, and (c) have opportunities for professional development" (Jackson et al. 2014, p.813). While evaluation systems are potentially effective tools for teacher improvement, the authors noted, most were viewed as being poorly designed and poorly implemented. The study described that many were viewed negatively because almost all teachers are given top ratings or because of other systemic problems like "vague district standards, poor evaluation instruments, overly restrictive collective bargaining agreements, and a lack of time [as well as] the absence of high-quality professional development for evaluators, a school culture that discourages critical feedback and negative evaluation ratings" (Donaldson 2009, p. 2, as cited in Jackson et al. 2014). Nevertheless, they pointed to promising examples of well-implemented evaluation systems leading to sustained performance

gains and real increases in teacher skills, as seen in the Chicago and Cincinnati Public Schools (cf. Steinberg & Sartain, 2015; Taylor & Tyler 2012).

On the question of financial incentives, Jackson and colleagues analyzed evidence suggesting that teacher performance pay can often improve student outcomes, especially the outcomes on which rewards were based (Neal 2012). Yet they also discussed the common problems of merit pay, like gaming the system and diverting valuable teaching resources to a limited set of outcomes at the expense of other valuable dimensions of the teaching profession. Significant research suggests that the effects of value-added measures and teacher performance or incentive pay will depend greatly on the design of such systems (Koedel, Mihaly & Rockoff, 2015; Pham, Nguyen & Springer, 2020).

Other literature attempts to address the recognized need for further empirical data around teacher observations as performance measures, especially in high-stakes systems. Well-designed classroom observations can be informative to school leaders for personnel decisions, including retaining and supporting the development of teachers (Garrett & Steinberg, 2015; Goldring et al., 2015; Jacob & Walsh, 2011). Cohen and Goldhaber (2016) showed how research generated from high-stakes vs. low-stakes settings may lead to different conclusions, and they describe a troublesome lack of clarity and consensus about the definition of quality practice and quality demonstration of practice. The situational nature of teaching complicates the attempt to standardize observational instruments, as responsive teaching may vary depending on the students - a point also affirmed by Lazarev and Newman (2015). Cohen and Goldhaber discussed the suggestion of adjusting the observation score for student demographics and prior achievement, yet they also highlighted the problem that adjusting scores without accounting for the non-random assignment of teachers could obscure real differences in teacher quality.

The Cohen and Goldhaber (2016) analysis of rater reliability indicated that administrators often may not keep multiple dimensions of quality in mind while observing, and content-specific aspects may be particularly subject to bias. Research that suggests ways to improve rater reliability may work in low-stakes environments, but it seems to not transfer easily to high-stakes contexts. Accuracy of scores can be affected by school culture, the administrator's existing relationship with the teacher, and administrators' prioritizing of various organizational demands over a strictly objective rating according to the measuring instrument. In most examples of inaccuracy given by the researchers, the inaccuracy favors the teacher with a padded score.

Ho and Kane (2013) emphasize the importance of involving multiple observers, along with a system to check and compare the feedback given by different evaluators. Even without increasing the number of observations, having multiple observers throughout the year increases reliability of scores. Their analysis also suggests that the "element of surprise" in unannounced evaluations may not be necessary or helpful for teacher development, as it shifts emphasis to accountability rather than improvement.

More specifically relevant to the question of bias, Sporte and Jiang (2016) investigated the extent to which value-added and observation scores are related to characteristics of students in schools and to individual teacher characteristics in Chicago Public Schools. They found that teachers with the lowest evaluation scores are overrepresented in schools with the most disadvantaged

students and teachers with the highest scores are underrepresented in those schools. This applies to both value-added and observation scores, but the differences in observation scores are more pronounced, showing a stronger relationship between observation scores and student/school characteristics. More research is needed to understand why this is the case in Chicago, but the authors suggest that it could be either because it is more difficult to recruit and retain high scoring teachers in high poverty schools, or because it is more difficult for teachers to attain a high score in a high poverty school. Another important finding of their research shows that Black, Latino, and other minority teachers have lower observation scores than White teachers, but, they interpret this to be mostly due to the overrepresentation of these teachers in high poverty schools where observation scores are generally lower. Campbell and Ronfeldt (2018) also found that observation scores reflect differences in teacher identities and classroom compositions in ways that could indicate bias, or at least that observation scores are capturing additional factors beyond differences in teacher effectiveness (Campbell, 2020; Drake, Auletto & Cowen, 2019). In their study there were no significant differences by race/ethnicity on value-added scores in reading or math. Despite these indications of potential bias in the observation system, Spote and Jiang (2016) reported that, three years after their initial study of Chicago's evaluation system, most teachers still perceived feedback as fair and accurate; most teachers and administrators felt it encouraged reflection and improvement in practice, although most reported that it also increases stress and anxiety (Spote, Jiang, et al. 2016).

The positive perceptions of feedback in Chicago conflicted with the findings of Kraft and Christian (2021) in their study of teachers' perceptions of evaluation feedback in Boston Public Schools. They analyzed the effects of the district's attempt to provide substantial training to administrators to improve feedback quality and found discouraging results. The authors reported that before the administrator training program, "Teachers generally reported that evaluators were trustworthy, fair, and accurate, but that they struggled to provide high-quality feedback" (Kraft & Christian 2021, p.1). Even after a semester-long training program aimed directly at improving administrators' ability to provide high-quality feedback, the study found little evidence of improved perceptions of feedback quality, classroom instruction, teacher self-efficacy, or student achievement.

A major reason for this lack of effectiveness seems to be evaluators' difficulty in finding time to meet with teachers to provide feedback. After the training, although administrators rated the training favorably and felt better equipped to provide high-quality feedback, researchers found no significant effect on number of observations or number and length of post-observation meetings, nor did they find positive effects on teacher self-efficacy or student achievement. Nevertheless, their analysis showed that it is possible for administrators to provide high-quality feedback within the evaluation structure, and some administrators are considered far more effective than others. The two most important factors that shaped the perception of feedback quality were the administrator's tenure at a school and the racial match between administrator and teacher. There was a clear association between the perception of quality feedback and the congruence of race between teacher and evaluator.

To increase the probability of effective feedback, the authors recommended a team of content specialists with the necessary skills and time to provide actionable feedback in frequent conversations with teachers, diversifying the workforce of evaluators to enable more racial

congruence between evaluators and teachers, and cultivating school cultures of trust and collective commitment to improvement.

This report also builds upon a number of prior studies of IMPACT since its inception in 2009-2010. Dee and Wyckoff (2015) present results that indicate the effectiveness of IMPACT over the first few years of its implementation. They point to a significant increase in voluntary attrition of low-performing teachers, improved performance of previously low-performing teachers who remained in DCPS, and improved performance of high-performing teachers attributed to financial incentives.

Turning the lens towards the effect of teacher turnover on student achievement in DCPS, Adnot et al. (2017) present evidence that teacher turnover differential by teacher effectiveness under IMPACT has had an overall positive effect on student achievement in reading and math. The authors note that the positive results of replacing low-performing teachers depend upon the available supply of high-quality entering teachers to sustain the high turnover rate, and that retaining more high-performing teachers would be of great benefit. They also discuss the significant heterogeneity across their results, including the observation that high-poverty schools appear to improve as a result of teacher turnover under IMPACT.

Most closely related to this report, Dee and Wyckoff (2017) examined IMPACT in its first six years, asserting that, despite the contention and political consequences involved in implementing a high-stakes teacher evaluation system, the quality of teaching in DCPS has dramatically improved under IMPACT. They point to IMPACT's ability to effectively differentiate between low-performing and high-performing teachers and the motivating effects of the threat of dismissal and incentive pay for teachers on either side of the "effective" threshold. The various components of IMPACT, they find, have had a positive influence on teacher retention and performance, which in turn has led to improved student performance. Importantly, they point out that the early statistical gains achieved as many of the least-effective teachers exit will be difficult to sustain over time. Nonetheless, they contend that IMPACT shows that it is not politically impossible to implement an effective, high-stakes teacher evaluation system, but they also point out the necessity of continual change and improvement of any such system.

3 Methodology

As originally charged by DCPS, the primary purpose of this research was to understand teachers' perspectives of IMPACT to inform the improvements to the evaluation system. As such, teacher interviews were the main focus of the research. However, to triangulate data and better understand how multiple stakeholders' perspectives relate to teachers' perspectives, we also gathered data through school leader focus groups and a school leader survey and examined other forms of data given by DCPS to the AU team. More information about each data collection procedure is included below.

3.1 Teacher Interviews

The AU research team conducted 46 semi-structured interviews with teachers in spring of 2020. We developed the semi-structured interview protocol using the research questions, feedback from DCPS, and the literature on teacher evaluation to guide the formation of the interview questions (see Appendix B for the interview protocol). The interviews were approximately 1 to 1.5 hours in length and were conducted online via Zoom, due to COVID. In order to assure internal validity and reliability, triangulation of the data sources as well as member checks (Merriam, 1998; Morrow, 2005) were used. Further verification strategies were employed in an effort to ensure rigor including: (1) methodological coherence, (2) appropriate sampling, (3) collecting and analyzing data concurrently, and (4) thinking theoretically (Morse et al., 2002). Participants had the opportunity to member-check their transcripts for accuracy and for confidentiality (identifying quotes they would not like to be shared due to identifiability). The research team engaged in interrater checks by having two or more raters code five transcripts and discuss any disagreement in codes.

3.1.1 Teacher Interview Sample.

To obtain a sample both representative of DCPS teachers as a whole and also likely to offer a diverse range of perspectives on IMPACT, we mutually agreed with DCPS to select a stratified sample that would prioritize representation across wards of the District of Columbia and grade levels (elementary, middle, and high school). Once this stratified sample was selected, we then examined that stratified sample and intentionally sampled to consider representation of other important teacher identities and experiences, such as race, gender, level of experience, pathway into teaching, IMPACT rating last year, whether or not teachers have individual value-added ratings, subject (including Teaching of English to Speakers of Other Languages), grade level taught, whether a teacher is classified as general or special education, and Title I school status. This intentional sample was then invited to participate. Participation was voluntary and participants were assured confidentiality. Additional intentional samples were drawn throughout the study until there was sufficient participant representation, as much as possible, across the characteristics, above. Finally, we also conducted snowball sampling, asking participants whether they knew of other teachers with a different opinion from their own that we should interview in order to ensure we had sufficient variation of perspectives in our sample.

Overall, our sample of interviewees is quite representative of DCPS teachers on a number of dimensions. The sample has 22 Black teachers (48%), 20 female and 2 male; 16 White teachers (35%), 10 female and 6 male; 4 (9%) Hispanic/Latinx teachers, 3 female and 1 male; one American Indian or Alaska Native female teacher (2%), and 3 (7%) teachers of Other or Unknown race, 2 females and 1 male. Compared with DCPS overall, our sample is slightly more female (78% for our sample vs. 75% for DCPS) and does underrepresent Black male teachers (4% of our sample vs. 11% of DCPS). Our sample includes a mix of subjects, with 10 (22%) elementary/all subject teachers, 10 (22%) math teachers, 7 (15%) ELA teachers, and 6 (13%) Special Education teachers, with the remainder being a mix of other subjects. The sample includes a range of grade spans, including 28 (61%) elementary teachers, 11 (24%) middle

school teachers, and 7 (15%) high school teachers, and represents a geographic range across DC with teachers from all eight wards and slightly larger numbers of teachers from wards 3, 4 and 8. Our sample represents a wide range of experience, with 7% of teachers being new to DCPS, 22% having 10 or more years of experience, and the remainder roughly uniformly divided in between, comparable to the overall sample of DCPS teachers although with slightly fewer brand new teachers. Thirty-five (76%) are at Title I schools, a very similar rate to the district as a whole. Three participants received a Developing rating, 21 were rated Effective and 19 were rated Highly Effective in 2018-19, with three having no prior rating. We also examined whether the perspectives on IMPACT that our respondents held were representative of the perspectives that teachers shared on the DCPS Insight survey. Perceptions of IMPACT were slightly negative on average for teachers who took the Insight survey, and our interviewees on average had statistically more negative opinions about IMPACT on the Insight survey compared to the average Insight respondent (as determined by two-sample t-tests). However, our respondents did hold a range of views on the Insight items from positive to negative.

3.2 School Leader Focus Groups and Survey

In summer of 2020, the team conducted four focus groups with school leaders. However, the participation in the focus groups was less than desired: two focus groups with principals (2 participants each) and two with assistant principals (5 and 7 participants, respectively).

DCPS requested that we also conduct a survey to receive more comprehensive feedback from school leaders. In December of 2020, AU research team sent an anonymous survey to 276 school leaders of DCPS about their perceptions of the IMPACT teacher evaluation system (the population of school leaders who were not in their first year, and therefore had experience with IMPACT). 75 school leaders responded and agreed to take the survey (27% response rate), but only 63 respondents completed the survey beyond the demographic questions. Among respondents, the following were descriptives on the sample: 48% were Principals and 52% were Assistant Principals; 51% were Elementary School, 10% were Middle School, 24% were High School, and 15% were Education Campus; 73% were in Title I schools and 27% were not in Title I schools; 17% were from Ward 1, 3% from Ward 2, 17% from Ward 3, 18% from Ward 4, 4% from Ward 5, 10% from Ward 6, 14% from Ward 7, 16% from Ward 8, and 1% Other; 11% were in their current role for 1 year; 25% for 2 years, 13% for 3 years, 11% for 4 years, 40% for 5+ years; 1% were Asian, 45% were Black, 13% were Hispanic/Latino, 25% were White, 9% would prefer not to say; 75% were Female, 20% were Male; and 6% would prefer not to say.

3.3 Data from DCPS used in the Report

The sampling for qualitative interviews as well as quantitative analyses for this report were informed largely by administrative data provided by DCPS. We supplemented these data with an adapted version of the TNTP Instructional Culture Insight survey (TNTP 2021) (<https://tntp.org/teacher-talent-toolbox/insight-survey>) with questions specific to this IMPACT

Review co-developed by the AU research team and DCPS added to the survey. DCPS provided demographic data about teachers with information about the schools they taught in, subjects, grades, and student populations taught, and current and historical information about IMPACT ratings with scores for each component of IMPACT going back as far as 2009-2010.

4 Findings

In this report, we organize the findings according to themes culled from teacher interviews and then supplement perspectives from the school leader data (focus groups, survey) and other data from DCPS to triangulate the themes with these other forms of data. We also noted where themes emerged from the school leaders that were not seen from teacher interviews. However, the teacher interview themes were the primary organizing principle for the results because the foundational purpose of the analysis was to highlight teachers' experiences with and perceptions of IMPACT to support improvements to the evaluation system. As such, we describe broad themes of teacher interviews first, and then present findings from the school leaders and the quantitative analyses of DCPS data, as they relate to that theme.

In each thematic section, we give example quotations from participants and list specific suggestions that participants shared to improve IMPACT, as they are connected to each theme. We only included quotations from teacher participants who affirmatively confirmed that these quotations could be used by our research team, to ensure confidentiality. We used school leader comments from the survey whenever appropriate without affirmative confirmation because the school leader survey was submitted anonymously. The recommendations in this document are derived directly from teachers' voices or school leader voices. In certain sections, where illustrative, we separated perspectives by any of the characteristics of interest (e.g., ward, school-level, IMPACT rating).

The findings are organized, below, by research question. Themes in interviews overlapped across research questions, and therefore we included themes within the research question that most closely matched the theme. Where participants had suggestions for how to improve IMPACT within each theme, those suggestions (along with accompanying quotes) are included below the theme (note that there were not teacher participant suggestions associated with every theme). In each section, we shared quotes pulled from the entire teacher participant pool, and then specific quotes by certain teacher or school characteristics to highlight areas of agreement or disagreement.

4.1 Teacher Experience of IMPACT

Research Question 1: *How do DCPS teachers and school leaders perceive IMPACT as a feedback, evaluation, accountability, and incentive system? What do they perceive could be improved and how?*

Several different data speak to answering research question 1, including teacher interviews, the teacher Insight survey, school leader focus groups, and the school leader survey. In this section, findings of broader perceptions of IMPACT are shared first, followed by specific themes that relate to RQ1.

4.1.1 RQ 1 General Perceptions of IMPACT

Through interviews, teacher participants described their broad, general perceptions of IMPACT, and their perceptions varied widely. On the whole, across all participants, there were more negative perceptions of the evaluation system than positive perceptions among interviewees. Participants shared that their experience of IMPACT varied considerably by the way that individual school leaders implemented the process at the school level. Below are three quotes that illustrate broad teacher perceptions of IMPACT:

“The understanding is that in theory, it's supposed to be systematic and fair and applied the same way so that you have one rating system and one pool, like if a teacher at one school is held up against another teacher at a different school, you should be able to say, ‘Oh, they have these same ratings, that they should be the same.’ I don't think that's true. I know that's not true. That it often feels very subjective across the district, as far as principals being able to do observations, even though there's the rubric that people can go in and be like, “Well, I'm going to say that this is that way.” That they can apply the rubric in the way that they want to base on what they see.”

“I guess I think it's an important conversation to have. My perception overall of IMPACT is that I don't mind it, actually. I know talking to my colleagues and peers that it's either I love it or I hate it kind of thing based on how you've been evaluated. When you reached out, I was like I would love to be able to give my perspective because I think it could be a little unique.”

“My general perceptions, it's always at the forefront of what you do as a teacher, the fact that you're going to be evaluated by your effectiveness as the teacher. I don't necessarily agree with the way IMPACT is- I'm not a fan of the evaluation system at all, I'm just not.”

In addition to teacher interviews, teachers were asked two broad questions about their perceptions of IMPACT on the Insight survey. Teacher participants in the Insight survey responded somewhat more favorably to the question of whether IMPACT supported their professional growth, 26% Agreeing or Strongly Agreeing, 27% Somewhat Agreeing, and 46% Somewhat Disagreeing, Disagreeing, or Strongly Disagreeing (average of 2.38 on a scale of 0-5) than they did to the question of whether IMPACT made them feel valued, with 19% Agreeing or Strongly Agreeing, 23% Somewhat Agreeing, and 58% Somewhat Disagreeing, Disagreeing, or Strongly Disagreeing (average of 1.97 on a scale of 0-5).

Additionally, school leader participants shared their general perceptions in the School Leader Survey. 30% Agreed or Strongly Agreed that IMPACT is an effective evaluation system for DCPS, while 57% Somewhat Agreed, and 13% Somewhat Disagreed, Disagreed or Strongly

Disagreed. Some school leaders also described their broad perceptions of IMPACT in comments, which portrayed the variation in perspectives:

“It is the most comprehensive system of evaluation that I have had in my career as an educator. It includes students' input (survey), as well as data to demonstrate growth (TAS). The indicators of success in the class observations cover the most relevant aspect of teaching and learning. Teachers and staff have the opportunity to demonstrate their commitment to school activities and goals (CSC). Finally, it holds everyone accountable for student success (IVA) and for high standards of professionalism (CP). It is clear and objective.”

“Impact is an effective evaluation system because it covers a wide range of a staff's effectiveness. Specifically for teachers it covers Essential Practices, TAS, CSC, and CP. For teachers that teach in 3rd grade and above, effectiveness is also measured through standardized testing. It is also effective because Admin is trained and normed on the IMPACT system.”

“I believe IMPACT contains parts that are effective in evaluating teachers and staff, including the EP rubric, for example. I do not believe it is an extremely effective system because it places disproportionate weight on student test scores, and the frequency of high-stakes evaluations is too high to support a healthy adult learning culture.”

“I believe that it shifts teacher focus from "students first" to "me first", especially those at risk of separation or in a position to reach Highly Effective status.”

“I do not have an answer for a specific tool, however I know after using Impact for many years, it is time to consider changing the way teachers in DCPS are evaluated.”

Beyond general perceptions, there were three themes we identified in our coding that addressed research question 1: teachers' understanding of how IMPACT works, how teachers perceive that IMPACT connects to student learning, and the teacher experience of IMPACT creating negative school climates. Several other themes about the teacher experience of IMPACT as an incentive system and feedback system will be addressed in sections on other research questions.

4.1.2 RQ 1 Theme: Understanding How IMPACT Works

Teacher interview participants had mixed understanding of the purpose of IMPACT and the specific components that make up the IMPACT score, including making sense and use of the Essential Practices rubric. Interviews also revealed that some participants could articulate exactly the measures that make up their score, while other participants did not know how IMPACT works and lacked an understanding of how to find resources to help use IMPACT to contribute to their professional growth. Here are several quotations from teacher participants (across all teacher identities and effectiveness ratings) that illustrate this theme:

“The information as far as getting that ahead of time, I just know that these emails and books come off like this is what the Essential Practices are for this year. I'm unclear as to what that is beyond that.”

“Yes, I honestly don't fully understand exactly what it is. I think it also has to do with the pay scale or whether you get bumped on the pay scale...”

“If you have that rubric and you have that structure, it gives you a way, like a neutral thing to go back to say, “Okay, we're focused on whatever category.” I don't even know what the categories are here. I can remember the categories from my old school better than the new school because we don't use it...”

Teachers who were rated as Highly Effective during the 2018-2019 IMPACT cycle were generally able to describe in detail the various components of the instrument.

“There's a criteria, there's a rubric that we have to meet and then it's weighed from zero to four. Measure zero to four. Ultimately, my goal is to land between 3.5 and four, and it hasn't always been that way. That's the measurement scale but then there's also the CSC, the commitment to the school community, and also your TAS, that is equated into all of that. The most heavily weighted is your observation score. I think I answered your question.”

“I mean it's pretty straightforward. You get your IMPACT book in the beginning of the year, they tell you if anything has changed. They tell you the percentage of the weights and all that other stuff and they say, “Six weeks,” and then we start. That's it.”

“Then again, cycle two observations will start around January. After January, if you are in cycle three, you will get another observation around March or April, sometimes May if things are backed up. Then in June, you'll submit your second CSC binder for the second half of the year. Along with that, your TAS goals and the breakdown and data, and hope that everything works out the right way.”

Many teachers, including those who were rated as Highly Effective during the 2018-2019 IMPACT cycle, commented on roll-out challenges and inconsistencies across DCPS:

“They have these sessions, at the beginning of the year, and then you have to somehow find out where the sessions are, they're not really widely published, but they have been at different schools all over the city, and you just have to figure out. A lot of times the union will inform you, but the school district, I don't think does a very good job in publicizing it. Union does it because they want to make sure, hey, they're in the grievance thing, and they want to make sure that you're on point.”

Some school leader survey participants also described their understanding of IMPACT:

“IMPACT includes multiple data points for gauging teacher effectiveness (observation, student data, etc.) It requires that teachers be observed multiple times in a school year, which is an improvement from our previous system. It is aligned to a career ladder and has opportunities for performance pay.”

In interviews, teacher participants shared some ideas for improving teachers' understanding of IMPACT. Participants described a desire for greater depth in the way teachers learn about IMPACT and the training of administrators on implementing IMPACT, which could improve teacher understanding.

"I would say-- When I say least effective, please understand I'm not saying we need to get rid of it at all, because I think it's great. I would say for me, in my experience, the least effective would be how IMPACT was rolled out. I know that as a new teacher...it's like an all-day orientation, getting it one time is not enough because we got so much information that day. ...For me, it was that first step because I think that's good to have it, but as a new teacher, we need to have it presented to us maybe one or two more times prior to us starting the full cycle of IMPACT evaluations and our scores actually being inputted to the IMPACT platform."

"For another aspect for the IMPACT is what I had mentioned with the administrators making sure that they have a pretty good understanding and grasp. Then also making sure that there's some type of system...for how these administrators can establish their relationships with the new teachers, or teachers that are new..."

"...that one day of training on just IMPACT, like I said, I know these things are costs or what have you. It might be something for just new teachers, but I think that that would help to change the tide. Might take one to two to three years for everyone to get on the same board, but something like that I see would really help."

School leader survey participants resonated with the teachers' suggestion that more training would improve IMPACT. 47% of school leader participants Agreed or Strongly agreed that more training and norming experiences for school leaders would help improve their objectivity on Essential Practice assessments, whereas 29% Somewhat Agreed, and 24% Somewhat disagreed, Disagreed, or Strongly disagreed. Many school leaders also discussed their desire for better training in their comments and suggestions for improving IMPACT:

"IMPACT can be a fair system, but some areas of the system require more norming among school and district teams, especially now that instructional leaders are evaluating virtual instruction. The bar of what constitutes highly effective instruction and leadership is not clearly defined."

"IMPACT clearly articulates the skills and behaviors teachers should have. There needs to be more norming across administrators/evaluators to help with consistency of application."

"In order to reduce the subjectivity inherent in some components of IMPACT (observations primarily), we need more time to calibrate systemwide."

"Greater depth in the ways teachers learn about IMPACT (e.g., orientation) and the training of administrators on implementing IMPACT, which could improve teacher understanding"

"More training for staff and norming possibilities for school leaders."

4.1.3 RQ 1 Theme: Variation in Perspectives on Student Learning

Participants had vastly different perspectives on whether IMPACT contributed to student learning. The greatest proportion of the teacher interview respondents said that IMPACT had a detrimental effect on learning through a fear-based climate and stifling innovation. Another large proportion described that it did improve student learning because it held teachers accountable. A smaller group of participants described that IMPACT had no effect on student learning, because they would be intrinsically motivated to improve their practice regardless of the evaluation system in place. The variation of participants' belief in the effect of IMPACT on student learning is reflected, below, in quotations from teacher participants.

Some teacher participants described a negative effect:

“Therefore, I have provided better student learning or helped aided better student learning, but day-to-day, year-to-year IMPACT does not matter to students. In fact, I would say give me my observation and get out of my room so I can get back to teaching them the way that I think I should be teaching them.”

“I had to be careful this year, in particular, because I found myself as a teacher making decisions about how I would teach that were not based on whether it was the best that I thought was for the kids.”

“I was starting to make decisions based on what I thought my principal would want and not having the confidence to be able to verbalize, “I’m making this decision because I think the students need X, and here’s the rationale.” I was giving up and just being like, “I’m just going to do what they want me to do.” That wasn’t good because then, in reflection, I could see, I saw the data and I should have gone with my instincts. I also want to keep my career. I love my job.”

Several Teachers who were rated as Highly Effective also largely perceived that the IMPACT evaluation system has a negative effect on student learning.

“An observation can be great for teaching in the moment, but so much of teaching is planning and being planned, and not only just the daily plans, but long-term plans. I don’t know that that’s the most impactful on student learning because I feel like you could get a really good score but have your kids still not really learn anything or not everything that they need to learn.”

“These are important things, and teachers are in this environment themselves of a lot of those messages being contradicted. When we walk into the classroom, we’re feeling uptight, scared, threatened or we’re feeling great because we just got \$2,000 for being a good teacher. None of that trickles down to the students in a good way and you know we’re teachers, everything trickles down to the students. Our mood, our bearing, our body language, all of the implicit bias that we have, it all gets filtered down to the students. If we’re constantly immersed in this atmosphere of reward and punishment, how are we going to do the right thing with our students and tell them that they’re not working for reward and punishment? We can’t. I can’t. I snap at my students and I get scared, when I know they’re screwing up my TAS data. I know I’m not supposed to, I don’t want to, but I know I’m doing it. That’s it.”

"I don't think IMPACT matters to student learning one way. I've had administrators come in to observe me and when they leave, the student [unintelligible 01:04:39] and seen as if they were putting on a show. When the administrators left, they were going back to what they were doing. The students know it's a game and whether that's because the teachers tell them or not, they know it's a game, so they don't care. Again, like me, some teachers have grown a little bit stronger through IMPACT and have been better teachers."

Some teachers perceived that it had a positive effect.

"I think it's directly impactful to student learning because it's measuring how great of a teacher you are. Are you challenging students? I think it directly impacts students, and it can harm them or either benefit them? How great you are as a teacher will have a significant impact on how great your students are, how great they develop from beginning of the year to the end of the year, what skills do they teach. How did you form their thinking? ... it's much needed ... if we close the learning gaps of students, particularly the racial disparities that we know lie in education...It definitely impacts our students."

Additionally, many teachers (across various identities and effectiveness ratings) also perceived that the IMPACT evaluation system set common expectations and helped develop a shared language across DCPS.

"I think it sets out common standards, you know what you need to be thinking about when you plan a lesson, that is fine. I think the question is what are we trying to get the system to do? If we're trying to use it to get rid of bad teachers, then it's perfect. It's effective. If we're trying to make teachers grow and improve their practice, it's absolutely not."

"Everybody tries to make that highly effective goal-setting. People want to be highly effective. People want to be-- I hear a lot more, 'I want to be the best teacher for my students' than I do, 'I want to get that extra bonus incentive.' I think it makes people want to be better teachers. I think that they get bogged down with the minutiae of it all as it goes through though."

"I think the expectations are set with the school mission, the district mission. Those are our focuses. Every student should be loved, challenged and ready to thrive in the world. Those are the things that I think are first and foremost with teachers rather than the things on IMPACT. The things on IMPACT, one of them is the students are welcoming to each other or something. People know how to do that, they tell the kids to snap. [laughs] They tell them to snap. It's not they're snapping because they feel it."

"It's a strong tool, people might not like that. I would say on a scale of one to 10, it's about eight or nine, but the tool has to match the goal. If DC wants to be number one, you have to use a stringent tool. So that's one part of it."

Some teachers perceived that there was no effect on student learning.

"Nothing at all. I don't think IMPACT matters to student learning one way."

School leader participants shared their perceptions on IMPACT's effect on student learning in the survey. Like teachers' descriptions in the interviews, there was considerable variation among school leaders in their perceptions: 35% of school leader survey participants Agreed or Strongly Agreed that teacher IMPACT benefits student outcomes, whereas 36% Somewhat Agreed, 21% somewhat disagreed, and 8% Disagreed or Strongly disagreed.

School leader survey participants also shared the important role that IMPACT had in setting expectations for teachers. 53% of school leader participants Agreed or Strongly Agreed that IMPACT helps to set high expectations for teachers, whereas 34% Somewhat Agree and 13% Somewhat disagree, Disagree, or Strongly Disagree. Several school leaders also commented on the survey and during focus groups on the way IMPACT sets expectations for teachers, for example:

"IMPACT is a great way to hold ourselves accountable to high standards with measurable, concrete goals."

"It is helpful to have some common language for teachers to understand what is expected of them in a classroom."

"IMPACT provides direction to staff about the expectations of the district. Evaluation tools are one of the few levers that districts have to shape the culture of the staff. IMPACT is not an easy measure to achieve "effective" or "highly effective" (especially for principals), but it is fair, clear and effective at ensuring at least a minimum level of performance and professionalism from the staff."

4.1.4 RQ1 Theme: School Culture/Climate of Distrust, Anxiety, and Competitiveness

A preponderance of teacher interview participants shared that IMPACT created a negative culture and climate in participants' schools. The negative climates that participants described focused on issues, such as, unhealthy competition, stress/anxiety, and undermining trust/relationships. Below are some example quotations from participants.

"...where I would assume IMPACT should strengthen a school community. It actually creates a very competitive, and I feel sometimes mean-spirited environment in my particular school."

"It's a terrible, terrible anxiety-producing system. They intended it to be that way and it's clear evidence that the district does not trust the teachers."

" 'Oh, it's IMPACT time,' and so the sense of stress was heightened and that sometimes would be the only interaction you would have with administration."

“I think that it sometimes makes teachers not trust each other as much. Especially if they know that someone has highly effective and only needs one observation, as opposed to newer teachers or teachers with lower scores that have to do the three observations...I think it sometimes leads teachers not to really trust the administration, because they're afraid to seem like they need to ask questions, or aren't confident in different subject matters knowing that they might be evaluated then in that area...I think it may lead to lower teacher morale. I think it's something that most teachers worry about and stress about in DCPS.”

“Nobody trust nobody... IMPACT has become a tool that nobody trusts because the leadership wants the test score because their job is on the chopping block and they decide, okay, if my job is on the chopping block, I'm going to put your job on the chopping block too because these kids did not get proficient or advanced, or whatever they call it now.”

“When the feedback matches the observation, then trust is high. When the feedback doesn't match the observation, then yes, trust is hard to come by. That can be really stressful. I think stressful both for the person experiencing the negative ratings, but also for the people around because they start wondering, ‘All right. Who's next?’... I think it raises anxiety with teachers and I think that affects the climate in the classroom.”

We examined whether there were differences in coding patterns by ward and by teacher race and found that the perception that IMPACT contributed to a negative school and district culture transcended teacher identity and wards of the district. In the below section, we provide a few example quotes from teachers who identify as different races and also some example quotes from teachers in different wards.

Teachers who identify as Black shared:

“I'm thinking that the name of the evaluation system should be changed just because of the connotation that it has, and then the fact that one of the masterminds behind it, Jason Kamras, he's now down in Richmond and he was already told before he got on 395 South to not even bring that down there. Then he admitted that it instated a culture of fear. For the person who created it to say that it instilled fear, why is the district still married to it? That's an issue.”

“I feel like that's where IMPACT becomes negative and competitive amongst educators, because now we're all chasing these achievement scores and these things, and we're forgetting that there's so many other components to our work that actually help kids and improve communities.”

A teacher who identifies as Hispanic shared:

“I feel it could be difficult to trust people sometimes because you may never know what conversations may be heard when it's their IMPACT time. Like you don't know if it's-- Well, I didn't do so great because this was the block that someone was supposed to come and they didn't do their part.”

Teachers who identified as White shared:

“I think it definitely has a negative impact on our school culture. I can test personally, definitely makes us in the testing grades feel a lot less valued, appreciated than it does compared to teachers in other grades.”

“I think it may lead to lower teacher morale. I think it's something that most teachers worry about and stress about in DCPS. I don't know that it's really led to teachers learning a lot more and becoming better.”

“I think it raises anxiety with teachers and I think that affects the climate in the classroom.”

The perceptions that IMPACT contributes to a negative school and district culture was expressed by teachers across wards.

Wards 1-2:

“Like I told you, this could be cutthroat environment sometimes is because people don't want nobody to get in their other way of getting their bonus again. I feel that even if I prefer to have my salary increase to keep moving in that ladder if something is going to give because the bonus might be something that is not-- Well, everybody wants the money. I'm not going to say I don't want the bonus, but you can see that that could create an unhealthy environment. I think that's what I'm trying to say. An unhealthy environment.”

Ward 3-4

“When you tell me we're going to talk about IMPACT, all I can think about is how to improve it and ways that there's some positives to it, but a lot of it is just additional stress, problems. It's not seen as this positive tool in the teacher world as where I think it was maybe constructed like that. I have a principal and I love the way she's going about it and it still has that fear. It's just not a positive vibe. When you hear IMPACT, everyone just-- Their heart rate goes up, their blood pressure goes up. It's just a stressful kind of thing.”

Ward 5-6

“I don't think there's any lack of trust about the use of IMPACT, but the fact that we have IMPACT at all is based completely on a foundation of mistrust. IMPACT wouldn't exist if trust existed. It's a constant reminder that fundamentally, teachers are not trusted and that fundamentally, principals aren't trusted. If they trusted principals, they wouldn't give them

IMPACT to evaluate us with. They would trust them to take care of their people. Everybody is in agreement that we just go to work on a complete solid bedrock of mistrust and we have to somehow work with children in that environment. I think that's where we are as far as the issue of trust goes."

Ward 7-8

"Where I would assume IMPACT should strengthen a school community. It actually creates a very competitive, and I feel sometimes mean-spirited environment in my particular school."

"Everybody knows who the favorite people are. People know and it has been viewed as administration gave those ratings if they were favorable, and they could always take it away. I think there was definitely this stress, and anytime you saw administrator walking around with their laptop, it was like, oh, it's IMPACT time, and so the sense of stress was heightened. That sometimes would be the only interaction you would have with administration."

Teachers who were rated Highly Effective in 2018-2019 also echoed the perception that the IMPACT evaluation process contributed to a negative school and district culture:

"Yes, I think it sometimes leads teachers not to really trust the administration, because they're afraid to seem like they need to ask questions or aren't confident in different subject matters knowing that they might be evaluated then in that area. I've not really heard any teachers say anything positive about it. I also know that the administrators get stressed by having to do so many time-consuming reports."

"I think it may lead to lower teacher morale. I think it's something that most teachers worry about and stress about in DCPS. I don't know that it's really led to teachers learning a lot more and becoming better."

"I think IMPACT does a lot across DC to determine school culture, determine what the experience of teaching is in DC. It is probably the biggest unifying factor in determining district culture.?"

The theme of negative climate was also echoed in school leader comments, for example:

"I feel that IMPACT is a system that does not allow for an improvement of teacher practice. It is a very punitive system and causes much anxiety to administrators and teachers. This system makes teachers feel like it's an "I gotcha" system versus one that is more welcoming and announced."

"In addition, the amount of money that teachers can receive in their bonuses creates divisiveness in the school community."

“The mood and climate of a school drastically change during IMPACT season and it creates strained relationships between staff and admin.”

The theme of IMPACT creating a climate of competitiveness, anxiety, and distrust was even echoed by school leaders who perceived other parts of IMPACT to be useful:

“I believe that IMPACT outlines clear expectations for DCPS staff members. The rubrics and opportunities to ensure targeted feedback are helpful. I strongly believe that the bonus structure is flawed and creates unnecessary anxiety for staff. The power given to administrators also creates a challenging environment and morale tends to shift as teachers/staff interpret their scores.”

In interviews, teacher participants shared some ideas for improving school culture, including a desire for greater trust and autonomy for teachers.

“What they really need to do is just relax, trust their teachers, provide us and make sure that we have the materials and supplies and equipment we need.”

“Since then, now that we are actually an improving school system and now that they have cognitive teachers that are trustworthy, they need to let go. They need to relax and trust us like they do in Connecticut or in Massachusetts or New York...”

Some school leaders echoed this call for teacher involvement and trust in their survey comments, for example:

“Teachers need to be at the table if IMPACT is going to be revised, updated, etc. Giving voice and space to the recipients of this high stakes evaluation is important to help increase buy-in. Most professions may not give employees this opportunity, but we are directly impact children's lives and it is critical that DCPS is improving teacher practice in order to increase student growth.”

“Think about having teachers report how they will commit to the school rather than we telling them and they complying for compliance sake. A commitment must come from within and not be coerced for score.”

“A lot of what's missing in this is teacher voice. It's my professional evaluation. I don't understand why teachers don't have a voice in the areas they wish to grow in. You would see greater growth, you would see greater investment. Then entire school communities can work around goals, which means opportunities for school wide professional development, which means community buy in. Again, teacher voices is absolutely huge.”

“For this tool to support teacher's professional growth teachers should be empowered to self-assess and then compare their assessment with that of evaluators to determine an objective rating and next steps.”

4.1.5 Summary of Research Question 1 Findings

Overall perceptions of IMPACT were more negative than positive, but with a great deal of variation in perspectives across teachers as well as school leaders. Many teachers and school leaders perceived that IMPACT played an important role in expectation setting for DCPS. However, many teachers and school leaders perceived that IMPACT created an unhealthy environment of distrust, fear, and competitiveness in schools that trickles down into the classroom. These themes held true across teacher effectiveness ratings, teacher race, as well as ward.

4.2 Professional Growth

Research Question 2: *How does IMPACT facilitate DCPS teacher improvements? How can IMPACT be altered to better support teacher improvement?*

There were quantitative findings from the Insight survey, the school leader Survey, and DCPS administrative data that spoke to this research question. In addition, there were qualitative findings from teacher interviews and school leader comments on the survey.

4.2.1 RQ2 Quantitative Findings Pertaining to Professional Growth

We investigated from the Insight survey how teacher perceptions of how IMPACT affected their professional growth varied by teacher and school characteristic and component of IMPACT as shown in Tables 2 and 3. In general, teachers reported that Essential Practices and Contributions to the School Community contributed to their professional growth more than other components of IMPACT across wards and teacher racial and gender identities. Overall, female teachers had more positive perceptions of IMPACT contributing to growth than male teachers, with the exception of Black and Asian male teachers, who had the most positive perceptions of IMPACT contributing to growth overall, with White teachers and teachers in Ward 3 generally having the least positive impressions of IMPACT contributing to growth. This could be due to several factors, for example: survivor bias due to higher numbers of teachers of color and teachers in wards with greater student economic need and lower average household incomes having faced higher rates of dismissal previously, leaving the teachers who did the best under IMPACT and thus have more favorable impressions of the system in schools; differences in survey response rates across groups with a more positive self-selected sample responding in some cases; as well as different experiences with IMPACT among subgroups of teachers and across the city.

We also constructed two simple measures of changes over time in IMPACT scores themselves, though they may not fully capture elements of professional growth that are not currently measured by IMPACT, by comparing changes in Essential Practice rubric scores in Cycle 3 to Cycle 1 in 2018-19 and changes in IMPACT scores from 2017-2018 to 2018-2019. Note that the latter analysis could be subject to similar survivor bias as above, as the lowest scorers will be dismissed and not have a second year of scores and first-year teachers in 2018-19 will only have one year of scores; thus, these analyses should be considered exploratory and results interpreted

with caution. The purpose of this analysis is not to causally attribute growth to IMPACT itself, but rather to investigate descriptive trends and associations with other factors that are related to the growth measured by IMPACT. We show the distribution of EP cycle changes in Figure 1 and regressions of EP Cycle changes and IMPACT score changes on teacher and school characteristics in Tables 4 and 5, respectively. These regressions allow us to determine how growth measures are related to teacher and school characteristics independently of other characteristics (i.e., statistically holding them constant). This is important to disentangle the interrelated effects of several influences on how teachers experience IMPACT; e.g., Black teachers are 4 times more likely than White teachers to teach in a Title I school and 5 times more likely to teach in Wards 7 or 8³. Thus, differential experiences of IMPACT by teacher characteristics can be conflated with school characteristics if teachers in different schools experience IMPACT differently and teachers with different characteristics are likely to teach in different schools.

As Figure 1 shows, the majority of teachers do increase EP scores between the beginning and end of the year, with an average increase of about 0.11 points on a scale of 1-4. Coefficients in regressions are interpreted relative to omitted categories - White teachers, male teachers, non-Title I schools, Ward 1 schools, and elementary schools. Coefficients with one or more asterisks are statistically significant, meaning they are not likely due to chance alone. Teachers who identify as Hispanic or Latino, teachers in Ward 7, and middle and high school teachers are statistically more likely to show growth in EP scores during the year, compared with teachers in these omitted categories. Similarly, Black, Hispanic/Latino, Other or Unknown race, and middle school teachers are likely to show growth in IMPACT scores from year to year, conditional on having two years of scores, and teachers in Ward 2 are significantly less likely to do so.⁴

Beyond quantitative analyses, there were two themes that emerged from interview coding that spoke to research question 2: alignment with professional development and the consequences of high stakes incentives for teacher growth and innovation.

4.2.2 RQ2 Theme: Alignment with Professional Development

By contrast with the quantitative findings, teacher interviews revealed that many teachers in our study perceived that IMPACT does not support professional growth, development, and improvement for teachers. Most participants shared a lack of understanding of how IMPACT aligns with Professional Development (PD) opportunities at their school and district wide. Professional Development includes PD trainings, coaching, and feedback. Teacher participants shared that feedback did not align to training and goal setting, or that feedback and goal setting were not present in the process. We examined whether the perceptions of professional growth alignment with IMPACT varied across teacher race and Ward and found that this perspective was consistently held across groups, however, there was some variation in perspectives within

³ According to DCPS, they intentionally recruit teachers of color, especially Black teachers, to teach in schools that predominately serve Black children--an effort that is backed by research (e.g., Howell, Norris, & Williams, 2019; Lindsay & Hart, 2017)

⁴ Additional analysis connecting IMPACT and LEAP were conducted by the AU team and will be available at a later date.

each group. In other words, across teachers (by race) and Wards, our coding revealed similar patterns, where a majority of participants shared that there was a lack of alignment between IMPACT and PD, but a few participants did see a good connection. We share, first, some teacher quotations from participants who did not see a strong connection between IMPACT and professional development:

“I think the general perspective it's definitely the evaluation system. It's meant to be to find effective teachers and teachers who maybe aren't doing a great job. They say that it's supposed to have professional development implications, but it doesn't. Then I guess as far as talking about IMPACT, I think it's that you want feedback on it but then again I think my perception of that would be does the feedback really matter?”

“I was more so-- the purpose of those conversations and those evaluations was to help me improve my daily practice. Whereas here when I came, DCPS, it's almost like you're chasing a score, rather than actually improving your practice as a teacher.”

“I think that where it's falling or what people are feeling is the fact that IMPACT is just the measure and maybe teachers are not feeling that they're receiving specific instruction and professional development to hone their skills to reach the height of IMPACT.”

“That's the best way I can say it. The feedback that I will typically receive from, let's say a leader at the school is not aligned to what the district actually has outlined for that content area or for that grade level. Sometimes I'm always amused by the fact that my school will say this, I'll go to PD and I hear this. Then I'm like, well, what do I do? Then it becomes more of me just making a decision that works in the best interest of my students and myself.”

“I feel like I'm the same teacher I've been for [10+] years. I don't think IMPACT had to do with my professional growth if you will.”

“The PDs that we have on this, we'll do one PD for like 40 minutes on one section of the rubric, and then the next time we have PD a month later, which is crazy, we do PD once a month. We do PD a month later and it'll be on something totally different, not even on the rubric. Then three months later, we'll come back to the rubric and it's a different section and there's never any follow-up about the first thing that we did.”

“I think over the years they've had maybe one or two PDs during pre-service week where they discuss some aspects of the Essential Practices, but it's not really discussed consistently or constantly throughout the year.”

Many teachers who were rated Highly Effective in 2018-2019 shared the perception that the IMPACT evaluation process did not support their professional growth or improve their practice.

“I don't think I've ever thought of this. I don't really think any of them necessarily improve my practice. It tells me how I'm doing based on the criteria given, but I wouldn't say it necessarily improves my practice. I think my principal does a really good job of giving very specific and detailed feedback and suggestions. I don't always agree with them, but I think she

puts a lot of thought into it. I guess if any part was going to help me improve my practice, it would be the part where she comes in and observes and then gives me feedback.”

“Holistically, no. I know the district does offer some one-off PD sessions that are like, ‘Oh, teach one, whatever. That part of the rubric we’re going to have this afternoon class on it.’ It’s just not well integrated. The other thing that I think beyond it not being well integrated is it’s not differentiated at all. When we go to PD, our whole day long PD sessions or our half-day PD sessions, I’m sitting next to the same teacher, a teacher on my team that’s a first-year teacher that needs help with behavior management. Just things that I don’t need help with.”

“Every bit of growth that I’ve ever had has come through my own learning, not from [inaudible 00:14:56] at all. IMPACT is just a score to give me a raise, and basically put me in a certain category. Personal and professional growth? I haven’t got from IMPACT.”

“I’ve watched people get low scores and not receive the assistance they need to become better because the tools are not in place, or the person or the coach that’s available has a really heavy workload and can barely get to that teacher to work on certain things.”

On the other hand, a minority of teacher participants shared positive experiences with IMPACT and their improvement as a teacher.

“For coaching, they use the EPs to guide learning goals, one for our students, and learning goals for me as professional development. A lot of our professional development are around the Essential Practices. A lot of the work that we do at grade level is around the EPs”

“What I did with that information is it caused me to go online and then make sure that I had examples of what that looked like. I actually did go back to the video library and looked up in IMPACT where that particular area of the score was, being able to look at those examples and get feedback and things like that. I could see the practice in action. Yes, that’s what I was able to do.”

Below is an example quote from a teacher who was rated Highly Effective who was able to implement feedback:

“Has it changed my practice? I guess it depends on the component again because if I’m really receptive to feedback, if they tell me, “You can change this”, then I’ll seek out a way to change it. For things like CSC, I will say certain things I have changed, for example, what-- I don’t remember the name, the actual name of the component, but some feedback I got was that you could do like a breakfast with the families or something to get more in that area. I will say in certain ways I’ve been able to implement some of the feedback.”

School leader survey participants shared their perceptions of how IMPACT aligns with professional growth. Of participants, 22% Agreed or Strongly agreed that IMPACT supports teachers’ professional growth, whereas 44% Somewhat Agreed, 24% Somewhat disagreed, and 10% Disagreed. Many school leaders shared their thoughts on the lack of alignment between IMPACT and professional development or the unrealized potential of alignment, for example:

“IMPACT is an excellent tool to punish or terminate teachers who are not trying to provide a high level education for our students. There is a need for such a tool as we must ensure that teachers are teaching rigorous content to students. However, IMPACT is not a good tool for coaching and pushing good teachers to great teachers. It misses all the nuance and subtlety of behavior modification and progress.”

“I believe it can be a great tool if used properly. If we spend intentional time building teachers capacity around instructional practice and not just using it as an evaluation tool we would see different results.”

“IMPACT is effective because it provides a clear rubric for observing instruction. IMPACT is ineffective because it is not actually a tool that supports teacher growth and development.”

“Bottom line, I do not believe that IMPACT supports teacher growth. Teacher growth is supported by having a positive and supportive culture of teaching and learning. Teachers grow when there is a clear instructional plan guided by a principal who is an instructional expert, when PD is laser light focused on 1-2 top priorities and there is buy-in to want to improve and when the instructional coaching program is strong and effective. Great IMPACT scores will come as a result to that, but growth is not happening due to IMPACT.”

“IMPACT does not support teachers' professional growth because it is punitive. What grows teachers' capacity in their work is very good professional development that is coupled with great coaching, and a supportive learning community in which ideas and strategies can be shared.”

Much like the teacher interviews, a minority of school leader comments described positive alignment between IMPACT and Professional Development:

“It identifies each area that teachers can grow and improve. For teachers who are committed to the profession the area identified allows for teacher and principals to set goals for improvement.”

“I think IMPACT does support teacher's growth professionally. It is meant to provide unbiased, substantiated feedback around a teacher's performance or lack thereof.”

In interviews, teacher participants shared some ideas for improving the alignment between IMPACT and professional development, including improving the feedback process and the quality of feedback, and providing more consistent feedback. They also suggested improving the observation process by providing more flexibility and more transparency.

More Consistent Feedback:

“I think I would have teachers have the option of opting into every other week observations with the rubric that would be followed the next day by a debrief with like a feedback action step cycle that involves a follow-up observation the next time looking for that same thing.”

“I know that's not sustainable for the number of teachers that have to be evaluated, but I feel like if there were some ways where there were cycles or where maybe instead of those people who get three evaluations you could do less but more times where it's a true, and it's announced, we know they're coming, because for that amount of time, you can't fake doing good work for that amount of time on one day.”

“I think that I would change to have more observations, so that the principal or AP are in your room on a consistent basis and could give you things to work on for the next time, then follow up on those. Then, either take the scoring away from it or have the bottom score dropped an average. If they did five, average the top four. I think I would take away that, even though I enjoy only having them in my room once a year...Again, I just don't think that they're in your classroom enough to know what it looks like day-to-day and what you're teaching to the students. I think that curriculum and more weekly meetings and things like that would have a bigger impact on student learning.”

“In terms of the observations, one thing that I would really like to be different is-- The thing that scares me is the feeling that they're going to come in and they have to see everything at once in that one day. The rubric, it describes everything the teacher would ever do. The place where I feel I might have to put on some show is that I have to do all of those things somehow in 45 minutes or something. Something that would feel less anxiety-arousing would be if we were told that over the course of, I don't know, like six more informal observations, they were going to be looking for-- I don't know if I want to say six, maybe it's three. They were going to be looking for this range of practices just to allow a little bit more for the ordinary variation.”

By contrast, in survey comments, many school leaders called for reducing the number of observations or reducing observation components to reduce the time burden of the process:

“IMPACT is effective in the sense that it looks at many facets of a teacher/employee's performance, however it would be more effective if it were more focused, and less time consuming to implement.”

“All of the components support, however there are too many measures and no time to reflect upon what they mean and how to respond to them. Perhaps some measures should not be taken every year.”

“As a school leader, I find the amount of time spent completing IMPACT evaluations to be fairly significant. I've worked in other school districts where they focus more on informal observations and one summative evaluation at the end of the year. The time spent completing one IMPACT evaluation could be better spent observing multiple classrooms.”

“I also think there are too many observations and that it is too time consuming. I wonder if we can have a way where after 3 highly effective years in a row- you can opt out completely for a year (maybe only get CP if needed and CSC).”

“Fewer Observations”

“Reduce the number of formal observations for teachers that are effective or higher.”

Teacher participants shared their desire for an improved feedback process:

“I also feel that there needs to be more purposefulness when those observations are happening. I would like to see something more like the Danielson with a preop and postop after the observation itself happens just so that there is less wiggle room for administrators to say, “I didn’t see this. That may not have happened in the classroom.” “We went over my lesson before we met, and you knew everything that was going to happen, so you can’t really say that unless I didn’t say what I was supposed to say or didn’t do what I was supposed to.”

“Then if everybody had the opportunity to get the experience that I have with my administration, I think that would be helpful. Actually, giving constructive, useful feedback and just generally not feeling like someone’s out to get you but actually, they want to help you be better.”

“If you want to hold on to teachers and mentor them, make it a tool of-- like you say, after each IMPACT session, write down three goals. Write down and add some plan, put it in writing. Don’t just say, Oh, you earned a 3.5, bye. Okay. [laughs]”

In survey comments, some school leaders also described a call for improved observation and feedback processes, including, for example, pre-conferences and informal evaluations:

“Requiring a pre-conference or an informal evaluation at least before the first cycle”

“Increase the informal observation and elevate the day-to-day contributions that teachers make.”

“Improving the observation process by providing more flexibility and more transparency”

Teacher participants shared their desire for a stronger connection between IMPACT and professional development:

“I would have PDs directly connected to the different standards. If that teacher got maybe a two in a certain standard, the next time we have professional development, that will be the PD that that teacher would attend. That way, that we saw that there’s a deficit in that area, these are the PDs that line up with helping you improve in that area and give that suggested list to the teacher for them to attend at least one PD in that subgroup to help increase those scores.”

“When we go to PD, our whole day long PD sessions or our half-day PD sessions, I’m sitting next to the same teacher, a teacher on my team that’s a first-year teacher that needs help with behavior management. Just things that I don’t need help with... Especially for those of us who have been in it for a while this is my 12th year, I don’t need the same thing again. I got it. I’m ready to move on to something...It doesn’t make sense to me because I feel like they have all that information. In IMPACT, this teacher needs to work on Teach Six, this teacher needs to work on Teach Five or whatever it is. They could group us.”

“I think professional development. More professional development opportunities based on the feedback from that evaluation system. For example, somehow- I couldn't begin to tell you the 'how', but if I waved a magic wand, then hey, these teachers are all struggling in this area based upon this feedback on IMPACT. Here are some professional development opportunities for these teachers. The same way we differentiate in the classroom, it'd be differentiated.”

“... Like I said, I would have a layer of mentor teachers to have a mentorship program. I don't want anybody telling me that's what your LEAP leaders are because they are not mentors. They're placed in the schools to go, and I guess give feedback to the principals what you're doing and what you're not doing. I don't see LEAP leaders as mentors. I want true teachers who have gone through the rigors of teaching, and who understand what teaching is supposed to be, to be placed there to guide incoming teachers...That's what I want. I don't want to hear anything about LEAP leaders, or I don't want to hear anything about relay either. Because all those things, I think they're judgmental, they're punitive and they should not be.”

Some school leaders echoed calls for stronger alignment to professional development in their survey comments, for example:

“I believe it could be more effective if there were more intentional alignment to coaching cycles and if an informal evaluation and/or a pre-conference were required.”

“It is fair if used properly and provide on going PD with the tool. Everything should align to the tool used for evaluation.”

“More coaching cycles connected to IMPACT for novice teachers or teachers who are developing/minimally effective, facilitated by their evaluators. (Time for evaluators to actually engage in those cycles vs logistical work in buildings).”

“There should be accompanying support/resources to provide to teachers who are lower on the IMPACT rating. While teachers who do well should be given compensation and accolades for their performance, teachers who do not meet the criteria, or who require more support should be able to grow within the system and receive support outside of the school to improve their practice.”

4.2.3 RQ2 Theme: Consequences of High-Stakes Incentives for Teacher Growth & Innovation

Many teacher participants described in interviews that the way IMPACT was tied to their livelihoods (salary, bonus) led to unintended consequences, such as significant teacher stress, fear, anxiety, and a lack of opportunity for growth or innovation. Teachers suggested that the extrinsic reward and punishment system undermined intrinsic motivation and professional growth, and harmed relationships with school leaders. Some also described that it discouraged innovation and risk-taking among higher-performing teachers in favor of compliance. We examined whether this theme varied by teacher race and Ward, and found no patterns by Ward, but there was a pattern among White teachers compared to teachers of other races in this theme.

Below are some quotations, across teacher identities, ratings, and Wards, that demonstrate this theme:

“I think that that's why it stresses me. Because I think about that. My rent goes up every year. I can't tell my rent, "Oh, wait a minute, you can't go up because I didn't get a step increase because my principal said I wasn't effective." It's just things like that. These are things in the back of my mind that I stress about because it's like, if I don't get a step increase, then my pay is going to stay the same. My rent goes up and all my other bills, cost of living go up and I can't have bills higher than the income”

“I just find that teachers find it more punitive.”

“I feel like it's just put in place to help DCPS determine whether or not they want to keep you or not. There's, again, so much punitive action that could be taken if you're ineffective or anything below effective, and there's a chance that you could lose your job. I don't see how I couldn't necessarily feel valued because I'm living with the constant threat of if you don't do this, then you're gone.”

Many teachers who were rated as Highly Effective also echoed significant stress and anxiety around the process:

“The way that IMPACT sets up its rubric so that it tries to quantify teaching, I think that in itself is admirable. I think that the fact that they don't admit the irony of trying to quantify teaching is problematic, but I think it's interesting that people are trying to parse out what the different components of teaching is. Teaching is complex, so any effort to try to understand teaching better is an interesting idea. They did a decent job of laying out some preliminary pieces of how you would parse and analyze the process of teaching. The rubric and the process is a cool idea. I think that the use of it by attaching that to high stakes results and consequences and rewards is foolish. It doomed the entire process to complete failure. If they had left that rubric and that analysis process completely separate from the concept of rewards and punishments, they would have had a very useful tool that 11 years in, would have borne a lot of fruit. The fact that they made the decision to tie it to very high stakes consequences has completely undermined their own purpose and kept that rubric and that process from having any effect at all. It's also had the opposite effect of basically limiting teachers' relationships with their principals and with the administration and with master educators back in the day. Limiting those relationships is such a narrow focus that they've pretty much ruined any opportunity they have to do delicate work that that rubric was supposed to accomplish.”

“I think IMPACT does a lot across DC to determine school culture, determine what the experience of teaching is in DC. It is probably the biggest unifying factor in determining district culture. Much of culture is determined school by school. There are a lot of schools that do things very differently than DCPS, so there's a lot of individual school-based culture. As far as creating a district-wide culture, I can't think of anything bigger than IMPACT that does that because it motivates us by fear and it motivates us with very serious consequences and it's ever present. I've never met a single teacher in DC that isn't always considering IMPACT, even if it's some kind of very visceral level at every moment of their job. I think IMPACT has more of an effect on us than

any PD, than any emails from the Chancellor, than any district-wide initiatives, than LIFT or LEAP or any of that. IMPACT has a much greater impact on the experience of being a teacher than any other component.”

“Our administration is very good about minimizing the effect of those debriefs as much as possible and trying to approach them as genuine conversations about our practice. I think our administration is pretty good at that. They can do that all they want. Nothing's going to change the fact that they are primarily high stakes evaluations and it's only secondary that they are prompts for good conversations. My school happens to have a lot of highly effective teachers and many of us have just shrugged it off as a necessary evil. I generally have good conversations with my AP when we talk and then I throw out the paper she gives me and I never read them. I try to separate the way that this is meant to help me from the ways it's just meant to scare me.”

Additionally, school leader participants shared their perceptions of the IMPACT^{plus}, LIFT, and Bonus structure in the survey. Overall, school leader survey participants were more negative on these accompanying systems, when compared to their general perceptions of IMPACT. Of participants, 22% Agree or Strongly agree that the accompanying systems to IMPACT benefit student outcomes, whereas 34% Somewhat Agree, 20% Somewhat disagree, and 23% Disagree or Strongly Disagree. Likewise, many school leader survey participants commented on how the way IMPACT is tied to teacher livelihood has a negative impact on school culture, the classroom, and students, for example:

“Teachers use IMPACT performance to make more money. Thus, since the tool is not used the same way across the board, the focus is not on students, in on a rating that will provide a bonus.”

“IMPACT puts a lot of pressure on staff which causes challenges to teacher creativity and trust. Teachers want to ensure it is going to help their IMPACT which shifts the focus from student achievement.”

“IMPACT provides clear guidance regarding professional practices. It also provides common language. The fact that the evaluations are connected to people's livelihoods and salary bonuses, this skews the purpose of the tool.”

“When you tie performance to financial incentives, people focus on financial incentives and not professional growth. It becomes about compliance, how to score more points or arguing when points aren't scored.”

Some teacher participants shared a desire for removing the way IMPACT is tied to teacher livelihood, such as salary or bonuses:

“What I probably would say most, excuse me, if that it wasn't weighted to where if you're not effective, you don't get a step increase. Not weighing it to money, because you're playing with people's livelihood. ...I think that that's probably what I would go to if they had to do IMPACT. I would say do not make it tied to your livelihood. If you want to do the bonuses, that's fine, because that's a bonus.”

Many school leader survey participants also made this suggestion in their comments:

“I really hope that DCPS will eliminate or significantly reduce the bonus structure. I suggest that DCPS reinvest these monies for schools and students. Alternatively, DCPS could consider increasing the salary scale. I think the commendations can still be provided to staff (Highly Effective, etc.), however, the large bonus amounts staff receive or do not receive creates an unhealthy work environment.”

“Dramatically reduce/remove the exorbitant bonus structures”

“Remove bonuses linked to the scores.”

“Remove or reduce bonus checks, put money into base-pay increases.”

“There should be no more bonuses”

“Remove the bonuses and provide a system that will truly impact a teacher's pedagogy. Perhaps the bonus money can be made available to teachers in another way.”

4.2.4 RQ2 Summary of Findings

Findings from teacher interviews, teacher Insight survey, school leader focus groups, and school leader survey demonstrated that across stakeholders, and across teacher identities, ward, and effectiveness ratings, most participants saw a need for improvement in the alignment between IMPACT and professional growth. However, there was some variation among participations, and a smaller group of participants was able to make strong connections between IMPACT and professional growth. Both teacher and school leader participants shared that one significant threat to the potential for growth due to IMPACT is the high-stakes nature of IMPACT. Both groups desired more formative ties to authentic professional development, including improved feedback cycles and coaching.

4.3 Validity and Fairness

Research Question 3: *To what extent can the validity and fairness of IMPACT be improved, and if so how?*

There were findings from the Insight survey, teacher interviews, and the school leader focus groups and survey that related to teacher and school leader perceptions of validity and fairness. First, we share participants' broad perceptions of validity and fairness. Then, we delve into specific themes that related to this research question.

4.3.1 Broad Perspectives on IMPACT Validity and Fairness

On the Insight survey, teachers shared their perceptions of whether evaluation ratings accurately reflected their performance. A majority of teachers who took the survey were at least somewhat positive about the accuracy of IMPACT ratings. Thirty-eight percent Agreed or Strongly agreed that the ratings accurately reflected their performance, 27% Somewhat Agreed, and 34% Somewhat Disagreed, Disagreed, or Strongly Disagreed. Additionally, school leader participants shared their general perceptions of the validity of IMPACT as an evaluation system in the survey. 34% Agree and Strongly Agree that IMPACT scores accurately reflect teachers' performance in the classroom, whereas 51% Somewhat Agree, and 17% Somewhat Disagree, Disagree, or Strongly Disagree.

A few school leader survey participants also described their perceptions of how valid IMPACT is in their comments, where some described they saw IMPACT as valid, whereas others did not, for example:

“A 30 minute evaluation does not capture exceptional and quality instruction in a teacher's classroom.”

“I know too many really terrible teachers who are rated Highly Effective every year.”

By contrast:

“I believe the IMPACT system gives a full picture of a teacher's effectiveness.”

Teacher Interview participants shared some broad comments about their perceptions of fairness or unfairness in the IMPACT system. Several of these comments focused on the fairness of the rubric tool, itself, but the unfairness of the implementation of the process, for example:

“I think it can be fair and in theory should be fair, but I don't think it probably is entirely lived into as fair. I think it just so much depends on the administrator.”

“I think that the tool itself is fair. I would say the tool is fair. I think it's strong. I'm not sure, I would have to look back at other evaluation systems, they give more, I guess, a little more wiggle room, but I don't think at a point, that's something that we can do. It's fair in the sense of the tool. I think the unfair part might be who is actually using the rubric? I think that's where-- I guess because the district is so wide and there's so much variability across campuses.”

We examined whether the teacher interview codes about fairness were similar across teachers with different racial identities. Proportionally more Black teacher participants spoke about fairness than other racial groups. Likewise, Black teachers spoke more about fairness issues related to equity across schools, whereas White teachers spoke more about fairness issues related to favoritism or subjectivity from administrators. Here are some exemplifying quotations from Black teacher participants:

“I think if you are a teacher in a school that has those kind of populations, you can almost feel like IMPACT is- I won't say punishing you, but it doesn't seem fair in that way because they

don't-- not saying that it's like-- I don't know how to really describe it, but you feel like you've-- especially in terms of the test scores. You feel you've given your all. They've made a lot of progress, but they still didn't meet the measure. It doesn't feel fair. It almost feels like you're not always rewarded for the growth that you see in the students, so that doesn't seem fair. Then I know through conversation I've had with other people in other wards, it seems as though if you're in a struggling school that your IMPACT tends to just be lower. It doesn't seem fair in that way."

"...what doesn't make it fair is, while we're all held to the same-- and I would say by ward, we're all held to the same standards of achievement, which I don't think should change. If we're held to the same standards of student achievement, what's the resources and manpower that's put in place to ensure that every child in the district has an equitable opportunity to meet those student achievement goals? That's what makes it unfair."

Whereas, for example, from a White participant on the subject of fairness:

"If the administration is close to somebody, they don't have to sweat about it. You can have a lazy teacher if the administration's just going to give them good scores."

School leader participants shared their general perceptions of the fairness of IMPACT as an evaluation system in the survey. 31% of school leader participants Agreed or Strongly Agreed that IMPACT is a fair teacher evaluation system, whereas 38% Somewhat Agreed, 15% Somewhat disagreed, and 16% Disagreed or Strongly Disagreed. Some survey comments described either variation in fair implementation or the lack of fairness, for example:

"I think it's incredibly subjective and very high-stakes and can't be completely equitably across the district."

"If school leaders provide clear, objective evidence to support scores, it is fair. There is significant variance in the quality of feedback provided in IMPACT reports, suggesting it is not used objectively, but rather subjectively, and to meet compliance measures set by DCPS."

"The tool itself is somewhat fair; however, the implementation and scoring of IMPACT components, particularly EP observations, vary vastly across schools and even within schools based on the administrator."

In addition to these quotes, broadly, about validity and fairness above, our interview coding revealed other themes that related to participants' perceptions of validity and fairness. There were two themes that relate to validity: Subjectivity/Inconsistency of the Evaluation System and Manipulation/Gaming the System. There were also two themes that emerged on fairness: Perceived History as an Inequitable Evaluation System and Participant Questioning of Equity Outcomes. Our team also analyzed quantitative data to examine discrepancies in scores (and various IMPACT components) by Ward and teacher race, which are presented in the final fairness theme.

4.3.2 RQ3 Validity Theme: Subjectivity/Inconsistency of the Evaluation System

A strong majority of teacher interview participants perceived their evaluations to be highly subjective and/or biased, especially based on their relationship (and other teachers' relationships) with school leaders. When examining this theme across teacher identities and school characteristics, the theme was similarly present for teachers across races and across Wards. Several noted inconsistencies in their scores when transitioning from one school to another.

"It seems to be really subjective, just based on your relationship with the principal and the assistant principal and what kind of teacher they think you are, without spending a lot of time in your classroom."

"I think that a lot of the complaints or the trouble or the issue, when you boil it down to it, it's really tying into that favoritism piece."

"Oh, they have these same ratings that they should be the same." I don't think that's true. I know that's not true. It often feels very subjective across the district. As far as principals being able to do observations, even though there's the rubric, that people can go in and be like, "Well, I'm going to say that this is that way." That they can apply the rubric in the way that they want to based on what they see."

"Because when you believe you're doing the right thing, you're going to do the right thing, but I'm getting one result here and I got a different result somewhere else but I pretty much was doing what they have been doing."

"I think it depends a lot on how they're executed by both the teacher and the administrators."

Some teacher participants had very personal experiences where they felt bias and or subjectivity directly affected them.

"I know that some people feel like it's I got you, I'm going to IMPACT you out so to speak. I actually have had a coworker that was IMPACTed out. She and the principal had issues and she felt that it seemed to her that the principal was giving her a low score to purposely IMPACT her out so we say...I know-- that and I believe that there are some principals that can do that because we're human and at times we may not necessarily like someone or it could be that you might not think they're a good teacher, but I think sometimes it's more with feelings than per se with what they're really supposed to use it for."

"I don't know if that was sufficient because I've had...every school has been a different experience with leadership and what-- I'm me, I do the same thing, I'm me. ... Having that experience and having different leaders...Everybody has been different. It's like I don't know what I'm doing right and what I'm doing wrong."

Many Highly Effective teachers also expressed that they felt the IMPACT process was highly subjective and at times, biased:

“I don't know, because I still think there's a lot of variability in this system. I don't know if I would say that it's holding teachers accountable. Plus, there's just so much ways of working this system. I've been under three administrations in my time in DCPS.”

“There were a lot of people pushing back against that administrator for low scores in sixth and seventh grade. There was a claim that favoritism was being made with eighth grade. It seemed like he took that into consideration when he was scoring us at the end of the year because some of my other eighth grade colleagues were also getting lower scores than they had gotten earlier in the year, similar situation like high-quality instruction didn't change. We didn't really know what to make of that.”

“The observations are definitely skewed or biased towards whatever kind of classroom setting. I've noticed that in a resource class versus an inclusion class, they're looking for different things, but also when they come into my resource class, certain content administrators, especially, expect to see them grappling with content on the same level as kids in inclusion when that's not necessarily functionally or developmentally appropriate for them.”

“With the master educators at the beginning of when I first came into DCPS, I felt like it was very subjective. I had math master educators coming and observing me in English classrooms and giving me feedback on how to implement math practices in English.”

“Because I've had two different principals and several different IMPACT observers. I would say the one thing that is consistent is subjectivity. I don't think it's an objective tool. I think it depends on what lens you're looking through and whether or not to say some degree you have positive or negative feelings about the person you're observing.”

Additionally, school leader participants shared their perceptions about whether IMPACT is subjective in the survey. 29% of school leader participants said that teacher IMPACT was “Not at all” or “A little” subjective, whereas 44% said that it was “Moderately” subjective, and 27% said that it was “Mostly” or “Entirely” subjective. Many school leaders commented on this subjectivity in focus groups and in the survey, for example:

“It clarifies expectations for instructional practices. However, the tool is not implemented the same across the district, thus making results subjective.”

“It has clear expectations for teaching and learning. The challenge is the subjectivity in which administrators apply the criteria.”

“It is too subjective on what evidence you use and how much weight you give to that evidence.”

“IMPACT is absolutely subjective and can be subject to the administrator observing the teacher with whom they have a relationship or not. While I agree there should be standards and

measures we are all evaluated against, we have to be aware that there are many things that teachers do that are outside of the IMPACT system, and many things that can help and harm a teacher's ratings."

"It is completely subjective. It is subjective to the topic of the day, to the time of the year, to the time of day. It is also subjective to the 30 minutes I am there. I would love to see a whole class period from start to finish."

"It is too subjective. Five administrators could view the same lesson independently and rate it differently."

A minority of school leader survey comments described that IMPACT tries to minimize subjectivity, comparatively with other teacher evaluation systems:

"Any "evaluation" process can be subjective...but IMPACT is one of the most fair of the evaluation systems I've seen. Administrators and teachers can be subject to the whims and caprices of their supervisors, and in some ways, IMPACT may be used to settle scores or take out petty grievances. However, other evaluation systems in other states and districts are the same, or even worse with their subjectivity. IMPACT has the minimum amount of subjectivity, but it is still there."

"Almost all evaluations, other than completely blind ones, are subjective. IMPACT does provide a clear rubric against which teacher practice can be assessed, which adds objectivity and consistency."

"The criteria for evaluation are clear and therefore slightly less subjective than other evaluation systems and that is what makes it clear."

4.3.3 RQ3 Validity Theme: Manipulation/Gaming the System

A majority of teacher participants described that there is manipulation and "gaming the system" in order to influence the IMPACT score as desired. Manipulation was reported from both the perspective of the teacher "gaming the system" and the administration manipulating scores. The gaming/manipulation that teacher participants described ranged greatly, for example, from administrator favoritism playing into scores to teachers changing their practice just for the day of the observation to teachers changing students' scores.

Here are a few voices of teachers who perceive the tool is being manipulated by the administrator:

"If you can get data from anything, you can make the data say what you want it to say."

"Then DCPS kind of taught me, life is not all fair and equal...it kind of depends who's in that room, and who's sitting behind that computer, so you could think you have the most awesome

lesson in the world, but that person doesn't think so, they can manipulate those numbers, to think well, you didn't quite do this justice.”

Many teachers who were rated as Highly Effective in 2018-2019 concurred that there could be significant “gaming” by both teachers and administrators of the IMPACT evaluation process:

“Sometimes. I know that sometimes that admin can get pressure from their higher ups. I remember my first year teaching when I was under the IMPACT system. I did a really good job, and they said I did a really good job but the-- I can't remember. I think it was the instructional superintendent was saying, ‘How does this school have so many highly effective teachers, but the kids are not doing-- They're not making humongous gains in their test scores?’ So that weighed into my IMPACT, which I thought was very unfair. You said I did a good job. You recognized that I did a good job, but you you feel pressure that you have to knock me down a little bit because of what the instructional superintendent said.”

“If you can get data from anything, you can make the data say what you want it to say. In terms of having this many teachers, this is a good way to get data, ‘Oh, we've had this many effective teachers at this school, this many have left or these ones have never left.’ I'm sure of DCPS, it's a good tool to tout things about teachers, it's just data collection. I don't think it's a good measure.”

Some school leaders also described in their focus group or survey comments that administrators manipulate scores to the way they desire, for example:

“This process is still subjective and can be used for good or bad. Administrators have the opportunity to rate someone the way they like and use explicit verbiage to justify their reasoning.”

“It is very subjective as a school leader can be vindictive about their evaluation. The system leaves entirely too much gray area for a leader to abuse the process. If a teacher is doing extremely well, they can still receive a low score because it is a subjective system and does not promote teacher growth. The LIFT scale and incentives are not appropriate. This method leaves an opportunity for people to operate in an unprofessional manner.”

Similarly, teacher participants share how they themselves have learned to game the system:

“At the same time, I do know, there are ways to get around where I know there are some teachers whose-- there are some tests that can be manipulated, and so to speak.”

“Or teachers who doctor scores to make it look like their kids made a lot of progress when they didn't make a lot of progress.”

“When it comes down to it, she's the one who puts in all the scores and things like that. That's the one you got to kind of buddy up and be on good terms with, and it's a game like that”

“I'm driven to like beat the IMPACT system, if that makes sense.”

“I also think a lot of it's a show and I think that is the mentality that a lot of teachers have taken on in IMPACT, it almost reminds me of that movie Matilda when Ms. Trunchbull comes and they have to put away all their fun stuff and make sure that nobody ever saw that there was some learning that was interesting going on to the students, and it was this is the curriculum. Let's go back to this lecture-style way of teaching. That's what it reminds me of during IMPACT season...Again, it's a show, so because it's a show, you have to find all of the props that you need for a good performance, and you're not getting an accurate set of data.”

In interviews, teacher participants shared some ideas for improving the validity of IMPACT, including a desire for more objective observers, for observers with subject-matter expertise, and for more than one observer⁵.

“Yes, I don't know if they could find people like that. I'm sure that they exist. But feedback would be more powerful if you could actually find people committed to supporting teachers in such a targeted way. I don't think that administrators do a good job with that. I don't think that they're the best people for the job. One because they're overworked. Subject matter experts, I don't necessarily think that they're instructional experts all the time.Like my LEAP person, she had [been] teaching for well over 10 years in the context that I was teaching in. I think it was easier for her to do that job. I don't think that our evaluators have the tool that it's not necessarily their fault, but I don't think that they have the tools to give really good ... powerful feedback....I think that they tried that with Master educators. Again, I don't think that they were part of the same context. I think having [an evaluator] who knows the... unique context that you're working in again makes feedback better.”

“I probably would bring back the master educators. I just liked that it was actually a content person. Now it's like the, whatever, five master educators I had across the years, only one was terrible. I just think like they actually knew the content, which is more helpful in terms of improving my teaching. I learned things from those observations ...Plus, they're just an impartial observer. They don't want you to do well or poorly. They don't really care because they don't have a relationship with you versus I feel like with your principal, they want you to do well, they want you to be successful or at least I would think so...I really appreciated that aspect.”

“It would be your people who are doing your professional learning community, who are leading you in professional growth, professional development. Maybe they wouldn't be the ones who would do your observation and your IMPACT score. Maybe it would be like if I'm at school A and there was school B, but the same person leading in a professional community, they would come over, they wouldn't know me, they wouldn't know necessarily what my goals were and where I had started, so they couldn't penalize you for not meeting all your objectives, but that they were still working together...and that I'm being observed and give an informal feedback on the ways that we're talking about it in LEAP, then I'm being observed in the same exact way by a person who was trained to do that in that way, not an administrator who this is one more thing they have to do...”

⁵ Findings from the Spring 2021 Insight Survey also address this issue—results are available from DCPS.

“One thing that I was thinking of is that maybe there should be an IMPACT team that might come in and they would be highly trained. Then, the team would have a representative for each discipline area or each grade band... maybe that would add a level of neutrality to the use of it... For me, it would be that neutral IMPACT team. I think that that would be honestly for the win... I think, too, it would be easier frankly to handle [chuckle] complaints because then it would just be, “I wasn't given a fair score,” but then you can look and say, “Evaluator X, they've looked at this across 50 different teachers at the elementary level.” It would be very easy to pinpoint and then it would take that personal part out.”

“If there would be some way to do IMPACT so that you can get more than just one evaluators point of view. Because different people see different things differently.”

“I think that I would get people who really are experts in reading and math and have them be trained in at schools, not at the discretion of the principal. Then I would have them work really closely with teachers and be in classrooms, at least on a weekly basis to see the teachers in action set goals with the teachers and then work with them all year to see that those goals are met... I think I would take out the point system and I would take out the extra money and just really focus on teachers' growth, and then have something separate that administrators can use for teachers who routinely aren't performing at a high enough level.”

School leaders also shared a desire to include more objective observers in their survey comments, for example:

“Impact is fair because it measures a wide range of teacher effectiveness. When Admin is trained well on the IMPACT system, it is not a subjective process. I thought it was better when there were outside evaluators coming into the school to evaluate as it normed the process district wide.”

“As such, schools should have an outside evaluator administer TRC and DIBELS.”

“Teachers are at the disposal of the evaluator. Relationships and bias can play into this- good or bad. There has to be an objective, unbiased, third party who provides some layer of oversight in this process.”

“I think master educators made sense.”

“I think the format is good and in general should not be changed. I do think that the outside assessors were helpful in norming the scores across the district.”

“I'd like there to be an expectation for shared observation of instruction such as collegial learning walks.”

4.3.4 RQ3 Fairness Theme: Perceived History as An Inequitable Evaluation System

Given that IMPACT has been in implementation for more than 10 years, teacher and school leader participants who had been working in DCPS since IMPACT was initiated often commented on the history of IMPACT. Some participants who were around at the beginning of IMPACT saw its importance initially for exiting low performers, while others found it to be problematic in having harsh methods that were inequitable by ward and race. Several participants used the term “cleaning house”). We examined whether these perceptions differed by teacher race and ward and found that, while there was not a pattern of differences by ward, White teacher participants more often spoke of “cleaning house” than teacher participants of other races. Several teacher participants described that any benefits have waned over time.

Below are quotes of interview participants that demonstrate teacher perceptions of the early implementation of IMPACT:

“When IMPACT started, which a lot of people don't realize, it was a very, very vicious tool of the educational reform movement, and the number of Black teachers that were fired through that sys-- To me it's very problematic, it's very prejudicial, and it didn't help that the people who created it were not people of color, but they came and slammed this down into communities where students and families were already suffering...They took teachers away from them, sometimes in the middle of the year, took principals away in the middle of the year because it didn't fit into what Michelle Rhee wanted. There wasn't a, "Okay, well this school is doing poorly, so let's put a two-year plan in place and see how we can improve, let's do a needs assessment, let's get parent input, let's get stakeholder input, let's see what we can do so that this school doesn't close." It was like, "Okay, shut it down, shut it down"...

“However, as time progressed, I think IMPACT was a tool that, I call it a drive-by. A tool whereby it became punitive. I think it moved away from what it intentionally set out to do for teachers. It started being, "I got you. I got you too," where teachers just-- they became fearful of it, they resisted it for many, many reasons. The district did try to change a few of the components, but I didn't think it had the impact that it originally set out to have.”

“Especially East of the river, so many schools were shut down...I think with just how IMPACT was created, it definitely was a way-- because it wasn't just that they got those teachers of color out. It was that they replaced them with younger White teachers who were not from the area, who did not understand the culture of these students, and it was okay, and so because of its origins, it's always going to be problematic.”

Many teachers rated as Highly Effective in 2018-2019 concurred with the idea that IMPACT was useful in the early years to exit low performers:

“On the other hand, when I started in DCPS, there were a lot of teachers who weren't teaching even in my own school. Who were showing movies and the kids were doing coloring

sheets, and those types of things. I did feel initially that it helped get and retain better teachers and get rid of teachers who weren't actually teaching.”

“I think that initially, when Michelle Rhee brought this in, its initial purpose was to clean house and they succeeded. They fired a ton of teachers. Right now, the returns are so fleetingly diminishing that it's insane that they're still using it. Those are my initial impressions.”

“I have seen the good, the bad, the ugly and while I do not agree with Michelle Rhee's approach, how she handled the accessing of teachers and things like that, the teachers that needed to go, did go. That's just based on my interactions. It was like, “Wow, they did go.”

“I believe that teachers that were in an uproar were the ones they are the teachers that wanted to do bare minimum. I'm not speaking out of hearsay. I'm speaking what I have observed and so I'm appreciative of that. Michelle Rhee was not all the way wrong. I think her approach was a little bit harsh.”

School leaders discussed the history of IMPACT in focus groups, for example:

“I've been in DCPS [more than a decade] years now. Being a teacher and seeing really just a miscarriage of education, and schools, and people that were just-- I watched teachers that would get their kids open a line and march into the vending machine in the staff lounge that they owned so that all the kids could buy snacks, and they would bring them back and that was the lesson. Michelle Rhee came in and IMPACT came in and it was incredibly effective at saying, ‘Okay, if we're not going to teach, then that's it. You're not going to be in front of kids.’ It was great at that. It was a very good tool. They've done a lot of that work. Not saying that that work still isn't always necessary. As far as now in a new phase of education with well-intentioned people that still just need to refine skills who are dealing with all of the difficulties of high needs urban education, IMPACT isn't necessarily the tool for those folks. I don't think it's meeting that goal. There was a time and a place where IMPACT was so necessary.”

4.3.5 RQ3 Fairness Theme: Questioning of Equity Outcomes

Several teacher participants commented that IMPACT intends to create equal outcomes across ward and for all students in DC. However, some participants perceived that the differences in scores by race and inequities in student testing may reify inequities. They also discussed inequitable differences across wards in challenges students face and access to resources to support both students and teacher growth.

“Because if you choose to work in a certain area, then IMPACT could be the nail in your coffin, if you understand what I'm saying. If you work in an area, let's say, that's a school that you're provided with everything that you need to be successful, everything that you need to ensure that your students are proficient in every area, then, of course, IMPACT is going to be golden. However, if you're faced with schools that have less resources, then it's a little harder to be highly effective or effective in certain components.”

“In my particular school, we felt as though the teachers that are Caucasian teachers to have scored higher than the African-American teachers. It was very very unfair. We felt it...I just felt like it was used to pit the Blacks against the Whites... It was a real big problem yes.”

“I feel as if, although we play a very significant role in whether or not students achieve, to carry that much weight for our evaluations, to be that dependent on student achievement, considering all that goes into our work-- and I work in Ward 8 schools. There are days when your students can be on it, and they are excited and life is great. Then there are other days where the traumas that they face, the problems that they're dealing with, the gaps in their learning can pretty much throw them off course. That just because of the inconsistency with the children sometimes, the fact that our evaluations are so heavily focused on student testing poses a big problem for me.”

“Well, I know that I've had a few friends who taught in one ward and then they would move to the other, and they were like, “Well, my score increased significantly or it went down significantly,” but again, I'm the same teacher. I personally feel like teachers in ward eight are graded a lot harder. Wards seven and eight I honestly believe that. To me, I don't really understand why, but I'd be interested in some data to see why it is that the evaluation is so stringent...Whereas when you go to other schools, the kids may, especially in wealthier areas, to me it should be harder to get highly effective in wealthier areas because the school is already getting points for things that the kids come in knowing already. You should not be seen as a good school if all your kids came into kindergarten knowing how to read. You didn't do anything... We have a huge amount of homeless students in our school and in the ward, or not homeless, but home insecure to where they're bouncing around to different family members or community members' houses and don't have a stable environment. Those things matter, trauma matters. When your brain has stopped developing because it's trying to deal with this traumatic incident and you going through school never having gotten services for this incident...”

We also examined whether the codes on equity varied by teacher race. Proportionally, Black teacher participants were more likely to discuss equity concerns with IMPACT than other teachers in our sample. As one Black teacher participant stated:

“Some schools have lots of challenges, and if you're at a school that faces many, many challenges, it's going to be a lot harder for you to get the same score as someone who doesn't face those same challenges. There's definitely a discrepancy between people at different schools.”

Many teachers who were rated as Highly Effective in 2018-2019 were in agreement that the differences in scores by race and inequities in student testing may reify inequities.

“It's hard to know if it's really fair or if it's like, 'Wow, this teacher looks amazing because all the other teachers aren't as good in my school.' It's not that that teacher really is amazing. I don't know. Also, DCPS doesn't share that, a lot of information about how many highly effective teachers are at this school versus this school, because I think that would also highlight. There should be probably an even distribution. Where's the discrepancy amongst it?”

Also, just because I'm good at teaching like somewhere in Northwest doesn't mean I'm any good in Southeast because there are different ways of teaching and there's different communities and there's different, partially, family relationships and those kinds of things. Just because you're effective somewhere else doesn't mean you're going to be effective in a different place.”

“It's a good question. Is it fair across the district? That's what it's supposed to do...I do appreciate about it, it's supposed to use the same standards for every school, so that's great. In practice, I don't know if it's happening across schools. As far as all kinds of potential places for bias, I don't know either. It's hard for me to tell.”

“What I can say, though, is across the district, and I have the benefit of having worked at [two different DCPS schools]. I can see how IMPACT impacts those at the lower end and the upper end as far as schools are in the area. A large part of IMPACT is really where you work. If you work at a better school, you are prone to getting better scores, and if you work at a "worse" school, and I say worse in air quotes, because every school has its issues and most of the educators that I've encountered in my career are really, really good people who try very, very hard.”

“One of my biggest issues with IMPACT is, you're evaluated about observations, at best, once, at worst, three times a year. Even three times a year, at 180 school days, you're talking about once every two months? Once every three months maybe if they're staggered. That same student that might have been homeless the day before and feels comfortable being in my classroom, but they're hungry and they're tired and so their head's down, there's nothing in IMPACT that says, "We really appreciate that you've built that strong relationship with that student to get them to sit in that classroom and be there for you". It's just, "Student A wasn't paying attention.”

In Survey comments, some school leaders also questioned whether IMPACT was an equitable process, for example:

“Academic achievement data is also important but the disparities between high-performing schools and low-performing schools make for inherent advantages for certain DCPS staff members.”

“IMPACT makes no allowances for the realities that may exist in some classrooms, and the many factors that are out of the teachers' control. Teachers are all measured by the same standard.”

“It is effective for classroom-based practices but does not take into account all of the other inequities we battle at the school.”

Teacher participants shared several suggestions for how to improve fairness. Teachers suggested giving more resources to teachers at under-resourced schools, better understanding trauma-informed teaching, measuring closing the equity gap, and changing the name of IMPACT and redesigning to get away from historical inequities.

“The biggest thing about students is that students need different things. Different students need different things and finding ways to meet what students need and get it to them is a bit of an art. Some students need some things that just don't have a place in my 53-minute classroom. I found tutoring hours in the morning, three days a week 8:00 to 8:30, that worked. I think to meet the needs of students, we should have a little bit broader lens, I don't know and look at student growth. I think definitely my principal focuses a lot on closing the gap. I think that is an excellent goal.”

“People are wondering, maybe he needs special ed, he can't read...No, it's not that he can't read. It's just that he's witnessed violence every day of his life for the first six years and nobody's ever thought to address that. IMPACT doesn't factor that in. There's no how did you help these kids during their trauma? That's not on the rubric, which should be... Now I really like to focus on how to fix things, but it's hard to say because they want to create standardization, but children across the district, there is no standard. Who was the standard child across the district that you would use to measure how effective this teacher is? What are our standard practices that will work across the district to measure how effective this teacher is?”

“If you could wave that magic wand again, how would you make IMPACT most useful for students? Participant: “My heart is beating really fast giving that answer. I would make it where if a teacher is deemed highly effective, that those teachers are given to the students that need the most. If you're a highly effective teacher, then you should be able to help the students that actually need highly effective teachers. That's what I would do.”

“I think if we have a tool that helps students, there should be some piece where we're measuring closing the gap. Closing the gap shouldn't just be one strand of one CSC rubric, which is barely worth, I don't know, 10% or whatever, it's worth. It should be a major piece of the puzzle if that's our goal as a district.”

“I don't think that IMPACT should be changed to where only some schools are held to certain standards. I think we as a district should have overall vision of what student achievement should look like. When we think about the disparities and just the inequities in this city, making sure that we provide all schools with those resources, the manpower, the curriculums that are going to help students achieve those things, that's what I would like to advocate for.”

“I'm thinking that the name of the evaluation system should be changed just because of the connotation that it has”

“It's very interesting to me that DCPS is so married to this system, and I really hope that this analysis would maybe provide some explanation into that because it doesn't make sense.”

“So, because there's too many flaws, to me, it makes sense to just scrap it and get community input and start over again.”

One school leader also commented in the survey on the importance of measuring equity frameworks explicitly in the evaluation framework:

“Teacher evaluations are not supported by research to improve student performance. There is no culturally responsiveness embedded in the framework and its evidence in teaching and learning.”

In addition to the qualitative perceptions on equity outcomes, our team ran quantitative analyses, analyzing how scores on IMPACT overall and on various components of IMPACT varied by teacher and school characteristics, including teacher race and gender, school ward, Title I status, grades served, and subject areas. Through these analyses, we sought to determine if there are notable patterns of results that could indicate potential bias or unfairness in the system, either for or against particular groups of teachers or rating teachers serving students with greater needs more harshly, thus not accounting for differences in student needs and discouraging teachers from serving schools with greater needs. Because, as noted above, teachers of certain identities are differentially likely to also teach in particular schools (for instance, Black and Hispanic/Latino teachers are significantly more likely to teach in Title I schools and Black teachers are significantly more likely to teach in Wards 7 and 8), and teacher and school characteristics could be confounded, we also ran regressions to determine how IMPACT scores vary on these dimensions independently of (holding constant) the other dimensions, as well as variability in components of the Essential Practices rubric.

Tables 6-10 present average score by IMPACT component for these groups. Female teachers consistently score higher than male teachers, and White, Asian, and American Indian teachers consistently score higher than Black, Hispanic or Latino, or Other/Unknown race teachers (there were an insufficient number of Native Hawaiian or Pacific Islander teachers in the dataset to report results without risking breach of confidentiality), although teachers in the latter groups do receive higher average bonus amounts, likely because they are more likely to serve in high-needs schools eligible for higher bonuses. Notably, the gaps were greatest in Essential Practices, Core Professionalism, and Contributions to the School Community, and smaller in TAS and Value-added, and student survey responses were higher for Black female, American Indian female, and Other/Unknown race female teachers than other groups. We see similar patterns by ward, with lower scores overall in wards 5, 7 and 8 driven by differences in EP, CP, and CSC. Teachers in Title I schools generally scored lower than in non-Title I schools, and teachers in elementary schools scored slightly higher overall and on EP and in high schools on other components, with middle school teachers generally slightly lower than other levels in almost all categories. The regression in Table 11 shows that, even holding constant other characteristics, Black, Hispanic/Latino, and Other/Unknown race teachers, teachers in wards 5 and 8, and middle school teachers scored lower overall, and female teachers, and teachers in Ward 2 scored higher overall than other teachers. Finally, Table 12 shows differences in specific components of the Essential Practices rubric by teacher and school characteristic. These differences could indicate potential biased or coded language in the rubric or in observation protocols or in observers themselves. We also examined difference in individuals EPs to provide insight to the possibility of biased EP measures. The EP components are listed below for reference:

- EP 1: Cultivate a responsive learning community
- EP 2: Challenge students with rigorous content
- EP 3: Lead a well-planned, purposeful learning experience
- EP 4: Maximize student ownership of learning
- EP 5: Respond to evidence of student learning

Female teachers score higher on all EPs than male teachers, and Hispanic/Latino teachers and middle school teachers score lower than other teachers on all EP components of the rubric, on average. Black teachers score higher on cultivating a responsive learning community, though not statistically significantly so, but lower on components 2-5 on academic rigor, pedagogy, and assessment, indicating potential bias in the rubric or observations in which Black teachers are stereotypically perceived as more culturally competent or nurturing but not recognized for their contributions to academics.

4.3.6 RQ3 Summary of Findings

Broadly, teachers and school leaders were moderately positive about whether they perceive IMPACT as valid. However, subjectivity, bias, and gaming the system were all cited as threats to validity in teacher interviews, in school leader focus groups, and in the school leader survey. Teacher and school leader participants recommended the use of external, more objective, and subject-matter knowledgeable observers to reduce subjectivity and favoritism. Subjectivity and gaming were consistent themes for teachers across race, ward, and effectiveness rating as well as across school leaders. In terms of fairness, participants were somewhat less favorable. Some (particularly White teachers) felt that the subjectivity and favoritism were unfair. Other teacher participations (particularly Black teachers) articulated concerns for equity, particularly for under resourced schools.

4.4 Labor Market

Research Question 4: *To what extent does IMPACT relate to the pipeline to and through DCPS for teachers?*

In response to research question 4, our coding of interviews revealed a theme on the Effects of IMPACT on Teacher Retention/Attrition. School leaders also mentioned their perceptions of attrition in their comments on the survey. Additionally, findings from Fall 2020 Insight Survey include questions about teachers' perceptions of the extent to which the bonuses/additional salary affect their desire to stay in DCPS.

4.4.1 RQ4 Theme: Effects on Teacher Retention/Attrition

Teacher interview participants had mixed perspectives on how IMPACT relates to teacher retention and attrition. Those who were highly motivated by the monetary aspect of IMPACT stated that it helped to retain them. Others felt the negative culture and climate, including fear, would drive them away from DCPS, regardless of the financial incentives. Yet others were intrinsically motivated to be a teacher, so IMPACT did not change their staying in the district—that money wouldn't make them stay and stress wouldn't make them leave.

“I remember saying like, ‘This is going to make me better up until the moment when it makes me quit. It's going to burn me out.’”

“I am exhilarated by it. It's exciting to work with my kids and see them grow, learn and just become better people overall. IMPACT and I feel the politics behind it, and the number chasing, the-- what people say [favoritism], all of that just makes me, honestly, not want to be in DC anymore. I'm considering leaving DCPS. I've already decided to leave DCPS after this next school year.”

“Honestly, I would say yes. Because of the monetary reward on the other end of IMPACT. I have a goal that I want to reach within a certain amount of time and if I don't reach that goal... I do think that I could see myself finding another district to work in, but also with that if I do achieve that goal I could also see myself staying with DCPS because once I figured out the formula or how I can achieve a Highly Effective rating I would be more comfortable to stay.”

“I know I'm going to be a teacher, I'm going to stay and be a teacher. I think IMPACT is just reinforcing. I like DC and I want to stay here and IMPACT is making that possible. I know if that was to go away and if they suddenly drastically reduced teacher salaries, I may have to take a second guess and, I don't know, move to another city or consider some other alternative.”

Like the broader teacher perceptions, teachers who were rated as Highly Effective in 2018-2019 held varied perceptions on the way IMPACT related to teacher retention/attrition:

“Well, it's kind of made me want to stay in my current school with my current administration who really values me and gives me high scores. It makes me nervous about starting over with a new administration and what that might look like for me.”

“As long as I continue to get effective or highly effective, it makes me want to stay but we do have a few really talented substitute teachers at our school who left teaching full time for DCPS because of IMPACT, because they were given really low scores and then ended up leaving being teachers. Just a thought that that could happen to me, if a new administrator came in who didn't like me, who scored me much lower. That's something that I do think about.”

“The scores and the lift ladder make me want to stay because it's been a nice benefit for me.

“The first year, IMPACT definitely made me want to quit because I got terrible evaluations all year. It made me feel like, A, maybe I wasn't very good and B, no one cares that I just gave up a higher paying job with better benefits and shorter hours to come to this district. Nobody cares. They just want me out of here. I did think about quitting the first year, but no, it doesn't-- If I've survived this long with IMPACT and I've learned how to game it, then it's not going to change my career path right now.”

“Within the profession, it's not going to cause me to stop teaching. If I left DCPS, I could teach somewhere else.”

“Again, the district, it seems it just doesn't care sometimes and so people are like, "I'm tired of this, I'm going somewhere else." Even if I have to take a pay job, I'm going to a different district because I know they're going to treat me better.”

Participants on the teacher Insight survey also showed this variation in attitudes. According to the survey results, 7% would not stay in DCPS without the bonus structure and 41.6% said it contributes to their desire to stay but is not the main factor; whereas 39.4% of teacher participants said that it does not at all affect their desire to continue and 12% of teacher participants shared IMPACT has a slight or significant negative effect on their desire to continue teaching in DCPS.

School leaders largely discussed teacher attrition, as it related to the ability to exit less effective teachers. Several school leaders in both focus groups and survey comments listed this as one of the most effective/useful aspects of IMPACT for school leaders. Below are example quotes from school leaders:

“IMPACT allows me to recruit the best teachers (IMPACTIplus) and remove poor performing teachers (twice minimally, developing then minimally).”

“Moving out low performing teachers.”

“That I can use it to remove truly ineffective teachers”

“Identifies low-end outliers well. Poor discrimination between excellent and good teachers.”

“IMPACT allows for standards and high expectations and a way to dismiss teachers who should not be in the profession, it serves its purpose of allowing an avenue to ensure that ineffective teachers are no longer with DCPS.”

4.4.2 RQ4 Summary of Findings

Overall, teacher and school leader participants discussed two roles that IMPACT has played for teacher recruitment, retention and attrition patterns. First, several spoke about how IMPACT has the ability to exit low performing teachers. School leaders saw this as a particular asset to their work. Second, participants discussed that IMPACT has the possibility of retaining teachers through incentives, such as bonuses and the LIFT ladder. However, other teachers perceived that the high-stakes and anxiety producing environment may cause them (or others) to leave DCPS.

4.5 Perspectives on Specific IMPACT Components

Beyond using the teacher and school leader data to answer the four main research questions, the AU team also examined perspectives of teachers and school leaders about each of the IMPACT components individually.

4.5.1 Essential Practices

Overall, teacher interview participants held mainly positive perspectives about the Essential Practices. Participants described that among the specific components of IMPACT, EP was most tied to practice and improvement. There was more positivity about the rubric itself than the observers and the feedback process. There was concern about subjectivity in ratings and school leaders conducting the observations and being biased based on their relationships. There was concern that the feedback on the observations was not growth oriented.

Several teacher interview participants described that the EPs showed that the philosophical intent of IMPACT is positive, to evaluate teaching from a holistic perspective that tries to break down and measure the components of teaching.

“The way that IMPACT sets up its rubric so that it tries to quantify teaching, I think that in itself is admirable. I think that the fact that they don't admit the irony of trying to quantify teaching is problematic, but I think it's interesting that people are trying to parse out what the different components of teaching is. Teaching is complex, so any effort to try to understand teaching better is an interesting idea. They did a decent job of laying out some preliminary pieces of how you would parse and analyze the process of teaching. The rubric and the process is a cool idea.”

Several teacher participants shared that the IMPACT observation cycle provides a narrow snapshot of an educator's daily teaching and learning in a DCPS classroom. There are day-to-day issues that occur in classrooms that are not captured through IMPACT's observations and the instrument more broadly.

“The way they do the IMPACT observations. The way that they come in to observe you, it doesn't account for a lot of the daily issues that are happening classrooms, it's very cut, and dry. You come in, but they don't take into account the other issues that happen in the classroom that could affect an observation score.”

“I think the observations aren't very effective, because they're just so short 30 minutes out of my entire 180 days in the classroom just doesn't give a really clear picture of the type of teacher that I am. I think that the task, it just really depends on your class. I have a special education background, so I tend to get all of the special education students in my class who make a lot of progress for them, but it doesn't always show up on the way that the rubric is for that.”

Several teacher participants acknowledged the importance of the observations around the Essential Practices but also acknowledge that there is the potential that teachers could “perform” during an observation, therefore the observations may not give a clear picture of a teacher's day-to-day instruction:

“I think the Essential Practices is most effective just because it kind of outlines what- I'm hesitant to say good teaching- with what they expect to see in the classroom, it outlines that because I won't necessarily say it's good teaching because it comes in different forms, but it's an outline to go by. That's good to work with because you understand how you're being scored. The

test is I don't think as effective because in the arts room I have to find ways to evaluate students fairly and it's beyond just taking a test to me. For visual arts, it's beyond taking a test and saying, multiple-choice, whatever, because there are some students who are great at testing and memorizing but even with that, that doesn't say that you have mastered visual arts. I break it down by category, you do portfolio. Most years I've just done a portfolio but even with that, it's tricky because I'm trying to look at growth, but it only works if I have students for a whole semester versus in advisory."

"Well, the one thing that I do like in the Essential Practices is the time afforded for students, the inquiry piece, to ask questions and to be engaged. It takes the teaching part off of the teacher, it's more student focused. Growing up, it was more so you just do what your teacher says to. You do this, you check this off, you did this, check, check, check, whereas now, we're putting it on the students to be accountable for their learning."

"I think observations are most, I would also say important, effective is hard because it depends on who's observing you and how they do it. Some years that I felt that was my first three years I had a really, really strong administrator. He acted like a coach. He always had specific insights into what I was doing and why I was doing it and gave me meaningful feedback. It was, honestly, I think the most accurate representation of what I was doing as how am I interacting on a regular basis in the classroom of children. "

In general, there seemed to be agreement that there was some anxiety and performativity around the observations for the EP score. Teacher participants shared that it was just a snapshot, could be a bad day, not really tied in any way to longer-term goals, growth, professional development or coaching.

"I think having it connected as a professional learning. If you really try to make this as one, an evaluation system but also a way for professional growth. I think just somehow really connecting it to that. I know teachers are like, "I don't want people in my room all the time." If you're someone who really is trying to grow and to be better and to do good work, don't have it just be that one time and then never come back. Really work with teachers. Like, "Here's your evaluation," and then come back and do it again. Maybe it would be informal but you would still just have the opportunity to grow and to get ongoing feedback rather than just they're specific times a year. Maybe you could opt-in or opt-out or different things of how you want do it but let there be an incentive for that then."

"I think it also provides a really strong potential structure for improving student learning, but it's not being-- I think it's valuable to have a rubric and to be able to focus on components of that rubric and really drill down on something to get better at. The PDs that we have on this, we'll do one PD for like 40 minutes on one section of the rubric, and then the next time we have PD a month later, which is crazy, we do PD once a month. We do PD a month later and it'll be on something totally different, not even on the rubric. Then three months later, we'll come back to the rubric and it's a different section and there's never any follow-up about the first thing that we did. There's no accountability, if I went to the meeting and sat there for 45 minutes, nod my head and then left and never did anything about it, no one would even know."

Additionally, school leader participants shared a mostly positive general perception of whether the Essential Practices component of teacher IMPACT supports their teachers' professional growth: 54% Agree and Strongly Agree, whereas 27% Somewhat Agree, and 19% Somewhat Disagree (0% Disagree or Strongly Disagree). Many school leader participants shared their positive perceptions of the EPs in their survey comments:

“It is helpful to have a common language to discuss teaching quality. It ensures we observe all teachers regularly.”

“The categories for evaluation are sufficient. The language inside of the rubrics, however, might be better clarified through evidentiary drop-downs or check boxes. For example: EPI could have actual drop down choices to support the observational narrative provided. Each score would have corresponding drop-down boxes with aligned/expected evidence. This may help teachers feel that the rating is less subjective, although leaders already utilize the guidance provided in IMPACT guidebooks, which provides such examples.”

“The Essential Practices are clear and give a picture of teachers as a whole through observations, CSC, CP, etc.”

“EP does support teachers' professional growth. I believe that teachers take the feedback and mostly implements the feedback.”

“The EP's are centered on best practices in education. However, 3A and 3B can be a bit redundant with other areas of the rubric. If you are skillfully designing and facilitating the lesson, it will impact the other areas.”

“The Essential Practices are a clear and concise explanation of best practices that gives us a common language and standardizes expectations across the district.”

“EP observations are the most helpful aspect of IMPACT, however, this is not a truly normed process across schools.”

“EPs provide a framework to support teacher growth. Based on the skills identified within rubric (although not exhaustive), feedback can be grounded in the EPs.”

“Giving feedback to teachers identifying areas that need improvement based on the particular lesson that I observed. I must couple this with the knowledge that I have of the teacher's overall performance to ensure the teacher gets meaningful feedback.”

Some school leaders also shared negative perceptions of the EPs:

“The current EPs do not fully and accurately capture content that is seen in lessons. It does not take into account prior lessons, information that may be vital to the success of the current lesson being observed.”

“The Essential Practice rubric is extremely vague and leaves room for much interpretation by the school leader. I do not feel that it truly shows what administrators should be looking for in instruction and how a teacher should measure up their lessons to ensure that they are not only doing “good” teaching but also ways to improve their lessons based upon the rubric.”

In the school leader survey, participants were asked what aspect of IMPACT was most helpful to their role as leaders. A majority of participants shared that the EPs were most effective:

“IMPACT allows me to unpack teacher instruction and student performance against a specific rubric. The IMPACT rubric also helps structure debrief conversations with teachers. The IMPACT debrief session to me is the most impactful component of the observation, scoring, reporting, and debrief process.”

“Most helpful is the dedicated time to sit down and debrief a lesson. The use of a rubric helps to identify key components to focus on when discussing areas for improvement.”

“During IMPACT season, I’m allocated more time on teacher improvement. I’m given more space and time to meet and work with teachers to help their improvement.”

“What is most helpful are the teacher conferences after evaluations occur. It is an opportunity to engage teachers in meaningful conversations about their instruction and how strategies to support student success.”

“The definitions of the EPs is the most helpful and running through professional development and discussions on them helps staff.”

School leaders also offered some rather specific suggestions for improvement of the EPs:

“EP5B - shift language for 4 to 3, 3 to 2, 2 to 1 and add new language for 4 to include students supporting and extending for each other. This would follow the pattern for most other EPs where 1-3 are teacher driven and 4 is student driven. Students supporting one another and providing extensions happens in classrooms, even at the elementary level, during whole and small group conversations and should be named as a highly effective practice.”

“Improvements would be to revise rubric- the Teach 1-9 rubric used when IMPACT first started, allowed for better alignment to teacher practice and clearly lined areas of improvement. There is a lack of consistency across EPs. In most of them, when student voice is front and center that scores a level 4. However, this is not the case in EP5 and that EP rubric can and should be changed.”

4.5.2 Individual Value-Added (IVA)

Most teacher participants who discussed IVA in interviews were negative. Participants who were negative about IVA described unfairness based on ward, and lack of transparency about how the measure is created. The teacher who did describe IVA favorably during the interview praised the

measure as an objective measure that was focused on student learning but has not given permission for their quote to be used in the final report, therefore a favorable description of IVA is not included in this final report.

“I guess the testing formula of the IVA, if you're in the school for six years and you don't know what this thing contribute it's just foolish. It's not like a [inaudible 00:48:02] you can just find, which is just crazy that it can affect incentives, your bonuses and you don't know what it is. Elementary level too, fourth and fifth grade are the only grades that have it. What's my motivation if I'm the best third-grade ELA teacher and I get them all to a fifth-grade level? I have no motivation to go to fourth or fifth grade when it's extra element that is unknown can hinder me. It's only a negative to move up the fourth or fifth grade if you're successful, K1, second, third-grade teacher, which is more of an elementary only problem. Teachers, I mean, we've talked about like, "Are you going to go to fourth or fifth grade?" They go, "Hell, no. They still do IVA? Simple enough." [chuckles]”

“I said, I feel like I'm a pretty educated person. I don't understand it. It's like an algorithm. They don't tell you what the algorithm is. It's really confusing. It's like you're trying to be more transparent, but you're not really at all.”

“I feel as if, although we play a very significant role in whether or not students achieve, to carry that much weight for our evaluations, to be that dependent on student achievement, considering all that goes into our work-- and I work in Ward 8 schools. There are days when your students can be on it, and they are excited and life is great. Then there are other days where the traumas that they face, the problems that they're dealing with, the gaps in their learning can pretty much throw them off course. That just because of the inconsistency with the children sometimes, the fact that our evaluations are so heavily focused on student testing poses a big problem for me.”

“The IVA is a really difficult part to understand still. Partially, the large percentage of it, and then also just the fact that it's really hard to correlate to anything we do on a day-to-day basis in school.”

“I know that the scores, the student performance tied to it, that's a big sticking point, too. That's tough coming from a teaching environment like DC where there's not a homogenous population of knowledge. So, you're really dealing with a lot of variability. I think that piece, and especially with the poor situation now, not that we're factoring that in, but that's just going to create a gap. We're all going to have to live with this gap for the next 10 years, it's just a fact.”

“Not that much is value-added, and that's just proprietary formula for assessing the value that we added to our students [unintelligible 00:05:02] scores. Nobody knows how it works and nobody will tell us when we ask. [laughs] I'm not particularly skeptical on it, but that was notable to me when I was trying to understand it better for the benefit of my teachers when I was a staff developer.”

“Anecdotally, it did seem like teachers who I thought were better teachers generally had better value-added data. That's not exactly true. I don't know whether that part's fair. On the

other hand, I personally do feel like I appreciate having extrinsic objective measures, but just I don't know whether mathematically it's fair."

Additionally, school leader participants shared a mostly negative general perception of whether the IVA component of teacher IMPACT supports their teachers' professional growth: 11% Agree and Strongly Agree and 22% Somewhat Agree, whereas 27% Somewhat Disagree, 25% Disagree, and 14% Strongly Disagree. Many school leader participants shared their negative perceptions of IVA in their survey comments:

"IMPACT has very clearly outlined the essential practices expected of our teachers, and it holds them accountable for those practices. It also has some school-level flexibility through the CSC process, and some teacher-level flexibility with their goal-setting process. It is weakened by the inclusion of IVA for some subjects. Teachers try to avoid teaching in IVA courses (ex: English 1 and 2) and prefer to teach English 3, AP English, Calculus, and other subjects that are not covered by IVA. In addition, the high stakes of the IMPACT system reduces conferencing with teachers to discussions about points, and not about improvements to their teaching. I'm not sure how to fix this, but it is more dramatic here than in previous systems I've used as both a supervisor and a teacher."

"It has evolved to reward and penalize as appropriate. It captures teaching and learning actions observed via EP observations. It also captures students outcomes, professionalism, contributions outside of the classroom, which all is part of the teacher's role. However, the IVA component is inconsistent since all teachers do not have and testing changes each school year. Student achievement can be measured with TAS."

"The IVA component is variable, very unreliable from one year to the next and causes a great deal of anxiety in teachers leading to difficulty with retention in the testing grades."

"As I noted, IVA doesn't look at student growth as TAS does. We can see and encourage growth of all levels through TAS annas an educator that is more of a win than a standardized test normalized to White men in a school community of Black young people"

"EP model is a good framework. IVA is a blunt tool for a fine task and with such small sample sizes is subject to a range of validity concerns. IVA has also kept poor performing teachers in our school who had one year of outlying scores due to interventions outside class put in place."

"IVA does not accurately portray a teacher's value in the classroom."

"IVA does not support growth and is based on student test scores on PARCC. What happened the year before with kids' scores can impact teachers' IVA the next year and for too many teachers in my building, IVA does not reflect the quality of the instruction they received in the current year. Teachers are scoring 4s on EPs across the board and their IVA is pulling them down because last year the kids did well on PARCC too. On the other hand, teachers are scoring 2s on EPs and getting high IVAs because last year the kids bombed PARCC. Please eliminate IVA."

“IVA is a black box, so not helpful for students or teachers, and is inequitable since it only applies to few teachers.”

“IVA- One element from expectancy theory of motivation that I think is lacking here is instrumentality. It is hard for teachers to see if their efforts and work will actually result in the outcomes and rewards they desire. It's not transparent enough or quick enough for teachers to see those outcomes.”

“IVA should be connected to the Benchmark assessments, which are more aligned with what's being done in the classroom everyday. PARCC should not be a part of teacher's IMPACT scores, especially as the results reflect not just what one teacher has done, but what previous teachers have also done.”

“IVA also has a similar impact, teachers don't use the data to inform their instruction, as it comes after the fact. It is not always an accurate measure of student progress.”

“IVA and student surveys are inherently unequal since they only apply to certain teachers who already feel a lot of stress. IVA and TAS are both statistically questionable, which creates a lot of frustration and stress about them. From what I've read, IVA is only reliable at the sample size of a school, in which case the underlying tests could be much shorter.”

“Essential Practices definitely support professional growth. IVA doesn't.”

“Teachers focus more on their scores than what their scores mean using IVA”

“Decrease percentage for IVA in teachers in tested grades”

“Eliminate IVA since it causes undue stress, it is not timely since it is calculated after the school year, and it only burdens teachers in certain departments and grades.”

“IVA is another mechanism that does not show teacher growth and should be removed as it can be punitive to those who teach those specific testing areas. Perhaps there should be an incentive over a range of time for those teachers whose students continue to show growth academically.”

“Value Added - Is not a clear metric that provides teachers any usable information that will cause them to adjust practice.”

“As stated above, any link to PARCC should be removed. IVA should reflect students' growth from BOY benchmarks to EOY.”

“Evaluations of teachers in different groups (group 1 v. 2a v. 3, etc.) are not equal because the validity of student test scores is not equal across groups. For example, a 4th grade Math teacher may have consistently positive EP ratings but have a bad student performance year on PARCC, which has a penalizing effect on his final rating. Whereas a PE teacher, or even a 1st

grade teacher of reading - such a critical year - can have consistently high EP scores and have that carry their final rating. It is not fair across teacher groups.”

“The EPs, TAS, CSC, and CP. Those are the parts that make a school a school. Testing scores does not. It encourages teachers to teach to a test and not to their students. The standards are often more than we have weeks of schools, and for schools with specialize curriculum then to it is even harder. IVA should be removed for all.”

4.5.3 Teacher Assessed Student Achievement (TAS)

Teacher interview participants were mainly negative on the TAS component of IMPACT. Many participants described that these were not set by teachers, as intended, but set by school leaders. Some participants also described that these are not validated assessments.

“The problem with TAS is that all the rigor that the school district claims to value with IVA, and it's an incredible amount of rigor. I tried to look at the math. I got a hold of one of the White papers. It's like 50 pages long. I don't know that math and I've been teaching math for 25 years. That's not my thing. Nobody understands that math. They give us TAS, and we're supposed to create something equally valid, based on nothing. TAS is completely statistically invalid and shouldn't exist in its present form at all. There's such a wide range of what teachers use for TAS. It's fine as a thing, but if you're going to base whether or not I get 20,000 extra dollars or I get fired on TAS, you better make it valid. It's not valid. That's all there is to it. Why do you have an inclusion in there that's not valid, when you have claimed to have that commitment? It's clearly gone. It should not be in there.”

“The original idea was that teachers could establish a measure, like my kids are going to do X according to whatever measure the teacher decided, and it seemed very liberal... Of course, they don't trust us, so that's now been taken away. The TAS still exists, but our ability to choose the measure has gone away. Essentially, they have forced us into accepting other standardized measures or other computer-based measures that they see as less subject to subjectiveness. TAS used to be a guaranteed way of helping your score, but now it's yet another way of their demonstration of lack of faith in us. Anyways, for me, that's how my IMPACT goes, other teachers can have other measures.”

“The first thing I would do is get rid of TAS. I think TAS is a huge waste of time. What TAS is set out to do is great. I should be getting my students to grow by the end of the year by giving them a pre-test in the beginning of the year, and then a post-test at the end, or a test at the end. Their score has improved? To me that says that they've grown and so I should get credit for that. However, the way TAS is set up, I set the parameters, I grade all the papers, and I give you all the data. What most people do, and I'm sure I've done this, lots of people I've talked to do this, we make the numbers work.”

“Teacher Assessment, something, something. That's supposed to be the teacher-- That was originally built as like, the teachers decide how they want to measure their student's progress and then they submit those choices in the beginning of the year. It's another theoretically

quantitative measure, but the teachers have a lot of control over it. However, at both of the schools that I've been, the administration has ended up determining that for teachers."

"Yes, that's the stupidest one of all, that's actually the only category that I think is completely useless. That's the one where you say that your kids are improved by a certain percentage on a test that you create...That's so meaningless. We know all the questions on the test, so I could totally just put a question from that test on the do now every day for the whole year and the kids would have seen every single question on the test 100 times and they all get 100 and that makes up, I think it's like 25%."

"Goal setting with TAS. We don't really decide our goals so much because they're trying to prescribed by the district. We have two choices, we can either choose mastery or growth, and my principal chooses growth, which is fine, so I'm good with it. I don't feel like we're really involved with that at all. It's just like, "This is your choice, go ahead and do this one."

Additionally, school leader participants shared a varied perceptions of whether the TAS component of teacher IMPACT supports their teachers' professional growth: 29% Agree and Strongly Agree and 25% Somewhat Agree, 21% Somewhat Disagree, 26% Disagree and Strongly Disagree. Many school leader participants shared varied perceptions of TAS in their survey comments. Although the survey showed varied perspectives, those school leaders who commented on TAS only shared negative perceptions, similar to the teacher participants:

"Some components of IMPACT are effective. For example, the observation tool is an effective way to capture snapshots of what is happening in a classroom. There are other components that are not as effective. For example, TAS has not been used successfully and has no place in the evaluation system. This component has caused students to be segregated. Conversations around students, "I don't want that student. They will screw up my data. I only want AP students." This is not how our teachers should be discussing our students. In addition, students who have IEPs lose out on having a strong teacher because of TAS."

"TAS does require teachers to look at and think about data, and that practice supports teachers' professional growth, but not the TAS scores themselves. IVA and TAS are both statistically questionable"

"TAS does not support teachers' growth- it puts fear in them, worrying if students don't excel on an assessment that it will undo their evaluations. It does not allow for teachers to reflect on the work they are doing, but make them worry about their scores and not use data to inform instruction"

"TAS is a truly flawed portion of the evaluation as teachers can totally manipulate this data to get higher scores. It does not in no way show how teacher practice has improved but only how a student might have improved on a few assignments throughout the school year."

"TAS and CSC - Are seen as metrics or bars to jump over but do not make better professionals"

“Get rid of TAS, it is a joke.”

School leader participants also shared a few suggestions for improving TAS in their survey comments:

“TAS could be more helpful if there were more standardized measures across content and subject areas.”

“TAS needs better framing. Since I've been here it has not been a robust part of the evaluation process, and during training it is typically ignored. It seems like free "points" for teachers.”

4.5.4 Commitment to School Community (CSC)

There were mixed teacher perceptions on the CSC in interviews. Of the teachers who discussed CSC, those that held trusting relationships with their school leaders were more positive, whereas those with less trusting relationships were more negative on CSC. One overarching comment was that CSC required teachers to participate in after-hours, often non-contractual duties activities in order to receive a strong score. Additionally, many participants mentioned how CSC is very subjective and implemented very differently across schools.

“The CSCs, the commitment to school and community is very subjective. There are some of our teachers. Like we have a special-ed point in our CSC. Some of our specials teachers aren't even give an IEPs. So how can you mark them low on something they don't even get. You know what I mean? It's not as objective like they would like it to be, because it doesn't fit a majority of the teachers. I can bet every teacher was like, "This doesn't apply to me. This doesn't apply to me. This doesn't apply to me.”

“Let me tell you something. The CSC is pretty much the Wild West of the IMPACT zone. The reason being, every single school's CSC is completely different. Some of them have a Bible full of things to do to get a three.”

“When I thought that CSC was this completely insane, arbitrary, absurd set of demands that came out of a place of distrust between teachers, administrators, it did not make me better. It paralyzed me and terrified me.”

“Then there's a section called contribution to school and community. That is-- gosh, what is it? I can't remember, 50% or 10%. No, it's 10%. That's the means by which the district extorts unpaid labor from us, because you can't just do your job and get a four, and everyone wants to get a four. You have to do something above and beyond. They give you a rubric that they say you should be able to attain a four within the school day, within your normal paid hours. That's simply not true. You have to do other stuff, you have to run committees or have an after-school program or plan a parent night. You have to do it twice a year.”

“I think what to me has been most unfair was the CSC. The CSC rubric for our school is 15 pages long. It's incredibly confusing. The rubric is just overwhelming. My first year, I spent hours trying to get all this documentation to prove that I was committed to my school. Whereas some other schools, they just had like a simple list or like a checklist or you could just say, "I did this and I did that." I think that needs to be streamlined and then unified across different schools because it's not fair that I have to spend so much time and so much energy and do so many more extra meetings and extra commitments just to get that one thing. I think the core professionalism maybe should count for something more. I think you only get docked if you don't do it. I don't know why we don't get any credit for showing up on time all the time.”

“I know our CSC is really bad, is really, really difficult. We had to do 40 home visits in order to get the highly effective, which was absurd to me. It's just a lot. That doesn't seem fair.”

By contrast, school leader participants shared a more positive general perception of whether the CSC component of teacher IMPACT supports their teachers' professional growth: 35% Agree and Strongly Agree and 35% Somewhat Agree, whereas 10% Somewhat Disagree and 22% Disagree and Strongly Disagree. School leader comments CSC were quite varied. Some were positive:

“CSC, I think is great! It allows school specific rubrics of what matters to their community and is a clear way of developing teacher leadership and building a strong community where everyone contributes their best to ensure our students are experiencing a wonderful school.”

“The CSC component lifts up that which educational leaders envision as strong professional behaviors and contributions to the school community.”

“In [other state], addressing school culture and teacher core professionalism were things that had to be handled through a lengthy disciplinary process, where even the best evidence could be thrown away due to the slightest technicality. While that is also somewhat true in DCPS, having CP and CSC to help shape culture and behavior is astounding! I was completely shocked and happy to see that there were levers and methods to help shape behavior and culture. The best part is that we can be very clear about what matters, and with SCAC/Union, create a vision for the ideal behavior of an effective employee at our school. Once we agree to the CSC and share it with teachers, every staff member has a clear idea of our ideal environment. These are critical conversations that you don't always have in other districts.”

“The EPs and CSC are very helpful when providing feedback to staff about their performance and justify the need for certain school initiatives.”

Other school leaders shared more negative views of CSC in their comments:

“It is effective to hold teachers accountable [to] have a compliance mindset for their work. It is a stressor that does not support high performing teachers. It's a convoluted system that often requires a likely of inputs that do not directly impact student growth and performance. For example, CSC usually involves teachers spending hours gathering documentation of their work, hours I wish were spent supporting students instead.”

“CSC is far more time and effort than it is worth and only marginally changes teacher practice - a single observation, PD that may or may not be relevant. The best part is that it collectively adds to student culture through clubs, activities, fieldtrips that we include in the rubric.”

“CSC and CP seem to be compliance areas more than indicators of high standards for teaching and learning, but CP is more important for addressing major employee behaviors. CSC seems to be less useful from a student impact perspective than TAS.”

“TAS and CSC - Are seen as metrics or bars to jump over but do not make better professionals”

4.5.5 Student Survey

Many teacher participants expressed negative perceptions about the Student Survey, but there were also positive opinions about their value. When asked what they would like to change about IMPACT or which component was least effective, many of the participants focused specifically on eliminating or significantly altering the process for the student survey. Therefore, there was not consensus about the student surveys from the teachers who we interviewed.

Several teacher participants described negative views, linked to validity issues with student surveys or unfairness by position:

“Eliminate those. That's ridiculous. They're asking third graders their opinion of their teacher and then grading them with an algorithm that only people at MIT or Harvard or whoever we farm this out to have seen. I don't know how those scores work. I do know that if four students in my class give me a one, I get a one. Ridiculous. I have four students who want to give me bad grades. Of course, they do. These are children who have difficulty learning, and my job is to make them do things they don't want to do. They're going to react as children do, and I can take it because I'm a professional. However, if four kids who are angry at me give me a one, then I'm sunk. Eliminate that, because it's statistically ridiculous. It makes no sense.”

“I've gone through the survey and third graders don't understand what a lot of things mean on the survey, and so they're left up to their own, so they click, click, click, click, and you just hope cross your fingers like, okay, I hope they [chuckles] actually click what they meant to click. I remember a couple years ago a third-grade class I got it at 2.88. I'm like, are you serious? I remember being in the classroom every- what does this mean? I don't understand this. What are they asking me about?”

“What I would steer away from is the SSP, the students feedback...it's unannounced, and they pick the class for you, but say for instance, if you just gave a test which your minority of the class did not do well or did not study for it, and now they're about to-- That happened to me in both my classes. Now they're about to take this-- Your evaluations are really low. Of course, you have the opportunity to have two chances where you can break it up in one semester and then on the

second, but it still is a little biased because as a teacher, I'm just like, "Wow, I got really low scores in this class, and me and this class, we have a good time," but it was just happened to be the time of day or the time that what we were doing in that class [unintelligible 00:09:46] giving them poor scores on SSP. I hate to complain without giving a resolution of how it could be improved better, but I would say maybe we could increase the number from two to four student evaluations, because out of two, it's a 50/50 chance of your average. Both scores have 50% weight. I would say maybe increase it to four SSPs."

"I have eight-year-olds taking a survey that's like mostly agree, slightly agree. It's their first time taking an online survey that's like this at all and we're all like used to doing them for \$5-card Amazon gift cards. It's their first time ever doing a survey, and they're eight years old. They can't navigate a computer. They're doing the survey and that's 10% that ties into my overall score and it's my fifth-grade class or third-grade class, it doesn't matter, like I've had classes-- All my students love me. I'm very confident. I'm a good teacher in a good environment and my kids are, "Mr. [], we did that computer thing for you, great." I'm like, "Okay, cool. How'd it go?" They go, "It was fun, whatever." Then I still get a 2.2, or I have like, "Hey, Mr. [], you got good grades from us," and I'm like, "These are 10-year-olds." I can imagine what middle school and high school looks like because these kids know exactly what they're doing, whether they love them or not, catch them on a bad day...I think I would really love the data, but I just see it in such a negative light because it's tied to this overall evaluation...I wouldn't do it to the point where an eight-year-old can decide my end of year performance technology."

"Student survey, there's a couple problems that I have with it. First of all, if you're in the upper grades, it's 10%, which is the same percentage as one of your observations from your principal. That means mathematically that the opinion of a fourth grader on maybe a bad day can matter as much as what your principal sees from a lesson. That just seems insane to me. I think you can be informative, what your students think of you and I think it's important to constantly be evaluating that, but to make it a percentage of your final evaluation when they're taking just a survey on a computer which they're not-- I don't know how seriously they even take it. It's really hard to stomach."

"The student survey, I'm sorry it's crap. I teach middle school. One day they might like you, one day they may not. If they get the opportunity they are not going to be objective. They are going to hate you and they are going to write down everything. That was a big thing like in my-- I've heard kids say, "I don't like him, I gave him all ones." Well, you didn't read the question. You just went on what you liked. That's my big problem."

Some teachers shared more positive opinions:

*"I think the student surveys are very effective because I think it's important that students feel comfortable or if they have issues you deal with it. This is not something that a student says about me so now I'm mad, no. For me, I'm looking at it as, "Okay, Miss ***, this is a area that you want to work on," that's the student surveys."*

"One of the good things about the IMPACT for students, is the student survey. That I would keep. Because when I look at that student survey, I know if I'm reaching my students or not."

Several teacher participants shared some suggestions for ensuring students better understood the survey, for example:

“That needs to be simplified. I like that they have audio available so kids can listen to the questions on headphones but the reading comprehension and attention span-- The kids do not want to read that many questions on a survey and it's very confusing because it's like, "Does your teacher help you out?" Obviously, you want to say yes all the time, versus like, "Does your teacher yell at you when you're bad?" Then they should be saying no, my teacher doesn't yell at me. I don't think that whole administration process of the surveys, I don't think they're-- Make the question shorter or make the survey shorter.”

“...everyone doesn't work with the same type of student, the same demographic. Let's say students in my area, students at my school on average, maybe 31% of the school is considered proficient readers. They're considered proficient readers. However, you have a school across town where it's 90%, these kids can read on a grade level, et cetera. These students are given a survey to survey their teachers and they don't necessarily understand what's being asked. I had an issue with the 'how' and if we are accommodating students and modifying assignments and they receive special accommodations, then why wasn't that done for the survey”

“I do worry about the student surveys because I know we have kids at different linguistic levels, and sometimes, there's a lot of language around those questions. If there is a chance to hear the question, spoken with audio, that helps, I think, to increase effectiveness. If there is not audio with those questions, then that would, of course, lower it, or if there were even picture support. I'm an ESL teacher by training. I think of these things when I think of kids who have different linguistics. If their answers and something like that are so high stakes, then we want to make sure that they really understand the question, and they're answering the question accurately.”

“With the student surveys, I would make sure that all the comprehensible supports are in place to make sure the kids understand the questions fully. I would probably decrease the number of questions because I think after a certain number of questions kids get tired and then they maybe don't put as much mental energy into understanding the questions. I think it's important to have the student's feedback on things. Now, there is the double edged sword to that so if you're a Professor McGonagall type who's very strict, you may not be as popular with the students as a teacher who is maybe--Anyway, the questions have to be designed carefully for that.”

By contrast, school leader participants shared a more positive general perception of whether the Student Survey component of teacher IMPACT supports their teachers' professional growth: 35% Agree and Strongly Agree and 32% Somewhat Agree, whereas 17% Somewhat Disagree and 16% Disagree and Strongly Disagree. School leaders offered only a few comments on the student survey, for example:

“The survey is informative but should not be counted toward a teacher's IMPACT, particularly in elementary school when kids are young, sometimes fickle, and sometimes don't understand the questions they are being asked.”

“Student Survey data does not help a teacher grow. It is a way in which a student has an opportunity to bash a teacher because they do not like them. Or, students do not take this survey seriously so they choose whatever just to complete it. This should be revamped to capture more”

“Student surveys are a great way to get valuable info about the student experience to inform teacher practice.”

4.5.6 Core Professionalism (CP)

Few teacher interview participants discussed Core Professionalism, but those who did mentioned that this measure seems to be more of a floor rather than a ceiling for teacher behavior.

“Another aspect of scores, core professionalism. Again, that's no challenge for me because you show up on time and you do what they say and that's core professionalism. I don't get how anybody ever loses the point on core professionalism. It's like you showed up. I just don't get that one.”

“The last component for my score is-- I don't even remember what it's called, but it can only be taken away if you're disrespectful. CORE professionals ... I'm like, "What is that?" That aspect is, if you were disrespectful to another child, or a student, or a family member, another staff person, or if you're late a lot, or have an unexcused absence or something like that.”

School leader participants reported a mainly positive general perception of whether the CP component of teacher IMPACT supports their teachers' professional growth: 37% Agree and Strongly Agree and 35% Somewhat Agree, whereas 16% Somewhat Disagree and 13% Disagree and Strongly Disagree. School leader comments on CP resonated with teacher perspectives of keeping the bare minimum expectations, but suggested that CP did not support growth:

“The CP areas norm and highlight professional attributes that should be expected from all stakeholders and holds our teachers/staff accountable.”

“Core professionalism promotes adherence to fundamentals of professionalism, but does not support growth.”

“There aren't enough clear benchmarks for each category. This is the most compliance-based section, but doesn't have clear indicators for all parts.”

“The core professionalism does not really support professional growth and is often used as part of progressive discipline.”

“CSC and CP seem to be compliance areas more than indicators of high standards for teaching and learning, but CP is more important for addressing major employee behaviors.”

“We need this tool to ensure people show up to work on time and complete core job functions CSC and CP (not twice per year) is helpful.”

5 Limitations

Our sample represents a broad swath of DCPS teachers across a diverse range of identities and experiences (see introduction for more information). However, our sample does have limitations. Black male teachers, first-year teachers, and ELA teachers are all slightly under-represented. This is an important limitation, and we recommend additional study on these populations of teachers to ensure their voices are thoroughly considered when implementing improvements to IMPACT. Additionally, our participants (on average) had slightly more negative perceptions of IMPACT, as measured on the DCPS Insight survey, than the average teacher who responded to the survey. Interview research and survey research may be prone to self-selection bias.

Beyond sample, an additional limitation is the timeframe. Spring 2020 was not a typical time period for DCPS teachers given the timing of the COVID-19 pandemic. The timing of the recruitment for teachers coincided with school closures, a sudden transition to online learning, and changes in the 2020 IMPACT process. Although we asked participants to focus on a typical IMPACT cycle, rather than the COVID timeline, the timeframe may have influenced what teachers shared during their interviews with us.

6 Conclusion

This report detailed the findings of DCPS teacher and school leader perspectives of the IMPACT teacher evaluation system. Overall, general perceptions of IMPACT among teachers interviewed were more negative than positive, although perceptions varied. Likewise, findings from the Insight survey, representing more than 70% of DCPS teachers were slightly negative, on average, about IMPACT. Teacher interviews reveal that the negativity stems from the high-stakes evaluation environment that produces mistrust, fear, and competitiveness in schools. Many school leaders in both focus groups and survey revealed that they concurred with teachers about the concerns about school climate and trust.

One additional finding that raises important questions for the future of IMPACT is that neither teachers nor school leaders in our study perceived, broadly (although there were exceptions in each group), that IMPACT was meeting its goal (and full potential) as a system to support teacher growth. The quantitative findings did indicate that modest gains in EP scores happen across the school year. However, qualitative findings showed that there is room for growth in

the alignment between IMPACT and professional development opportunities. Several participants (both teachers and school leaders) commented that having an authentic professional development system may be impossible in a high-stakes evaluation system. This may lead to the question about whether it is possible to have one system that effectively accomplishes two of the stated mechanisms of IMPACT: to transition out low performing teachers and support teacher growth.

Several participants also commented on the different life cycles of districts and how DCPS may have needed a system that focused more on transitioning out low performers 10 years ago (high-stakes), but now DCPS may be ready for a system that *elevates the highest performers*. In other words, many participants suggested that perhaps the initial IMPACT system focused on raising the floor of teaching rather than breaking through the ceiling of high quality (which may require a more formative, trusting, and nuanced approach). Some school leaders were, however, still favorable about the ability of IMPACT to exit low performers in focus groups and in surveys.

By contrast, teachers and school leaders, on the whole, were slightly positive about their perceptions of the validity of IMPACT in surveys. Yet, the qualitative findings from both teachers and school leaders revealed concerns with subjectivity and gaming the system. Many school leaders were also very concerned with their perceptions of how subjective IMPACT is, given it is tied to teacher livelihood.

Both teachers and school leaders were less favorable about whether IMPACT was a fair system, citing both concerns with favoritism as well equity concerns for under-resourced schools. Many participants lauded the intentions of IMPACT for fairness, but described that in implementation and in practice there were several equity and fairness concerns. There were some teachers and school leaders that cited concerns that teachers received lower scores in less resourced schools. Concerns over the inequities in scores across schools were particularly cited by Black teacher participants.

In terms of specific components of IMPACT, there was, largely, agreement by both teachers and school leaders that the EPs were the most growth-oriented component of IMPACT. Likewise, there was by and large agreement by both teacher and school leader participants that they were concerned that IVA was biased and unfair. There were significant concerns with the validity of the student survey and TAS; and concerns about subjectivity and inconsistent implementation with CSC. Both teachers and (particularly) school leaders seemed to find CP to be a useful bar for performance, but not for growth. Participants across both stakeholders described appreciation for a multiple measures approach to evaluation.

Throughout the interviews, focus groups, and surveys, participants made several important recommendations to improve IMPACT, including:

1. Improving the way that feedback is given, for example, by having pre-observation conferences, setting specific goals, and focusing feedback on specific goals tied to subject-matter.
2. Multiple low stakes observations that are more closely tied to coaching.
3. Reducing the connection between IMPACT and high-stakes monetary incentives;

4. Considering external evaluators, rotating evaluators, evaluators with subject-matter expertise, and multiple evaluators;
5. Increasing formative professional development opportunities;
6. Greater depth in the ways teachers learn about IMPACT (e.g., orientation);
7. Improved training of administrators on implementing IMPACT, including more norming
8. Greater trust and autonomy for teachers and more inclusion of teachers' voices in the evaluation process;
9. Including trauma-informed and culturally relevant teaching in IMPACT measures;
10. Giving more resources to teachers at under-resourced schools;
11. Improving the observation process by providing more flexibility and more transparency;
12. Measuring closing the equity gap;
13. Participants, especially school leaders, recommended eliminating IVA as a means to reducing inequalities.⁶
14. Changing the name of IMPACT and redesigning to get away from historical inequities.

In sum, perceptions of IMPACT matter. They matter to school culture and climate, to motivation and buy-in, they are related to retention and job satisfaction, and they matter because they exemplify the lived experiences of teachers and school leaders in DCPS. Teachers and school leaders expressed appreciation for their voices being heard and considered in the upcoming and future evolutions of IMPACT.

⁶ Per DCPS analyses, the IVA has more equal outcomes by Ward and Title I than other components, whereas there are more disparate outcomes in the EP component.

7 References

- Adnot, Melinda, Dee, Thomas, Katz, Veronica & Wyckoff, James. (2017). Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1), 54. Retrieved from <http://dx.doi.org/10.3102/0162373716663646>.
- Cohen J & Goldhaber D. (2016). Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educational Researcher*, 45(6):378-387.
- Chetty R, Friedman JN & Rockoff JE. (2013). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *NBER Working Paper*. 19424
- Dee, T.S. & Wyckoff, J. (2015), Incentives, Selection, and Teacher Performance: Evidence from IMPACT. *Journal Policy Analysis Management*, 34: 267-297.
- Dee, Thomas S & Wyckoff, James. (2017). A Lasting Impact. *Education Next*, 17(4), 58. Retrieved from <http://educationnext.org/journal/fall-2017-vol-17-no-4/>.
- Donaldson ML. (2009) So long, Lake Wobegon? Using teacher evaluation to raise teacher quality. *Rep., Cent. Am. Prog.*, Washington, DC
- Ho, A. D. & Kane, T. J. (2013). The reliability of classroom observations by school personnel. Seattle, WA: Bill & Melinda Gates Foundation.
- Howell, D., Norris, A., & Williams, K. L. (2019). Towards Black gaze theory: How Black female teachers make Black students visible. *Urban Education Research & Policy Annuals*, 6(1).
- Jackson CK. (2013). Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina. *NBER Working Paper*. 18624
- Jackson, C.K., Rockoff, J., & Staiger, D. (2014). Teacher effects and teacher-related policies. *Annual Review of Economics*, 6(25), 801-825.
- Jiang, Jennie Y & Spote, Susan E. (2016). Teacher Evaluation in Chicago: Differences in Observation and Value-Added Scores by Teacher, Student, and School Characteristics. Research Report. *University of Chicago Consortium on School Research*, Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=eric3&NEWS=N&AN=ED589723>
- Kraft, Matthew A., & Alvin Christian. (2021). Can Teacher Evaluation Systems Produce High-Quality Feedback? An Administrator Training Field Experiment. *EdWorkingPaper*: 19-62. Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/ydke-mt05>

- Lazarev, V. & Newman, D. (2015). How teacher evaluation is affected by class characteristics: Are observations biased? Paper presented at Association for Education Finance and Policy annual meeting, San Antonio, TX.
- Lindsay, C. A., & Hart, C. M. (2017). Teacher race and school discipline: Are students suspended less often when they have a teacher of the same race?. *Education Next*, 17(1), 72-79.
- Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education*. Revised and Expanded from "Case Study Research in Education." Jossey-Bass Publishers, 350 Sansome St, San Francisco, CA 94104.
- Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology*, 52(2), 250.
- Morse, J. M., Barrett, M., Mayan, M., Olson, K., & Spiers, J. (2002). Verification strategies for establishing reliability and validity in qualitative research. *International Journal of Qualitative Methods*, 1(2), 13-22
- Neal D. (2012). The design of performance pay in education. *In Handbook of the Economics of Education*, Vol. 4, ed. EA Hanushek, S Machin, L Woessmann, pp. 495–550. Amsterdam: North-Holland.
- Sporte, Susan E & Jiang, Jennie Y. (2016). Teacher Evaluation in Practice: Year 3 Teacher and Administrator Perceptions of REACH. Research Brief. *University of Chicago Consortium on School Research*, Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=eric3&NEWS=N&AN=ED568433>
- Taylor E.S. & Tyler J.H. (2012). The effect of evaluation on teacher performance. *American Economics Review*, 102:3628–51
- The New Teacher Project. (2021). *About Insight. Instructional Culture Insight*. <https://tntp.org/teacher-talent-toolbox/insight-survey>

8 Appendix A

AU SOE—DCPS IMPACT Evaluation

Research Questions

The research questions fall into four main areas: the teacher experience of IMPACT, professional growth, validity, and labor market.

Teacher Experience of IMPACT

Overarching research questions:

How do DCPS teachers and school leaders perceive IMPACT as a feedback, evaluation, accountability, and incentive system? What do they perceive could be improved and how?

Professional Growth

Overarching research questions:

How does IMPACT facilitate DCPS teacher improvements? How can IMPACT be altered to better support teacher improvement?

Validity and Fairness

Overarching research question:

To what extent can the validity and fairness of IMPACT be improved, and if so how?

Labor Market

Overarching research question:

To what extent does IMPACT relate to the pipeline to and through DCPS for teachers?

9 Appendix B

Interview Protocol

Thank you for participating in this study. We expect this interview to take 45-90 minutes.

This interview will be audio recorded and transcribed to provide more accurate data analysis. After initial transcription, researchers will provide you an opportunity to check for clarity.

At any time you may choose to no longer participate.

As a reminder from the consent form, your participation is completely voluntary and confidential. DCPS will not be made aware of your participation in this study and your participation will not have any consequences or benefits for your employment in any way. DCPS will, however receive results from the study for the purpose of improving the IMPACT evaluation system. Any results given to DCPS or published will be reported either in aggregate or in a way that no individuals can be identified from the results. You will have an opportunity to review any final reports or results prior to having them sent to DCPS.

After the interview you will be asked to complete a brief demographic questionnaire.

Do you wish to continue?

General Perceptions and Understanding of IMPACT

1. Can you start by telling me your general perspectives about IMPACT? What comes to mind when I tell you today, we are going to talk about IMPACT?
2. Describe your understanding of how IMPACT works and what measures go into your score?
 - 2.1. What is your understanding of how IMPACT works across DCPS?
3. How does IMPACT work in practice, specifically at your school?
 - 3.1. How integral is IMPACT to you the daily work of your school?
 - 3.2. How does IMPACT relate to the relationships and trust in your school?
4. How were you evaluated at a previous teaching job/administration (if you had one)? How does that evaluation experience compare to the current experience of IMPACT?
5. Think about each of the different measures that make up your IMPACT score (i.e., Essential Practices observations (EP), Teacher-Assessed Student Achievement (TAS), Individual Value Added (IVA), Student Surveys (SSP), Commitment to School Community (CSC), and Core Professionalism (CP). Which parts do you believe are most effective? Which are least effective? Why?

5.1. If you could change any of these measures, which would you change? Why? And how?

6. Do you find the evaluation results you receive are reflective of your performance? Why or Why not?

6.1. Did you experience feeling valued in the IMPACT process and if so how? If not, why not?

Experiences with IMPACT for Professional Growth

7. Please share an experience you have had receiving IMPACT feedback? From that please describe your experience with the follow up and outcomes of the feedback experience?

8. Think about each of the different parts of the IMPACT feedback process as you experienced them. Which parts do you believe are most effective in your professional growth? Which are least effective? Why?

8.1. how you obtain information about IMPACT

8.2. observation procedures

8.3. receiving feedback and debriefing (EP observation conference)

8.4. goalsetting

9. How does your experience of receiving feedback in IMPACT connect to your other professional growth opportunities in DCPS (e.g., professional learning and coaching)?

10. Do you feel IMPACT has changed your teaching practice? If so, how? If not, why not?

10.1. Do you think IMPACT has changed your collaboration with teachers across expertise? If so how? If not, why not?

10.2. How did the IMPACT process provide you with feedback on the subject-matter of the specific lesson you taught?

11. Does IMPACT make you think differently about your career path, and if so how?

11.1. Has IMPACT had any influence on your retention as a teacher within DCPS? In the teaching profession broadly?

11.2. What has been your experience with the other DCPS programs that relate to IMPACT? How do they relate to your career path? Specifically:

11.2.1. IMPACT*plus* (bonus structure)

11.2.2. LIFT (the teacher career ladder including the salary schedule and reduced assessments)

11.2.3. Standing Ovation (yearly celebration of highly effective teachers)

The Effectiveness of IMPACT for Teachers, Students, and DCPS

12. What effect do you think IMPACT has on the effectiveness of DCPS broadly? Why?

12.1. Do you think IMPACT is effective in keeping DCPS teachers accountable? Why or why not?

- 12.2. What effect do you think IMPACT has on student learning? Why?
- 13. Has IMPACT improved the quality of teachers at your school if so how? If not, why not?
 - 13.1. How does IMPACT relate to expectation-setting for teachers in your school?
 - 13.2. What role has had IMPACT had on inclusive instructional practices with students with disabilities?
- 14. Do you think IMPACT is fair as an evaluation system? Why or Why not? Consider fairness across themes including:
 - 14.1. Title, gender, race/ethnicity, ward, school, staff/student?

Closing Opinions

- 15. If you had a magic wand to make IMPACT most useful for DCPS what would you do?
 - 15.1. Most useful for teachers?
 - 15.2. Most useful for students?
- 16. Is there anything you want to share about your experience with IMPACT that I did not ask about?
- 17. Is there a colleague you think I should talk to who might provide a different perspective on IMPACT?

10 Tables

Table 1.

Characteristics of Interview Sample (n=46) vs. DCPS

Interviewed teachers				<u>All Teachers</u>			
Race	F	M	Total	Race	F	M	Total
American Indian or Native Alaskan	1	0	1	American Indian or Native Alaskan	9	3	12
	2.17	0	2.17		0.22	0.07	0.29
Black	20	2	22	Asian	113	29	142
	43.48	4.35	47.83		2.76	0.71	3.47
Hispanic/Latino	3	1	4	Black	1,506	466	1,972
	6.52	2.17	8.7		36.75	11.37	48.12
Other/Unknown	2	1	3	Hispanic/Latino	228	94	322
	4.35	2.17	6.52		5.56	2.29	7.86
White	10	6	16	Native Hawaiian or Pacific Islander	3	0	3
	21.74	13.04	34.78		0.07	0	0.07
Total	36	10	46	Other/Unknown	243	84	327

78.26 21.74 100

	5.93	2.05	7.98
White	978	342	1,320
	23.87	8.35	32.21
Total	3,080	1,018	4,098
	75.16	24.84	100

Subjects	Freq.	Percent	Subjects	Freq.	Percent
All Subjects	10	21.74	All Subjects	1,007	24.9
Art/Music	1	2.17	Art/Music	241	5.96
CES	1	2.17	CES	76	1.88
ELA	7	15.22	CTE	81	2
ELL/ESL	2	4.35	Child Development	1	0.02
Health/PE	1	2.17	ELA	729	18.03
Math	10	21.74	ELL/ESL	198	4.9
Science	1	2.17	Health/PE	173	4.28
Science/Social Studies	4	8.7	ILS	2	0.05
Special Education	6	13.04	Math	578	14.29
World Languages	3	6.52	Other	42	1.04
			Science	38	0.94
Total	46	100	Science/Social Studies	334	8.26

Social Studies	29	0.72
Special Education	371	9.17
World Languages	144	3.56

IMPACT Rating 2018-19			IMPACT Rating 2018-19		
	Freq.	Percent		Freq.	Percent
Developing	3	6.98	Developing	328	9.34
Effective	21	48.84	Effective	1,477	42.06
Highly Effective	19	44.19	Highly Effective	1,645	46.84
			Ineffective	5	0.14
Total	43	100	Minimally Effective	48	1.37
			No Rating	9	0.26

Ward	Freq.	Percent	Ward	Freq.	Percent
1	6	13.04	1	494	12.05
2	2	4.35	2	230	5.61
3	8	17.39	3	534	13.03
4	9	19.57	4	756	18.45
5	4	8.7	5	392	9.57
6	3	6.52	6	627	15.3

7	6	13.04	7	504	12.3
8	8	17.39	8	561	13.69
Total	46	100	Total	4,098	100

First IMPACT			First IMPACT		
Year	Freq.	Percent	Year	Freq.	Percent
2009-2010	10	21.74	2009-2010	950	23.18
2010-2011	2	4.35	2010-2011	156	3.81
2011-2012	3	6.52	2011-2012	181	4.42
2012-2013	3	6.52	2012-2013	168	4.1
2013-2014	2	4.35	2013-2014	239	5.83
2014-2015	2	4.35	2014-2015	323	7.88
2015-2016	7	15.22	2015-2016	424	10.35
2016-2017	3	6.52	2016-2017	335	8.17
2017-2018	7	15.22	2017-2018	406	9.91
2018-2019	4	8.7	2018-2019	419	10.22
2019-2020	3	6.52	2019-2020	497	12.13
Total	46	100			

Primary Grade Span	Freq.	Percent	Primary Grade Span	Freq.	Percent

Elementary	28	60.87	Elementary	2,545	62.1
High	7	15.22	High	893	21.79
Middle	11	23.91	Middle	660	16.11
Total	46	100	Total	4,098	100

IVA Impact	Freq.	Percent	IVA Impact	Freq.	Percent
No	33	71.74	No	3,391	82.75
Yes	13	28.26	Yes	707	17.25
Total	46	100			

Title I	Freq.	Percent	Title I	Freq.	Percent
Not Title I	11	23.91	NA	10	0.24
Title I	34	73.91	Not Title I	901	21.99
Title I Targeted Assistance	1	2.17	Title I	3,121	76.16
			Title I Targeted Assistance	66	1.61
Total	45	100			

Pipeline	Freq.	Percent	Pipeline	Freq.	Percent

N/A (Traditional, other, or unknown)	39	84.78	N/A (Traditional, other, or unknown)	3,797	92.65
Teach For America	5	10.87	Relay GSE	29	0.71
Urban Teachers	2	4.35	Teach For America	170	4.15
			Urban Teachers	102	2.49

Table 2

Perceptions of IMPACT Supporting Professional Growth, by Teacher Race, Gender, and Component of IMPACT

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
	American Indian or Alaska Native F	American Indian or Alaska Native M*	Asian F	Asian M	Black F	Black M	Hispanic/ Latino F	Hispanic/ Latino M	Native Hawaiian or Other Pacific *	Other/ Unknown F	Other/ Unknown M	White F	White M
My experiences with IMPACT support my professional growth.	2.71	--	2.74	3.17	2.47	2.70	2.49	2.78	--	2.54	1.76	2.15	2.04
The Commitment to School Community (CSC) component of IMPACT supports my professional growth.	3.29	--	2.76	3.39	2.50	2.83	2.46	2.89	--	2.56	2.23	2.17	2.13
The Essential Practices component of IMPACT supports my professional growth.	2.86	--	2.85	3.30	2.71	2.91	2.61	2.93	--	2.79	1.98	2.41	2.37

The Individual Value-Added component of IMPACT supports my professional growth.	2.43	--	2.30	3.13	2.14	2.61	2.11	2.71	--	2.13	1.52	1.65	1.70
The Teacher Assessed Student Achievement Data goals (TAS) component of IMPACT supports my professional growth.	2.86	--	2.79	3.00	2.53	2.75	2.29	2.78	--	2.45	2.13	2.00	1.82
The Student Survey component of IMPACT supports my professional growth.	2.71	--	2.22	2.91	2.24	2.48	2.12	2.55	--	2.19	1.83	1.88	1.76
Observations	7	2	78	23	2	299	169	76	2	165	52	755	262

*Data suppressed because fewer than 5 individuals in cell.

Table 3*Perceptions of IMPACT Supporting Professional Growth, by Ward and IMPACT Component*

	<u>Ward</u>							
	1	2	3	4	5	6	7	8
My experiences with IMPACT support my professional growth.	2.21	2.40	2.24	2.33	2.43	2.40	2.51	2.61
The Commitment to School Community (CSC) component of IMPACT supports my professional growth.	2.25	2.54	2.40	2.33	2.43	2.49	2.62	2.55
The Essential Practices component of IMPACT supports my professional growth.	2.51	2.75	2.52	2.57	2.64	2.58	2.70	2.81
The Individual Value-Added component of IMPACT supports my professional growth.	1.91	2.00	1.71	2.04	2.24	2.00	2.22	2.24
The Teacher Assessed Student Achievement Data goals (TAS) component of IMPACT supports my professional growth.	1.98	2.41	2.15	2.27	2.51	2.37	2.57	2.58
The Student Survey component of IMPACT supports my professional growth.	2.29	1.99	1.94	2.13	2.21	1.90	2.30	2.23
Observations	369	180	370	557	321	425	342	368

Figure 1

Distribution of Changes in Essential Practices Scores, Cycle 1 to Cycle 3, 2018-2019

Table 4

Regression of Changes in EP Scores from Cycle 1 to Cycle 3 on Teacher and School Characteristics

VARIABLES	Essential Practices Score - Change Cycle 1 to Cycle 3
American Indian or Alaska Native	-0.0641 (0.187)
Asian	-0.0204 (0.0397)
Black	-0.0179 (0.0167)
Hispanic/Latino	0.0478* (0.0269)
Native Hawaiian or Other Pacific Islander	-0.0328 (0.324)
Other/Unknown	-0.0333 (0.0277)
Female	-0.0114 (0.0156)
Title I	-0.0103 (0.0441)

Students % Black	-0.000402
	(0.000517)
Students % Free/Reduced-Price Lunch	0.00129**
	(0.000524)
Ward = 2	-0.00558
	(0.0412)
Ward = 3	0.0378
	(0.0408)
Ward = 4	0.000324
	(0.0256)
Ward = 5	0.0284
	(0.0384)
Ward = 6	0.0513
	(0.0337)
Ward = 7	0.0705*
	(0.0400)
Ward = 8	0.0422
	(0.0403)
High School	0.0361*

	(0.0190)
Middle School	0.0605***
	(0.0195)
Constant	0.0362
	(0.0398)
Observations	2,362
R-squared	0.019

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Table 5

Logistic Regression of Increase in IMPACT Score from 2017-18 to 2018-19 on Teacher and School Characteristics

VARIABLES	Increase from 17-18 to 18-19 IMPACT Score
American Indian or Alaska Native	-0.272 (1.091)
Asian	-0.159 (0.320)
Black	0.272** (0.132)
Hispanic/Latino	0.499** (0.205)
Native Hawaiian or Other Pacific Islander	-
Other/Unknown	0.472** (0.211)
Female	-0.216* (0.125)
Title I	-0.568

	(0.377)
Students % Black	-0.000844
	(0.00390)
Students % Free/Reduced-Price Lunch	0.00870**
	(0.00422)
Ward = 2	-0.663**
	(0.337)
Ward = 3	-0.275
	(0.322)
Ward = 4	-0.200
	(0.199)
Ward = 5	0.235
	(0.286)
Ward = 6	-0.0301
	(0.250)
Ward = 7	0.0432
	(0.308)
Ward = 8	-0.0993
	(0.314)

High School	-0.0844
	(0.153)
Middle School	0.354**
	(0.151)
Constant	-1.437***
	(0.317)
Observations	2,265

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6*Mean IMPACT scores by Teacher Race and Gender and IMPACT Component*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	
	Ameri can Indian or Alaska Native F	Ameri can Indian or Alaska Native M*	Asia n F	Asian M	Black F	Blac k M	Latino F	Hispa nic/ Latino M	Hispa nic/ Latino M	Nativ e Hawai ian or Other Pacifi c Island er M or F*	Other/ Unkn own F	Other/ Unkno wn M	Whit e F	Whit e M
IMPACT score	355.13	--	345.12	334.74	337.76	325.47	339.30	325.43	--	340.32	327.92	350.47	341.90	
Bonus	12000.00	--	9530.61	10363.64	11563.49	9580.65	7793.10	8444.44	--	11154.55	11000.00	8085.71	7787.23	
EP Score	3.43	--	3.39	3.27	3.34	3.22	3.34	3.18	--	3.36	3.23	3.48	3.40	
EP1	3.61	--	3.59	3.49	3.61	3.50	3.55	3.46	--	3.61	3.50	3.65	3.58	
EP2	3.29	--	3.33	3.20	3.21	3.15	3.24	3.09	--	3.24	3.20	3.38	3.34	
EP3	3.55	--	3.46	3.35	3.38	3.22	3.37	3.19	--	3.41	3.27	3.54	3.45	
EP4	3.36	--	3.25	3.11	3.20	3.06	3.21	3.02	--	3.21	3.07	3.36	3.27	
EP5	3.33	--	3.33	3.22	3.32	3.16	3.31	3.16	--	3.33	3.10	3.45	3.36	
EP Cycle 1	3.39	--	3.36	3.24	3.29	3.17	3.27	3.11	--	3.34	3.22	3.43	3.36	
EP Cycle 2	3.42	--	3.37	3.17	3.32	3.20	3.34	3.14	--	3.34	3.14	3.44	3.37	
EP Cycle 3	3.27	--	3.36	3.23	3.33	3.24	3.33	3.23	--	3.32	3.27	3.44	3.37	
CP Overall	0.00	--	-0.90	0.00	-1.48	-2.24	-0.68	-1.11	--	-1.18	-1.25	-0.48	-1.48	
CSC Overall	3.78	--	3.70	3.59	3.60	3.49	3.72	3.58	--	3.62	3.54	3.76	3.70	

TAS													
Overall	3.75	--	3.58	3.58	3.50	3.41	3.49	3.45	--	3.50	3.53	3.59	3.56
IVA													
Overall	--	--	3.80	3.31	3.11	3.16	3.25	3.09	--	3.29	3.25	3.18	3.12
IVA													
Reading	--	--	4.00	3.20	3.08	3.17	3.35	3.00	--	3.40	3.38	3.18	3.01
IVA													
Math	--	--	3.76	3.42	3.13	3.14	3.08	3.13	--	3.19	3.19	3.18	3.20
SSOP													
Overall	3.93	--	3.05	3.12	3.30	3.15	3.17	3.20	--	3.33	3.10	3.26	3.17
Observations	8	3	89	27	1285	384	176	72	3	203	64	806	277

*Results suppressed for privacy/confidentiality due to fewer than 5 individuals in cell

Table 7*IMPACT Scores by Ward and IMPACT Component*

	Ward							
	1	2	3	4	5	6	7	8
IMPACT score	339.12	352.38	348.94	339.57	330.54	346.38	333.07	331.02
Bonus	11579.7	5984.5	2562.0	9529.4	11612.2	9203.2	12712.82	16206.7
EP Score	3.37	3.52	3.48	3.35	3.25	3.42	3.28	3.29
EP1	3.61	3.67	3.64	3.62	3.50	3.63	3.56	3.52
EP2	3.27	3.49	3.40	3.19	3.17	3.31	3.17	3.21
EP3	3.42	3.52	3.55	3.39	3.26	3.47	3.30	3.32
EP4	3.23	3.41	3.38	3.23	3.08	3.27	3.11	3.13
EP5	3.31	3.50	3.46	3.31	3.22	3.39	3.26	3.25
EP Cycle 1	3.33	3.49	3.45	3.30	3.20	3.37	3.22	3.23
EP Cycle 2	3.32	3.49	3.46	3.33	3.21	3.39	3.25	3.27
EP Cycle 3	3.34	3.47	3.43	3.34	3.24	3.37	3.31	3.29
CP Overall	-0.66	-0.73	-1.02	-1.02	-1.77	-0.88	-1.28	-2.36
CSC Overall	3.71	3.81	3.77	3.63	3.52	3.70	3.55	3.52
TAS Overall	3.32	3.66	3.62	3.52	3.52	3.66	3.45	3.44
IVA Overall	3.25	2.98	3.16	3.24	3.12	3.14	3.14	3.13
IVA Reading	3.35	2.84	3.12	3.19	3.06	3.15	3.13	3.12
IVA Math	3.13	3.14	3.18	3.26	3.12	3.12	3.18	3.13
SSOP Overall	3.31	3.21	3.09	3.32	3.24	3.18	3.35	3.17

Observations	411	411	411	411	411	411	411	411
--------------	-----	-----	-----	-----	-----	-----	-----	-----

Table 8*IMPACT Scores by School Title I Status and IMPACT Component*

	(1)	(2)
	Not Title I	Title I
IMPACT score	328.36	336.43
Bonus	4203.18	12369.82
EP Score	3.29	3.32
EP1	3.49	3.57
EP2	3.26	3.20
EP3	3.31	3.36
EP4	3.15	3.17
EP5	3.25	3.29
EP Cycle 1	3.27	3.27
EP Cycle 2	3.24	3.29
EP Cycle 3	3.21	3.31
CP Overall	-2.41	-1.34
CSC Overall	3.58	3.60
TAS Overall	3.35	3.50
IVA Overall	3.01	3.19
IVA Reading	2.97	3.19
IVA Math	3.04	3.18
SSOP Overall	3.08	3.28
Observations	1438	2607

Table 9*IMPACT Scores by School Level/Grade Span and IMPACT Component*

	(1) Elementary	(2) Middle	(3) High
IMPACT score	342.33	333.37	336.68
Bonus	9426.74	9481.01	11077.84
EP Score	3.40	3.30	3.30
EP1	3.62	3.55	3.55
EP2	3.28	3.19	3.24
EP3	3.45	3.33	3.32
EP4	3.27	3.15	3.13
EP5	3.37	3.26	3.25
EP Cycle 1	3.36	3.24	3.25
EP Cycle 2	3.37	3.28	3.27
EP Cycle 3	3.36	3.31	3.29
CP Overall	-1.02	-1.34	-1.73
CSC Overall	3.64	3.66	3.63
TAS Overall	3.51	3.45	3.60
IVA Overall	3.17	3.15	3.19
IVA Reading	3.13	3.18	3.18
IVA Math	3.19	3.11	3.21
SSOP Overall	3.19	3.18	3.33
Observations	2148	514	735

Table 10

IMPACT Scores by Subject and IMPACT Component

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	All Subj ects	Art/M usic	CES	CTE	ELA	ELL/ ESL	Healt h/PE	IL S*	Math	Othe r	Scie nce	Scie nce/ Soci al	Socia l	Speci al	World Langu ages
IMPACT score	344.60	339.87	355.56	323.24	337.64	342.10	342.06	--	337.21	330.85	342.73	333.51	321.36	341.85	334.11
Bonus	9088.30	8971.43	13960.78	9538.46	10494.88	8120.88	9184.21	--	11404.65	7916.67	9615.39	9346.46	15000.00	10082.35	6236.36
EP Score	3.41	3.38	3.41	3.22	3.38	3.37	3.39	--	3.38	3.21	3.34	3.27	3.11	3.31	3.28
EP1	3.63	3.56	3.65	3.56	3.62	3.62	3.64	--	3.61	3.45	3.47	3.49	3.42	3.59	3.41
EP2	3.31	3.34	3.17	3.13	3.27	3.24	3.27	--	3.29	3.15	3.37	3.22	3.06	3.17	3.21
EP3	3.45	3.43	3.53	3.22	3.42	3.43	3.45	--	3.41	3.21	3.40	3.31	3.15	3.34	3.35
EP4	3.28	3.25	3.28	3.06	3.24	3.22	3.24	--	3.27	3.05	3.17	3.11	2.80	3.14	3.14
EP5	3.38	3.34	3.42	3.11	3.34	3.34	3.34	--	3.33	3.17	3.27	3.22	3.11	3.33	3.27
EP Cycle 1	3.37	3.34	3.37	3.21	3.33	3.34	3.33	--	3.33	3.18	3.21	3.20	3.02	3.27	3.22
EP Cycle 2	3.38	3.34	3.36	3.14	3.35	3.33	3.34	--	3.37	3.13	3.39	3.26	3.24	3.28	3.25
EP Cycle 3	3.36	3.36	3.37	3.15	3.36	3.37	3.39	--	3.37	3.21	3.33	3.24	3.36	3.32	3.21
CP Overall	0.81	-1.84	-1.00	3.68	-1.31	-0.36	-1.53	--	-1.31	1.18	0.45	1.41	-6.36	-1.10	-1.44
CSC Overall	3.65	3.61	3.62	3.49	3.67	3.69	3.55	--	3.64	3.61	3.73	3.66	3.69	3.64	3.66
TAS Overall	3.54	3.66	3.69	3.44	3.41	3.55	3.73	--	3.44	3.47	3.67	3.53	3.77	3.50	3.69
IVA Overall	3.07	2.55	.	.	3.16	2.79	.	--	3.19	3.65	3.70	3.18	.	3.57	3.20

IVA															
Reading	2.92	1.90	.	.	3.16	2.95	.	--	3.16	3.65	.	3.17	.	4.00	3.20
IVA															
Math	3.17	3.20	.	.	3.07	2.64	.	--	3.19	.	3.70	3.18	.	3.35	.
SSOP															
Overall	3.20	3.07	3.06	3.18	3.30	3.33	3.19	--	3.29	3.17	3.35	3.24	3.28	3.26	3.06
Observations	865	201	70	68	604	167	144	2	465	34	22	283	11	319	118

Table 11*Regression of IMPACT Score on Teacher and School Characteristics*

VARIABLES	IMPACT score
American Indian or Alaska Native	9.874 (9.949)
Asian	-2.243 (3.090)
Black	-8.092*** (1.355)
Hispanic/Latino	-9.605*** (2.268)
Native Hawaiian or Other Pacific Islander	17.93 (18.12)
Other/Unknown	-7.301*** (2.208)
Female	10.09*** (1.302)
Title I	6.632* (3.410)

School % Black students	0.0125
	(0.0416)
School % students receiving free or reduced-price lunch	-0.193***
	(0.0412)
Ward = 2	7.843**
	(3.148)
Ward = 3	2.125
	(3.138)
Ward = 4	-1.221
	(2.100)
Ward = 5	-6.241**
	(3.134)
Ward = 6	2.564
	(2.677)
Ward = 7	-3.190
	(3.262)
Ward = 8	-5.773*
	(3.278)
High School	-2.325

	(1.552)
Middle School	-8.955***
	(1.647)
Constant	347.6***
	(3.116)
Observations	3,334
R-squared	0.097

Table 12*Regression of Essential Practices Rubric Scores on Teacher and School Characteristics*

	(1)	(2)	(3)	(4)	(5)
VARIABLES	EP1	EP2	EP3	EP4	EP5
American Indian or Alaska Native	-0.0152 (0.122)	0.170 (0.136)	0.0181 (0.136)	0.0805 (0.136)	0.0316 (0.126)
Asian	-0.0344 (0.0380)	-0.0224 (0.0422)	-0.0339 (0.0422)	-0.0684 (0.0423)	-0.0768* (0.0393)
Black	0.00484 (0.0167)	-0.0870*** (0.0185)	-0.102*** (0.0185)	-0.0791*** (0.0185)	-0.0779*** (0.0172)
Hispanic/Latino	-0.100*** (0.0279)	-0.111*** (0.0310)	-0.167*** (0.0310)	-0.154*** (0.0310)	-0.114*** (0.0288)
Native Hawaiian or Other Pacific Islander	0.245 (0.223)	0.0915 (0.248)	-0.137 (0.247)	0.0992 (0.248)	0.143 (0.230)
Other/Unknown	-0.00817 (0.0272)	-0.0829*** (0.0302)	-0.0855*** (0.0301)	-0.112*** (0.0302)	-0.107*** (0.0281)
Female	0.0801*** (0.0160)	0.0446** (0.0178)	0.108*** (0.0178)	0.101*** (0.0178)	0.116*** (0.0165)

Title I	6.92e-05	-0.202***	0.0650	0.109**	0.111**
	(0.0420)	(0.0466)	(0.0466)	(0.0467)	(0.0433)
Students % Black	-0.00116**	-0.00128**	0.000335	-0.00115**	0.000572
	(0.000512)	(0.000569)	(0.000568)	(0.000570)	(0.000529)
Students % Free/Reduced- Price Lunch	-0.000733	-0.00181***	-0.00251***	-0.00316***	-0.00274***
	(0.000507)	(0.000563)	(0.000563)	(0.000564)	(0.000524)
Ward 2	0.0141	0.0268	-0.00532	0.0832*	0.111***
	(0.0388)	(0.0430)	(0.0430)	(0.0431)	(0.0400)
Ward 3	-0.0543	-0.237***	-0.00563	-0.0103	0.0563
	(0.0386)	(0.0429)	(0.0428)	(0.0430)	(0.0399)
Ward 4	-0.0203	-0.118***	-0.0775***	-0.0328	-0.0371
	(0.0259)	(0.0287)	(0.0287)	(0.0288)	(0.0267)
Ward 5	-0.0722*	0.0109	-0.151***	-0.0599	-0.0857**
	(0.0386)	(0.0428)	(0.0428)	(0.0429)	(0.0398)
Ward 6	0.00965	-0.0209	-0.0410	-0.00796	-0.00869
	(0.0329)	(0.0366)	(0.0365)	(0.0366)	(0.0340)
Ward 7	-0.0106	0.0235	-0.119***	-0.0211	-0.0578
	(0.0402)	(0.0446)	(0.0445)	(0.0447)	(0.0414)

Ward 8	-0.0419	0.0547	-0.103**	-0.000729	-0.0680
	(0.0403)	(0.0448)	(0.0448)	(0.0449)	(0.0416)
High School	-0.0359*	0.00282	-0.107***	-0.0963***	-0.0942***
	(0.0191)	(0.0212)	(0.0212)	(0.0212)	(0.0197)
Middle School	-0.0475**	-0.0723***	-0.120***	-0.127***	-0.118***
	(0.0203)	(0.0225)	(0.0225)	(0.0225)	(0.0209)
Constant	3.697***	3.695***	3.602***	3.464***	3.425***
	(0.0384)	(0.0426)	(0.0425)	(0.0427)	(0.0396)
Observations	3,334	3,334	3,334	3,334	3,334
R-squared	0.043	0.105	0.094	0.106	0.089
