



Inside school turnaround: What drives success?

Kirsten Lee Hill¹ · Laura Desimone² · Tonya Wolford³ · Adrienne Reitano³ · Andrew Porter⁴

Accepted: 10 January 2022 / Published online: 29 March 2022
© The Author(s) 2022

Abstract

This study proposes an empirically grounded theory of how school reform implementation relates to effectiveness, useful for developing and studying many approaches to school reform both in the U.S. and abroad, and also for assessing how policymakers and implementers might leverage various aspects of implementation to create effective school improvement models at scale. Guided by a new framework that links Bryk and colleagues' (2010) five essential supports and as reported by Desimone's (2002) adaptation of Porter and colleagues' (1986) policy attributes theory, we use a mixed-methods approach to study the implementation and effectiveness of school turnaround efforts in the School District of Philadelphia. We explore the relationships among key turnaround model components, approaches to model implementation, and academic achievement using a matched comparison design and estimating a series of regressions. Qualitative methods are used to contextualize findings and offer explanatory hypotheses.

Keyword Education policy · Effective schools · Mixed methods · School reform · Turnaround

School turnaround: An important part of the policy landscape we know little about

School turnarounds came to the forefront of education reform in the United States in the Obama Administration's American Recovery and Reinvestment Act of 2009 (ARRA).¹ The prominence of school turnaround on the federal education policy agenda (e.g., Ayers et al., 2012; Kutash et al., 2010; Trujillo & Renée, 2015) makes

¹ This act outlined four models of “interventions for struggling schools” as part of the State Fiscal Stabilization Fund, School Improvement Grants (SIGs), and Race to the Top (U.S. Department of Education, 2009a, p. 35). These four models, discussed by the U.S. Department of Education as options for turning around our nation's lowest-performing schools, stem from, and refine, the five school restructuring options outlined by the No Child Left Behind Act of 2001 (NCLB) (American Institutes for Research, 2011; U.S. Department of Education, 2003). The four models are described as follows: (1) turnaround—

✉ Kirsten Lee Hill
kirsten@kirstenleehill.com

Extended author information available on the last page of the article

it critical to evaluate not only its effectiveness in improving student achievement but also the mechanisms through which the reforms operate. Codifying the turnaround process through the framework we propose will additionally be beneficial to other school reform efforts, both nationally and abroad, which may not use “turnaround” nomenclature, but often reflect similar mechanisms for generating improvement (e.g., Liu, 2017).

Research on school turnarounds is limited. There are a number of case studies chronicling change efforts (Calkins et al., 2007; Peck & Reitzug, 2018; Peurach & Neumerski, 2015; What Works Clearinghouse, 2008) and a few causal studies that reveal inconsistent success (Heissel & Ladd, 2018; Strunk et al., 2016a, 2016b; Zimmer et al., 2017). Consequently, there is a lack of consensus about what works to turnaround schools and why (de la Torre et al., 2012; Tannenbaum et al., 2015; VanGronigen, & Meyers, 2019). In fact, some argue that there is a lack of compelling evidence that school turnarounds could even possibly be effective (Murphy & Bleiberg, 2019).

Research in this area is further complicated by semantics as the term “turnaround” can refer to both a specific approach to school reform that involves replacing a school’s principal and at least 50 percent of the staff, as well as an overall strategy for improving low performing schools (Dee, 2012; Herman, et al., 2008; What Works Clearinghouse, 2008; Zavadsky, 2012). This ambiguity in terminology makes understanding reform models’ theories of change critical to interpreting results.

One of the critical weaknesses of this literature, which we intend our work to address, is the lack of a cohesive conceptual framework for understanding and interpreting turnaround efforts. School turnaround efforts have been critiqued for both their lack of grounding in a conceptual framework and systematic investigation into the reasons behind success and failure (McCauley, 2018; Meyers & Smylie, 2017). We build on this work by proposing the Integrative Framework for School Improvement, appropriate for studying school turnaround and other school reform efforts.

Our proposed framework provides an organizing schema to describe and study each component of a turnaround/reform effort. One of our central goals in this work is to move beyond the semantics of turnaround. We believe the labels and names applied to various efforts risk creating artificial differences that limit our generation of knowledge of how to improve schools—knowledge that is globally useful. The framework we propose drills down to the core components of reform initiatives, offering a simple and common language we can use to distill reform efforts so that across schools, districts, states, and countries, we can learn from one another about what does and does not work to improve schools. This paper presents the results of a study that tested this framework in the School District of Philadelphia, yielding research that sheds light on what does and does not work to improve schools.

Footnote 1 (continued)

hire a new principal, replace at least 50 percent of the staff, and implement a comprehensive plan for school improvement; (2) restart—reopen the school under the management of a charter or education management organization; (3) closure—close the school and re-assign students to higher-achieving schools; and (4) transformation—replace the principal and implement comprehensive instructional and organizational reforms (e.g., U.S. Department of Education, 2009a, p. 35).

Conceptual framework: Integrating essential supports with policy attributes

Advancing studies of school turnaround requires a strong conceptual framework that can be applied to various models of school reform to better understand the mechanisms through which reforms operate and what impacts these mechanisms have on the reforms' successes and/or failures. School turnaround's goal of improving schools and the comprehensive nature of the reform (targeting various aspects of a school including leadership, school climate, and instruction) (Center on School Turnaround, 2017) point to the value of considering the extensive bodies of literature on effective schools and comprehensive school reform (CSR) in developing a conceptual framework to guide a study of school turnaround.

To study the success of schools, it is critical to understand both the key components of a reform strategy (such as its approach to instruction and school climate), as well as the mechanisms used to implement the strategies. To this end, we propose the Integrative Framework for School Improvement which joins Bryk and colleagues' (2010) work on the five essential supports with Desimone's (2002) adaptation of Porter and colleagues' (1986) policy attributes theory. To put it simply, the five essential supports capture the key components of a reform strategy whereas the policy attributes capture how schools choose to implement these components.

Decades of research have identified key components of effective schools (e.g., Purkey & Smith, 1983); research by Bryk et al. (2010) on school improvement operationalized key components of effective schools by identifying and developing measures for five essential supports for school improvement. Bryk and colleagues found that *each of the following five essential supports are necessary* for school improvement:

- *School leadership*: the extent to which a leader sets shared goals and high expectations and is able to engage and motivate key stakeholders.
- *Parent-community ties*: the extent to which teachers engage families in supporting their child's learning, facilitate families' engagement with school events, are familiar with students' culture and community, and to which schools have partnerships within the community to ensure the needs of the whole child are met.
- *Professional capacity*: the quality of teachers and professional development opportunities, staff's dispositions and beliefs, and the extent to which staff collaborate and have collegial relationships.
- *Student-centered learning climate*: order and safety within a school, the extent to which teachers both challenge and support students, and the quality of relationships among staff and students.
- *Instructional guidance*: how curriculum and content are organized and delivered, including a balance of basic and more applied skills.

They found that weakness in even one of these areas significantly reduced the likelihood of success, defined as "academic productivity," a construct of statewide standardized test scores that considers value-add change over time (Bryk et al.,

2010). The five essential supports provide a useful framework for understanding the key components of various school improvement models. By measuring the strength of the five essential supports in different school improvement models, we can identify critical components of each model's reform strategy and evaluate where a school ought to focus its efforts to improve. In other words, we can use measures of the essential supports to paint a picture of a school's approach to improvement—for example, does its reform strategy emphasize school climate over instruction?

In studies of school reform, research has long recognized the importance of studying both implementation and outcomes as there is consensus that the success or failure of an intervention or policy is best understood within the context of how it was implemented (e.g., Berends, 2000; McLaughlin, 1987). Our Integrative Framework for School Improvement employs the use of Desimone's (2002) adaptation of Porter and colleagues' (1986) policy attributes theory which posits that implementation of any policy or reform is improved through being more specific, consistent, authoritative, powerful, and stable (Porter, 1989; see also: Porter, 1994). Desimone (2002) suggests that rather than a "true" value of policy attributes driving implementation (i.e., a value as reflected in legislation, policy statement or expert review), it is educators' perception of the attributes that influence their behavior, and we adopt this perspective. The policy attributes are defined as follows:

- *Specificity*: degree of detail, clarity, and guidance. Specificity facilitates implementation by making clear goals, roles, and what stakeholders need to do in order to successfully implement policy/programming.
- *Authority*: buy-in and resources. Authority facilitates implementation by providing necessary support—people, time, and money.
- *Consistency*: degree of alignment with other reform efforts and policies within an organization, school, or district. Consistency facilitates implementation by ensuring that everyone is on the same page and that various initiatives work together.
- *Power*: rewards and sanctions/penalties. Power facilitates implementation by exerting pressure to implement.
- *Stability*: consistency of people, policies, and contextual factors over time. Stability is critical because volatile improvement efforts signal that changes are temporary, which weakens commitment to policies/programming.

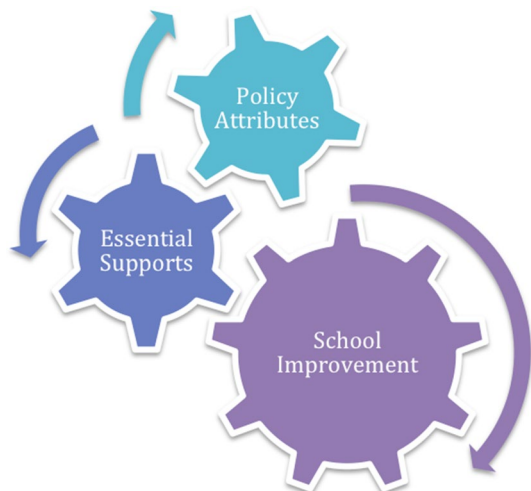
Whereas the five essential supports tell the story of *what* a school is doing, the five policy attributes look under the hood of school improvement, shedding light on *how* a school implements their overall reform strategies. Merging these two frameworks provides common language for exploring both the "what" (essential supports) and the "how" (policy attributes) of school reform. The Integrative Framework for School Improvement allows researchers and practitioners alike to unpack the idea

of “success” in school reform and use careful examination of reform components and implementation strategies to gain an understanding of why a particular school improvement model is or is not associated with gains in academic achievement, the most common metric of success for school turnaround efforts.

By including measures of the essential supports and policy attributes in analyses, we join other researchers in challenging the standard assumption that improved student achievement as measured by standardized tests adequately captures the complexities of school improvement (e.g., Datnow & Stringfield, 2000; Hallinger & Murphy, 1986; Purkey & Smith, 1983; Rosenholtz, 1985; WWC, 2008). In fact, this approach is uniquely suited to capture the complexities of school improvement allowing us to move beyond a focus on student achievement to answer questions related to identifying how schools build and support reform components, and how the components work in tandem for school success.

Our Integrative Framework for School Improvement, shown in Fig. 1 below, hypothesizes that school improvement is mediated by the essential supports and policy attributes and allows for interactions among all these aspects of reform. We can think of school improvement (the goal of reform efforts) as its own gear for creating change, turned by the gears of the essential supports and policy attributes. This framework is a mechanism for systematizing our understanding and evaluation of various school improvement efforts, including school turnarounds. Additional information on the Integrative Framework for School Improvement can be found in Hill, Desimone, Wolford, & Reitano (2017).

Fig. 1 Conceptual framework



Research questions and study context: The role of implementation in school turnaround success

Guided by our Integrative Framework for School Improvement, we address three research questions: (1) What are the relationships between the various approaches to school turnaround and academic achievement? (2) To what extent do the essential supports and policy attributes mediate the relationship between school turnaround and academic achievement? And (3) How do teachers' and principals' descriptions of their schools' approaches to improvement map on to the essential supports and policy attributes, and help to explain the relative successes and challenges of each school improvement model?

Study context

This study is set in the School District of Philadelphia, a large urban school district in the Northeast (hereafter the District), a leading recipient of School Improvement Grants (SIGs) in 2010. The study takes place within the context of a larger, district-wide study on school improvement efforts that was conducted as part of the Shared Solutions Partnership, a researcher-practitioner partnership between The University of Pennsylvania Graduate School of Education and the School District of Philadelphia. The District identified the need for a study of what the District refers to as its “school turnaround model”—The Renaissance Schools Initiative. This is the name of the District’s primary initiative to improve low-performing schools. One of the central goals of the study was to apply the Integrative Framework for School Improvement to shed light on how the different school turnaround models that make up the Renaissance Schools Initiative operate in the District and also identify ways in which these models can be improved upon.

While referred to as a turnaround model, the Renaissance Schools Initiative is comprised of both District-operated turnarounds as well as restarts (charter-operated schools for which there are a number of providers). In the District, both restarts (formally called Renaissance Charters), and turnarounds (formally called Promise Academies) hire new principals,² replace at least 50 percent of staff, and implement a “comprehensive plan for improvement” (U.S. Department of Education, 2009a, p. 35). The critical difference between the restart and turnaround models is that restarts are operated by charter or education management organizations whereas the District continues to operate the turnaround schools. Colloquially, the District refers to all of these schools as “turnarounds.” In our analyses, we put the thorny semantics of “turnaround” aside and instead consider each turnaround operator independently as a model of school improvement.

² An exception is made if a principal has been at a school for two or fewer years when it is identified for turnaround (Stratos et al., 2014). One school in our sample did retain its principal post-turnaround.

Research design

Our study uses a theory-driven evaluation with a quasi-experimental design to conduct a fine-grained analysis of the implementation and effectiveness of two central approaches to school improvement in the District—namely, restart and turnaround models—as compared to a group of traditional public schools in the District that were not selected to partake in targeted improvement efforts. Theory-driven evaluations can be used to move past discussions about whether or not a program worked, to discussions about how a program or intervention operates (Chen, 1990; Smith, 1990). Additionally, they can be used diagnostically to inform improvements to a program or intervention that is being evaluated (Mark, 1990). This type of evaluation is guided by a conceptual model, and commonly leverages mixed methods to address different components of the model in a holistic way (Chen, 2006). With our conceptual model as a guide, we designed a study to examine the relationship among the policy attributes, essential supports, and student achievement using both quantitative and qualitative methods. Specifically, we use a matched-comparison design, which intends to replicate a randomized control trial (RCT), which is commonly accepted as the most rigorous way to show causality (Song & Herman, 2010). We could not randomly assign treatment because the District determines what schools will engage in its turnaround efforts. Consequently, we used a “matching” process to create two groups that mimic the groups that would have been created in an RCT (Stuart & Rubin, 2007, p. 155). To identify our matches, we created a composite indicator for every school in the District using key demographic and achievement variables and used the “nearest neighbor” method to select the comparison school with the smallest difference in composite indicator for each treatment school (Stuart, 2010). The schools in the treatment condition (those engaging in District-initiated improvement efforts) were selected because they were the lowest performing, academically, in the District. As such, the “nearest neighbor” matches were not all that near, at least not in terms of academic achievement. Baseline analyses demonstrated that not only were English Language Arts (ELA) and math achievement significantly worse in the treatment condition (as we would expect since these were the schools identified as needing improvement) but there is a difference of approximately 1 standard deviation between the two group means, *in favor of the comparison group*. For results of these baseline analyses see Table 1.

We employed mixed methods, integrating analyses of administrative data, survey data, and interview data to capitalize on the strengths of both qualitative and quantitative data (Johnson & Onwuegbuzie, 2004). By looking at achievement data alongside survey and interview data (Desimone, 2012), we were able to gain an enhanced understanding of how different school improvement models work (or do not work) to improve student achievement and additionally paint a picture of what various manifestations of school improvement look like.

Table 1 T-Tests to compare baseline differences between treatment and comparison groups

Variable	Treatment group	N	Mean	Mean difference	Significant
ELA Achievement PSSA (Grades 3–5)	Comparison	16	0.4834	0.9541	0.0001927*
	Treatment	17	−0.4708		
Math Achievement PSSA (Grades 3–5)	Comparison	16	0.5587	1.0867	0.0000094*
	Treatment	17	−0.5280		
ELA Achievement PSSA (All Available Grades)	Comparison	17	0.4487	0.9135	0.0000632*
	Treatment	17	−0.4648		
Math Achievement PSSA (All Available Grades)	Comparison	17	0.5023	1.0193	0.0000087*
	Treatment	17	−0.5171		
Percent ELL	Comparison	17	9.8824	5.3529	0.0928976
	Treatment	17	4.5294		
Percent IEP	Comparison	17	17.5095	2.6312	0.1647917
	Treatment	17	14.8783		
Percent Economically Disadvantaged	Comparison	17	98.4102	3.0481	0.1318681
	Treatment	17	95.3622		
Percent American Indian/Alaskan	Comparison	17	0.5294	0.3529	0.4452020
	Treatment	17	0.1765		
Percent Asian	Comparison	17	1.2353	0.9412	0.1341596
	Treatment	17	0.2941		
Percent Black	Comparison	17	58.5294	−23.4118	0.0445551
	Treatment	17	81.9412		
Percent Hispanic	Comparison	17	35.4118	20.8824	0.0646057
	Treatment	17	14.5294		
Percent Multiple Races	Comparison	17	3.0000	0.8235	0.2134346
	Treatment	17	2.1765		
Percent Native Hawaiian/Pacific Islander	Comparison	17	0.0000	NA	NA
	Treatment	17	0.0000		
Percent White	Comparison	17	1.9412	0.9412	0.3248127
	Treatment	17	1.0000		

*Significant after correcting for False Discovery Rate

Bold: Higher mean

We collected data during the 2014–15 school year. While teacher surveys were administered to all schools as part of a district-wide initiative, our sample included a total of 34 schools: 17 treatment schools which are implementing specific improvement models and 17 matched comparison schools. The District turnaround initiative rolls out in yearly cohorts. Our quantitative sample consisted of all schools from the first (Group 1) and fourth (Group 2) cohorts of the initiative. Group 1 schools were “seasoned” implementers and were in their fifth year of implementing their turnaround model. Group 2 schools were in their second year of implementation of their turnaround model in the 2014–15 school year. Thus, we were able to study schools that were just starting to implement and those that

had been implementing for several years, to examine a trajectory of implementation, and also to look at how implementation and effectiveness might differ based on the amount of time the school had engaged in turnaround (Berends, 2000; Cohen & Ball, 1999; Murphy & Bleiberg, 2019; Meyers & Smylie, 2017; Player & Katz, 2016). Within the treatment group there are 9 Renaissance Charters (six in Group 1; three in Group 2), 8 Promise Academies (four in Group 1; four in Group 2), and 17 matched comparison schools. Three Renaissance Charter operators were represented: Mastery Charter Schools (Mastery), Universal Companies (Universal), and Young Scholars. We refer to each of these operators as well as the Promise Academies as an “improvement model.” Both Renaissance Charters and Promise Academies were required to hire new principals,³ replace at least 50 percent of their staff, and create a comprehensive plan for improvement. The Renaissance Charters are operated by external operators and Promise Academies are operated by the District.

A qualitative study undertaken by the District in 2013–2014 sought to describe key difference among each operator (Stratos et al., 2014). Mastery schools operate using a no-excuses model that implements rigorous, Common Core aligned curricula consistently across its various schools; Universal takes a more community-based approach to reform, giving school leaders greater autonomy over policies; Young Scholars is a data-driven model that focuses on college-readiness, positive student behavior, and extended learning time; and the Promise Academy model entailed minimal school-based autonomy over instruction and climate policies, additional instructional and professional development time, and prescribed intervention programs (Stratos et al., 2014).

To supplement the teacher survey data, we conducted site visits and interviewed principals and teachers at six schools selected from the larger sample of 34 schools. The goal of these site visits and interviews was to capture teachers’ and principals’ perspectives on schools’ unique approaches to improvement and what successes and challenges they associate with their schools’ models of reform. We focused our qualitative data collection on *only those schools in Group 1* since they had five years of experience with the reform and therefore were the most likely to exhibit features of their given turnaround models. Additionally, prior research had shown that the Promise Academies from Group 1 were the only Promise Academies that maintained some semblance of the Promise Academy model after extensive budget cuts and instability in the District (Stratos et al., 2014). Within the Group 1 schools, we narrowed our focus to two schools per model. The inclusion of two schools per turnaround model helped us differentiate between features unique to a school and those that are representative of the particular turnaround model. Specifically, we studied two schools operated by Mastery, two operated by Universal, and two district-operated Promise Academy schools. Young Scholars did not have two schools in Group 1, and therefore was excluded from our qualitative sample. In addition to Mastery and Universal being the only charter providers in Group 1 that met our criteria of

³ One school in our sample did retain its principal post-turnaround.

having at least two schools, they are also the two largest providers of elementary school restarts in the District.

In the District, the number of elementary/middle schools engaging in improvement far exceeds that of high schools. Research that has found notable differences in effective reforms for high school and elementary schools (e.g., Edgerton & Desimone, 2019; Firestone & Herriott, 1982; Lee et al., 1993). To best facilitate meaningful comparisons, this study focuses on elementary schools only.

Data

The data for this study draw on a larger data collection effort undertaken by the Shared Solutions Partnership. Data sources include achievement data, administrative data, survey data, as well as interview data.

Achievement data

The primary outcome variable in our quantitative analyses is school-level achievement on the Pennsylvania System of School Assessment (PSSA). Grade-level mean scaled scores were used to control for prior achievement, and state-reported PSSA proficiency bands were used to construct our school-level outcome variables. Proficiency bands are used by both the Pennsylvania Department of Education (PDE) and the District to evaluate the performance of schools and make accountability decisions. For example, it is proficiency bands that are used in identifying schools for turnaround or closure. This routine and institutionalized use of proficiency bands for accountability purposes makes them a practical outcome measure for our study. English Language Arts (ELA) and math outcome variables for the analysis were constructed with a two-step process using the state's 2014–2015 achievement data. First, we multiplied the percent of students scoring in each proficiency band by the numbers below and summed them to create grade-level PSSA scores for ELA and math, separately: Below Basic = $\times 1$; Basic = $\times 2$; Proficient = $\times 3$; and Advanced = $\times 4$. In the second step, we created a school-level achievement outcome variable for ELA and math, separately, by taking the mean of the grade-level scores created in the first step.

Administrative data

We included covariates in regression models to mitigate bias from the estimates of treatment effect (Steiner et al., 2010). In addition to measures of prior achievement (McCoach, 2006), we included variables to capture student race, ethnicity, gender, economic disadvantaged status, and indicators for English Language Learners and students with learning disabilities. Noting constraints on power due to a small sample size and a heavily specified model and being mindful of the risk of overfitting (this is when a model begins to describe random error and reduces generalizability) (Babayak, 2004; Breiman & Freedman, 1983), baseline analyses informed the creation of an index of IEP/ELL, which captures the percent of students who are identified as

English Language Learners (ELL) and those identified as needing an Individualized Education Plan (IEP) to include along with percent white students as covariates in the final regressions. IEP/ELL is a school-level variable equivalent to the percent of students who are identified as ELL at a school plus the percent of students with IEPs.

Survey data

We use data from a district-wide teacher survey that was jointly developed as part of the Shared Solutions Partnership. This survey was administered to all teachers in the District online via an e-mail link between May 18th and June 19th, 2015. Teachers were not required to participate in this data collection, and e-mail/web-based surveys typically have lower response rates (e.g., Mertler, 2003; Sheehan, 2001). We used best practices to help boost the response rate within the study sample such as as: crafting a well-designed and clear survey; prominently featuring the District and University logos on the survey; offering financial incentives; and making repeated contact and follow-ups with participants via e-mail and the phone (Fowler, 2014). Within the study sample, the average response rate for a school was approximately 63 percent of teachers, with a low of 11.9 percent and high of 96.67 percent (standard deviation equal to 23.22).⁴

Survey questions were drawn primarily from surveys with known validity and reliability, such as those used by Bryk et al. (2010) in Chicago. We used expert review, focus groups, and cognitive interviews to further refine the questions and items (Desimone & Le Floch, 2004). Survey development took approximately six months, and each survey went through over 10 rounds of intensive review and revisions.

In total, this survey measured 10 key constructs (5 essential supports and 5 policy attributes) and 19 sub-constructs total within the essential supports and policy attributes. Each construct was comprised of multiple items, which increases validity and reliability (Mayer, 1999). Many of the constructs included sub-constructs. For example, the construct “School Climate” consisted of the sub-constructs “Bullying,” “Respect,” and “Challenges.” To view a full list of the constructs, sub-constructs and their reliabilities, please see Table 2.

Using the survey data, we created school-level scores to measure the essential supports, policy attributes, and their sub-constructs. All survey items were on Likert scales, and were scored on 0 to 3- or 0 to 5-point scales (depending on the total number of answer categories), with zero indicating the lowest (or most negative response).⁵ When appropriate, items were reverse coded. To create measures of the essential supports and policy attributes for our analyses we used a three-step process. First, recalling that our 10 (total) essential supports and policy attributes constructs are made up of 19 (total) sub-constructs, we created teacher-level scores for each sub-construct on our survey by averaging the scores on each item within a sub-construct. Next, we created teacher-level construct scores by averaging a teacher’s

⁴ This is higher than the response rate for the District. A total of 5423 teachers (53 percent) responded to the survey.

⁵ Most commonly used answer categories were: Strongly Disagree, Disagree, Agree, Strongly Agree and Never, 1–4 times a year, 5–7 times a year, Monthly or about monthly (8 or 9 times a year), Weekly or about weekly, Daily or almost daily.

Table 2 Constructs/Subscales

School leadership	Items	Cronbach's alpha
Expectations and feedback	5	0.939
Inclusive leadership	5	0.945
Overall	10	0.961
Parent/Guardian-community ties	7	0.896
Professional capacity		
Innovation	5	0.878
Peer collaboration	7	0.843
Quality of PD: Content focus and duration	11	0.883
Quality of PD: Collective participation, active learning, and learning span	6	0.69
<i>Overall</i>	29	0.895
Climate		
Bullying	6	0.745
Respect	10	0.678
Challenges	10	0.829
<i>Overall</i>	26	0.871
Instruction		
Procedural	5	0.673
Conceptual	7	0.843
Student engagement	7	0.804
Teacher self-reflection	12	0.789
<i>Overall</i>	31	0.865
Specificity (around instructional guidance)	11	0.84
Consistency		
Curriculum and instruction	5	0.781
Professional development	5	0.734
<i>Overall</i>	10	0.792
Authority		
Institutional	13	0.892
Normative	6	0.835
<i>Overall</i>	19	0.897
Power	2	0.228
Stability		
Mobility/turnover	6	0.743
Student absenteeism/tardiness	3	0.917
<i>Overall</i>	9	0.847

scores for each of the constructs' sub-constructs. Lastly, we created school-level scores by averaging all the teachers' scores for the constructs and sub-constructs, respectively. We calculated Cronbach's alphas for each of the constructs and sub-constructs to assess the internal consistency of the survey items. All scale reliabilities for this study, with the exception of one (for Power), fell within the 0.67 and

0.961 range (average = 0.811), indicating an acceptable internal consistency between items within each construct and sub-construct, without item redundancy (Gliem & Gliem, 2003; Nunnally & Bernstein, 1994; Streiner, 2003). The construct “Power” had an alpha of 0.228, which is not unusual as it was a two-item indicator and alpha is, in part, a function of the number of items in a scale (Cortina, 1993).

We checked for multicollinearity within and between the essential supports and policy attributes and in cases in which correlations were found to be high (defined as greater than 0.8), we revisited the survey questions to ensure that they were measuring different components (conceptually) and also to check where the items fell on the survey, as items asked near one another are more likely to be correlated. For correlations that exceeded 0.9, we additionally looked at the scatterplots at both the school- and teacher-level, which confirmed that variables are not collinear and did not pose a threat to our analysis.

To account for partial response bias and help ensure validity of the survey scores, we required that teachers respond to a minimum of 50 percent of items within a sub-construct in order to be eligible for a sub-construct score. Similarly, for a teacher to receive a construct-level score, they must have a score for at least half of the sub-constructs that comprise the construct. This decision considered response descriptives as well as the number of items within sub-constructs, how many survey questions a sub-construct spanned, and how many dimensions of a construct we measured. The survey was designed to be precise and efficient—we used the minimum number of items necessary to capture each sub-construct and construct so as to reduce teachers’ burden of taking the survey. Due to complexity of the carefully chosen items that make up each sub-construct (and of sub-constructs that make up each construct), weighting and imputation did not seem appropriate (Brick & Kalton, 1996). The threshold of 50 percent was chosen to maximize use of available data while maintaining construct validity.

Interview data

We visited each of our case study schools in the spring of 2015. In keeping with the standard of research and best practices set by the District’s qualitative study during the prior school year and to ensure minimal interruptions to schools’ daily activities, all interviews and observations for a given school took place on the same day. Comprehensive Site Visits, lasting between approximately four and seven hours were coordinated with each school’s principal by the District. To reduce selection bias, increase consistency within our sample, and also ensure a more representative sample of interviewees at each school, we asked that principals arrange interviews with: (1) A teacher in a leadership position; (2) A 3rd grade teacher; (3) A teacher in lower-elementary (Grades K-2); and (4) A teacher in upper-elementary (Grades 4–8). These criteria were chosen based on hypotheses that teachers in tested grades and subject areas may have a different experience at school than those in non-tested grades and subjects; similarly, teachers who are new may have a different experience than those who are more experienced.

The interview protocol was structured according to the five essential supports and included questions that spoke to each of the five policy attributes. We asked principals and teachers questions about their school's overall approach and philosophy for each essential support; their perspective on their school's strengths and challenges in each of these areas; and, additionally about the policy attributes. The interviews were semi-structured in nature in order to ensure that the participants speak to aspects of school improvement that we were interested in learning about, while also allowing for new information about implementation and effectiveness to be brought to light (Weiss, 1994). Interview protocols went through multiple revisions with Shared Solutions staff. Each interview lasted approximately 45–60 min. We interviewed a total of 24 teachers and 6 principals in our six case study schools—4 teachers and 1 principal per school.

Analytic approach

Quantitative

To determine which school improvement models, key components, and implementation strategies are associated with higher performance on the PSSA, we used blocked, stepwise regression models that included indicators for school improvement model type, prior academic achievement, percent of the student population that is white, and a composite variable (IEP/ELL) reflecting the percent of students who are identified as ELL and those identified as needing an IEP. At the start of the study, we examined what we term “baseline” differences between the treatment and comparison groups on these key demographic and achievement variables to guide us in selecting variables to include in our models.

In our regressions, we used a blocking method, which allowed us to specify different regression approaches for different parts of the model. Because we are interested in examining the relationship between various school improvement models and achievement, we entered a dummy variable for school improvement model type (Mastery, Universal, Young Scholars, and Promise Academy) into the equation. Furthermore, noting the importance of our covariates, in estimating our regressions, we entered baseline achievement data, percent white students, and our IEP/ELL variables. These variables comprise the first block of the regression equation and were included in all regression models estimated for this study in order to disentangle the effects of the essential supports and policy attributes from those of the model and aforementioned covariates.

For the next block, we entered all 10 of the main survey constructs, and specified stepwise regression. This type of regression is used when the ratio of the number of variables to sample size is high and also to find which predictor variables most efficiently explain variance in an outcome variable (Babyak, 2004; Leigh, 1988; Streiner, 2013). Stepwise regression does not favor any variable, but rather uses only statistics to choose which variables enter into an equation; this lack of theory and strict reliance on math has led some to criticize its use (Thompson, 1995). In our case, this study does not purport to be causal, but rather is exploratory in nature

and because—to our knowledge—there is no theoretical or hypothesized basis to presume that certain essential supports or policy attributes should be favored over others, stepwise regression was deemed the most powerful and preferred approach for detecting significant effects and better understanding the relationship among the various predictors.

We supplemented our regression analyses with multivariate analyses of covariance (MANCOVAs) and descriptive statistics with the goal of identifying each school improvement model's strengths and challenges as measured by the constructs on the teacher survey. The MANCOVAs additionally allowed us to explore meaningful differences among the models on these variables.

Qualitative

We analyzed the qualitative data in two phases. First, interview transcripts were coded on a rolling basis in the order in which the site visits were conducted. All interviews were first deductively coded (Miles et al., 2014) using the five essential supports and five policy attributes. We then used inductive coding to identify themes in the data that could help to explain successes and challenges in essential supports and levels of policy attributes.

The purpose of closed coding the interviews using the essential supports and policy attributes was to provide rich descriptions of each school improvement model's key components and strategies for their implementation. Our survey analysis identified the relative strength of the essential supports and levels of policy attributes present in schools implementing different improvement models. As we went through the interview transcripts during phase one of coding, we also added more codes to our coding structure in order to capture aspects of teachers' and principals' experiences with school improvement models that were not already accounted for by our conceptual framework (Miles et al., 2014). This open coding shed additional light on the successes and challenges associated with implementing the schools' improvement models. New codes that emerged during analysis included: capacity building, autonomy, adaptation, collaboration, communication, resources, responsibility, support, trust, and values/philosophy.

For our analysis, we drew on the results from the quantitative analysis of survey and academic achievement data to narrow our focus within the qualitative data, considering only parts of interviews coded for the essential supports and policy attributes that were identified as significant in our regression models. This analysis was conducted at the operator and not the school level. The purpose of the study is to understand different school improvement models, namely the Mastery, Universal, and Promise Academy models for improvement in terms of the essential supports and policy attributes and *not* to explain an individual school's approach to improvement. To conduct this analysis, we ran queries conditioned on school operator and the construct of interest—a particular essential support and/or policy attribute. Next, we inductively coded within these selections to identify themes in the data that could help to explain successes and challenges in essential supports and levels of policy attributes.

Limitations

While we did equate groups on critical characteristics known to be associated with student learning, our study is not causal. Unobserved and unmeasured differences between the study groups weaken the internal validity, and this should be kept in mind in interpreting results. At the same time, equating groups on critical characteristics known to be associated with student learning contributes to the confidence that any relationships we find may be causal. Additionally, lack of access to both student-level achievement data and grade-level mean scaled scores is a limitation of this study. We made numerous attempts to acquire this data from the state and to collaborate with the District's charter school office to collect achievement and administrative data for the Renaissance Charter schools but none of these efforts were fruitful.

While the policy attributes can permeate all aspects of a school improvement model, each of the policy attributes was only asked about within the context of a single essential support. This was an effort to capture maximum variance using a minimum number of items to keep our survey concise and minimize burden on teachers. Finally, while we did provide principals with specifications to reduce selection bias and ensure a more representative sample of interviewees at each school, we acknowledge that principals may be biased in their choice of teachers.

Results

For the sake of simplicity and to minimize redundancies in this paper, we restrict interpretation of our results to our regression model for grades three through eight in English Language Arts (ELA). The ELA model had the best fit, the R-squared values for the models predicting ELA achievement were consistently higher than those predicting math achievement, and math achievement tended to be lower for schools regardless of model. Differences in ELA achievement were on average two times the size of differences in math achievement. A potential explanation for schools' higher performance in ELA is the intense focus that the District has placed on ELA achievement. One of the District's official goals outlined in their improvement plan is that "100% of 8 year-olds will read on grade level" (The School District of Philadelphia, 2015); there is no comparable goal for math.

Also, in interpreting our results, it is important to note that prior to implementing a specific improvement model, all schools in our treatment group were performing significantly worse in ELA and math than the comparison schools. Further, there were no significant differences in student populations served, as measured by percent white students or our IEP/ELL index, for any of the school improvement models from the baseline year to the year of data collection.

RQ1: What are the relationships between the various approaches to school turnaround and academic achievement?

On average, regression models conditioned on school improvement model, controlling for prior academic achievement, percent white students, and IEP/ELL, explained approximately 68.4 percent of variation in academic achievement. In general, when individual school improvement model types are included as predictors, inclusion of the essential supports and policy attributes does not explain additional variance in academic achievement. Because the essential supports and policy attributes drop out of the regression models when we include indicators for the individual school improvement models, we hypothesize that these features capture the essence of a school's approach to improvement.

As shown in Table 3, Mastery was found to be a significant predictor of achievement. This means that schools being turned around by Mastery were associated with higher levels of student achievement in ELA. Specifically, if a school is engaging in Mastery's school improvement model, it is predicted to score approximately 51 points—or 13 percent—higher in ELA achievement ($p=0.000$), holding percent white students, IEP/ELL, and prior achievement constant. Universal is also outperforming the comparison schools in ELA ($p=0.006$) such that if a school is engaging in Universal's school improvement model, it will score approximately 21 points—or 5 percent—higher on our ELA outcome variable. While not significant ($p=0.693$), Promise Academies are performing 1.97 points lower in ELA achievement as compared to comparison schools. In this model, prior ELA achievement is a significant predictor of current ELA achievement ($p=0.002$), holding all else constant such that a one standard deviation increase in prior ELA achievement is associated with a 10.14-point increase in performance on our outcome variable. Percent white students is also a significant predictor ($p=0.009$) such that each one percent increase is related to a 2.95-point increase in ELA achievement.

Robustness check. To address concerns about response biases due to differential response rates, we did a robustness check on our analyses by re-running the regressions for only schools that had a response rate of 50 percent or higher. Results were found to be robust to response rates: Mastery remained a significant predictor for all outcome variables and Universal remained a significant predictor of the all grade achievement outcomes.

RQ2: To what extent do essential supports and policy attributes mediate the relationship between school turnaround and academic achievement?

Hypothesizing that the essential supports and policy attributes may not enter into the final stepwise regressions because the school improvement model indicators designating Mastery, Universal, Young Scholars, and Promise Academy effectively capture particular essential supports and policy attributes, we ran the same stepwise regression again but replaced the indicators for each school improvement model with a variable for “Treatment Group.”

Table 3 Summary of regression model output

Outcome	Control			Key Predictors												
	R2	Adjusted R2	Treatment Group	Percent White	IEP/ELL	ELA Baseline	School Leadership	Parent/Guardian-Community Ties	Professional Capacity	School Climate	Instruction	Specificity	Consistency	Authority	Power	Stability
ELA Achievement (PSSA All Grades: 3-8)	0.675	0.587	-3.974	3.975*	-0.414	0.079	x	x	x	x	x	28.123*	x	27.009*	18.834*	x
ELA Achievement (PSSA All Grades: 3-8)	0.808	0.756	Mastery	Percent White	IEP/ELL	ELA Baseline	Universal	2.952*	-0.222	10.142*	Young Scholars	9.13	-1.973	Promise Academy		

Key
 * sig at .05
 . sig at .1
 x excluded from model

In this revised regression, Treatment Group was not a significant predictor of achievement, however, three policy attributes entered into the regressions as significant predictors of our outcome variables: Specificity, Authority, and Power. All three of these policy attributes were statistically significant in the ELA achievement model for all grades (three through eight). Of the three policy attributes that entered into the equations—by a slight margin—Specificity was associated with the largest improvement in achievement such that a one-point increase in Specificity (out of total of three points) is associated with a 28.12 point (approximately 7 percent) increase in ELA achievement ($p=0.022$). Similarly, a one-point increase in Authority (out of total of three points) is associated with a 27 point (6.75 percent) increase in ELA achievement ($p=0.011$). Power was found to be a significant predictor at the 0.10 level ($p=0.065$) such that a one-point increase in Power (out of total of three points) is associated with an 18.83 point (4.7 percent) increase in ELA achievement. These results will be further contextualized and discussed in the following sections on “levers” for improvement.

RQ3: How do teachers’ and principals’ descriptions of their schools’ approaches to improvement map on to the essential supports and policy attributes, and help to explain the relative successes and challenges of each school improvement model?

As Table 3 shows, we found that it is the policy attributes and not the essential supports that have the strongest relationship with student achievement. While the essential supports were not found to be significant in our regression analysis (and as such, for the purpose of space, we do not present our qualitative findings on them), they provided the structure through which we were able to ask teachers and principals about their schools’ approach to improvement. For full qualitative results, including findings on the essential supports, see Hill (2016). While all five policy attributes can be used in any one reform effort—e.g., climate efforts can be specific, consistent, authoritative, powerful, and stable—we asked about each of the attributes in the context of a single essential support, due to the necessity of making our survey effort efficient and concise.

Our data help us better understand how different school improvement models are implemented. Table 4 provides an overview of which models scored “high” and “low” on each of the significant policy attributes. The column, “True Max,” indicates the highest score a model could receive for each of the constructs and sub-constructs. As evidenced in Table 4, the Mastery model—the school improvement model that was significantly predictive of academic achievement—scored the highest in all the policy attributes identified as significant by our regression models. Likewise, the Promise Academy model, which was consistently, though not significantly, related to lower academic achievement, tends to score low in the policy attributes. The Universal model tends toward the middle of the policy attributes scale.

We believe that understanding these attributes is vital to creating successful models for school improvement. We use our interview data to explore what these

Table 4 Policy Attributes: Highs and Lows

	Promise Academy	Universal Companies	Mastery Charter Schools	True Max	Comparison Group Mean (for reference)
	Mean (n=8)	Mean (n=3)	Mean (n=4)		
<i>Specificity</i>	0.602634	0.808621	1.113416	3	0.465365
<i>Authority</i>	1.596975	1.686462	2.093698	3	1.520361
<i>Power</i>	1.359311	1.347684	2.068490	3	1.359411
Key	Low				
	High				

*Sig at .05

’Sig at .1

X excluded from model

significant policy attributes *look like in practice* so that we can think critically and practically about how schools can use these attributes to generate gains in student achievement. We draw on data from our six site visits to illustrate what it may look like to be “high” and “low” in the policy attributes that regressions identified as significant predictors of student achievement. Using these combined survey and interview results we identify three levers for school improvement.

Lever one: specificity: striking a balance between directives and discretion.

In this study, the policy attribute Specificity was measured within the context of the essential support of Instruction. To gauge the level of Specificity used in a school’s approach to instruction, teachers were asked about how much control they had over different aspects of instruction in their classrooms—for example: determining course objectives, choosing books and other materials, selecting content, and setting pacing. When we identify a school improvement model as more specific, what we mean is that the model provides teachers with clearer guidance, materials, and training that leave less room for individual interpretation of school policies and practices. It is important to note that Specificity, like all the policy attributes, is a continuum with both advantages and disadvantages at both ends. For instance, a highly specific model that provides teachers with clearer, materials and training may improve implementation fidelity but leave less room for teacher creativity and autonomy. Such nuances make additional qualitative work—such as interviews—critical to interpretation of these results.

As shown in Table 4, schools operated by Mastery scored the highest on specificity (1.11) and Promise Academies scored the lowest (0.603). Practically, this is the difference between more teachers indicating that they have “a great deal” of control and more teachers indicating that they have “a little control” with regard to instruction. For all Specificity survey items, see Table 5. Despite the difference between these scores, teachers at Mastery schools and Promise Academies often used similar language to describe their school’s prescription for instruction.

Teachers at Mastery schools described a uniform instructional policy and prescribed strategies and materials that they used in their classroom but emphasized that they have a degree of discretion and flexibility in implementing these prescriptions as they see fit. Adele, a Mastery teacher, said:

It’s literally like they have unit plans for us. They make it up for us, they give us the curriculum, but we have free rein because we know our kids. And that’s what I love about Mastery is the fact that they acknowledge that you know your kids, you know what they need.

Mastery teachers noted how the Mastery model had recently increased flexibility for teachers, transitioning from a very rigid sequence of lesson delivery to increased choice and more engaging instruction. Lauren, a Mastery teacher, said:

Last year...it was either the Mastery way or the highway... This year, it seems that they’re embracing more individuality with the teaching ... So, last year was, you have to do it this way and lessons must look like this, this, this, and this. This year there is a structure to the lesson, but they, they do leave some room for individual growth, for you to express yourself as a teacher...

Like Mastery teachers, Promise Academy teachers also highlighted language around mandates, required curriculum, and needing to follow standards—in fact, in interviews they referenced this type of specificity to a *greater* extent. However, while teachers were likely to mention that while their school improvement model had specific guidelines, they noted that they were not likely to always follow them, rather they’ve “learned to kind of veer off to what works best for our kids” (Jane, Promise Academy Teacher). Similarly, Maxine, another Promise Academy teacher, explained that:

[E]ven though it’s mandatory, we still have so much to work with, we can still kinda go off on our own. Find other resources. You know, I can do centers in the middle of a lesson. I mean I just do whatever is best for my kids. But we have so much like wiggle room to kind of, you know, do what we need to do for our students.

Respondents from both models describe an instructional environment that has structure but that they flex within in order to meet the needs of their students. A careful reading of the interviews, though, highlights that Mastery teachers appeared to feel their model provides prescriptive guidance but allows for flexibility, while the Promise Academy model may in theory be very specific, but teachers make the choice not to comply with the model’s requirements, as evidenced in the quotes

Table 5 Survey items**Specificity**

How much control do **YOU** have over the following in your class:

- Determining course objectives
- Choosing books and other instructional materials
- Selecting content, topics, and skills to be taught
- Selecting the sequence in which topics are covered
- Setting the pace for covering topics
- Determining how classroom space is used
- Setting standards of behavior in my classroom
- Choosing the teaching methods and strategies I use with my students
- Determining the amount of homework to be assigned
- Choosing criteria for grading students
- Choosing the evaluation and assessment activities used in my class

Institutional authority

To what extent do you consider each of the following factors a challenge to student learning in your school?

- Shortage of highly qualified teachers
- Lack of high-quality professional development opportunities for teachers
- Shortage of instructional support staff (e.g. teacher aids and reading specialists)
- Shortage of other support staff (e.g. nurses, counselors, and security)
- Lack of teacher planning time built into the school day
- Lack of support for teaching special education students (i.e., students with IEPs)
- Lack of support for teaching English Language Learners
- Lack of school resources to provide the extra help for students who need it
- Lack of computers or other technological resources
- Lack of support from parents and guardians

To what extent do you consider each of the following factors a challenge to student learning in your classroom?

- Insufficient class time to cover all of the curriculum
- Inadequate textbooks, materials, or other non-technological instructional resources
- Lack of computers or other technological resources

Normative authority

To what extent do you agree or disagree with the following statements?

- My school has clear strategies for improving instruction
- Teacher morale is high

To what extent do you agree or disagree with the following statements?

- Teachers at my school have high expectations for students
- Teachers at my school support the idea that all students can learn
- Teachers feel responsible when students in my school fail academically
- Most of my colleagues share my beliefs and values about what the central mission of the school should be

Power

To what extent do you agree or disagree with the following statements?

- My school or District/CMO recognizes or rewards me based on my teaching and/or student achievement

Table 5 (continued)

 My school or District/CMO penalizes me based on my teaching and/or student achievement

above. This may be because the Promise Academy is a District-managed model which has less authority (buy-in) than externally managed models which tend to choose teachers who know about and believe in the model.

Levers two and three: Institutional and normative authority.

The District-wide teacher survey included two sub-constructs within Authority: Institutional and Normative. In order to delve more comprehensively into this significant finding, we consider each sub-construct separately.

Institutional authority: access to resources. The survey asked questions about teachers' perceptions of the extent to which various shortages of resources were a challenge to student learning in their schools and classrooms. Resources asked about on the survey included: staff, materials (such as textbooks and technology), and time. Please see the Table 5 for a list of the questions. The greater the challenges that teacher reported due to lack or shortages of these various resources, the lower the model's score for Institutional Authority.

Mastery turnaround schools had the highest score for institutional authority (1.89) and Promise Academy had the lowest (1.28), which reflects the difference between more teachers reporting that a lack of various resources is “a slight challenge” as compared to more teachers reporting that this lack of resources is “a moderate challenge.” Our interview data show that what teachers and principals identify as challenges in terms of supports offered by their school is relative to their school context and how they define “a challenge.” For example, Lauren, a Mastery teacher, laments about the challenge of having only “two or three deans on our floor.” Kylie, another Mastery teacher also referenced human resources as a challenge, explaining: “I feel like if we had a little bit more manpower it might feel a little less like daunting on us, cause if we have like a difficult student in the classroom sometimes it is hard to teach the whole class...” In contrast, a Promise Academy principal noted that the school does not even have one full time police officer but they “make the best of it.”

Overall, feedback from Promise Academy staff was mixed, emphasizing inconsistencies in how the model is implemented. Some staff emphasized both a lack of personnel as well as a dearth of material resources, as Darcy explains:

I think the challenge is also not always having the materials that are necessary to meet the needs of those students. You know, when you have children that range in level from C to M [denoting reading levels] you have to be able to pull a great variety of materials. We do have a really good lending library here which was, you know, put together for us, but sometimes there's just not enough bodies and there's not enough materials to be able to reach everybody effectively.

At the same school, another teacher, Maxine noted: “Prior to coming to a Promise Academy, I felt like I had no support with regards to resources and behavior with the students; but being here, it’s a different, it’s night and day.”

As Institutional Authority was found to be a significant predictor of student achievement, these findings suggest that rather than the actual, finite amount of resources available at (or to) a school, it is teachers’ and principals’ perceptions of the resources and their mindset toward challenges that is critical for improvement. This finding is consistent with Desimone’s (2002) adaptation of policy attribute theory that focuses on *perceptions* of policy attributes rather than their true levels and her argument that “it is district, principal, and teacher knowledge and interpretation of the attributes that directly influence practice” (Desimone, 2002, p. 440).

Normative Authority: Buy-in and presence of a strong professional culture. The District-wide teacher survey asked questions about the extent to which teachers agreed or disagreed with various statements about the professional culture at their school. Particular areas of professional culture asked about included teacher morale, shared beliefs and values, and responsibility. The more teachers who strongly agreed with these statements, the higher the model’s score for Normative Authority. Please see Table 5 for a list of the questions.

Mastery was again the highest scorer at 2.28, compared to the lowest score from Universal at 1.93. This is the difference between more teachers indicating that they “strongly agree” with these statements about their school’s professional culture and more teachers indicating that they “disagree” with these statements. At Mastery schools, it was mainly the principals that spoke to the presence of Normative Authority. Interview data suggest that teachers are hired because they are a good fit with the organizational values, which strengthens Normative Authority. A Mastery principal explained:

I think Mastery has such a rigorous process to want more people, to make sure that you, you become a part of the organization, but I do think most people that are here are very committed to the students and the families.

A second important theme was the degree to which teachers mentioned the development of relationships within their schools, as Mastery teacher, Tim, explains, “these people are really family from the, from the basement all the way up to the 3rd floor.”

At Universal schools, there were inconsistencies in the interview data, with some teachers reporting shared values and expectations and others expressing the opposite. Leo, a Universal teacher, said:

I think you kind of have a split here. I think you have the majority of the staff that has a shared vision. Most of us who have that shared vision have been here for the five years. You do have your dissenters. You have some people who you know, kind of work in isolation, and that works best for them.

The conflicting views on this indicator may explain why the Universal model scored low in Normative Authority on our survey.

Lever Four: Power (incentive systems).

To gauge the level of Power used by a school in implementing their improvement model, our survey asked teachers about the extent to which they agreed or disagreed with two statements—one about rewards and one about penalties. Higher scores in Power imply that more teachers at the school agree that their school will reward or punish them based on their teaching or student achievement. See Table 5 for the survey questions.

The models scoring the highest and lowest on Power were Mastery (2.07) and Universal (1.35), respectively. More teachers at Mastery schools strongly agreed that there were rewards and/or penalty systems in place, whereas more teachers at Universal Schools disagreed. Mastery was the only model in our study described as having an explicit incentive system in place—monetary incentives were linked to instruction, the most specific aspect of the model.

It noteworthy that a simple two-item indicator was found to have such a significant relationship with academic achievement (see Table 3). This finding is even more noteworthy when one considers our qualitative data, which suggest that higher scores on Power were driven by the presence of rewards rather than penalties. Overall, in our interviews, teachers and principals at all models maintained that rather than penalties, schools focused on how to support teachers in improving, suggesting that it is the use of Power in the form of rewards that is beneficial for school improvement. Mastery used a performance-based pay scale that ties salary to teaching evaluations. In addition to monetary rewards, teachers at Mastery also referenced verbal recognition and praise as types of rewards.

At schools implementing the Universal model of improvement, teachers were inconsistent. Sam, a Universal teacher, explained:

But I mean they give us rewards, like here like all the time on different things because they know how hard we work, they know what we're going through, so like our, and our administration is great with that, with rewarding us.

Conversely, Leo reported:

There's nothing. I mean I think that's why most of us don't really care, because there's nothing attached. I mean if you do what you're supposed to do, you get a contract for next year. I mean I guess the only penalty would be not to receive a contract.

In terms of penalties, Mastery teachers were at a general loss—suggesting that perhaps not getting the rating you wanted was a penalty. Across all models, teachers and principals maintained that schools focused on how to support teachers in improving, not penalties.

Discussion

Lessons in turnaround from other sectors emphasize the importance of being systematic in distilling key components of initiatives and seeking to understand their successes and failures (Murphy, 2010). The semantics of various reform efforts both hide and confound commonalities, severely limiting what we can learn from the many actors engaging in innovative approaches to school improvement around the world. This study makes important contributions to the research on school turnaround by providing a conceptual framework that can be used to understand and systematically investigate turnaround efforts (McCauley, 2018; Meyers & Smylie, 2017).

Rather than simply declaring Mastery a successful turnaround, our Integrative Framework for School Improvement enabled us to identify Specificity, Authority, and Power as possible explanations for the higher achievement of Mastery. Furthermore, our mixed methods approach allowed us additional insights into these potential mediating variables. Revealing, for instance, that high Specificity did not preclude flexibility and that with regards to Institutional Authority, actual level of resources and challenges were less important than respondents' comparative view of resources, and their mindset about challenges.

In addition to providing a framework for thinking about what models are doing to generate improvement and how they are doing it, we offer diagnostic tools (survey instruments and interview protocols) that researchers and practitioners alike can use to identify, measure, and assess the relative contribution of the essential supports and policy attributes for *any* school improvement effort. This work is complementary to that undertaken by The Center for School Turnaround (2017) which identified four domains for rapid improvement (which overlap with the five essential supports) and provided recommended practices and questions for self-reflection based on analysis of existing literature on effective schools and school improvement.

Below we outline what we believe are the two key takeaways from our analysis:

1. *Do not get caught up in labels, in order to improve schools we must focus on the components of reforms:* In studying school turnaround, it is the theory of action that matters, not nomenclature. By grouping various approaches to school turnaround together just because they share the same label, critical differences in outcomes as well as predictors are lost. To conduct the most accurate and meaningful evaluations of school turnaround, researchers must look to the key components and implementation attributes of a model rather than its namesake as these will provide far more insight into whether or not the model successfully generates improvements and how it does so.
2. *How schools supported their efforts (i.e., policy attributes) drove differences in achievement; not the essential supports:* All schools were engaging in essential supports in different ways; however, it wasn't the type of support or variation in these supports that predicted differences in student achievement, rather, it was the way they implemented their efforts in terms of the policy attributes. Specificity,



Fig. 2 School improvement thumbprint

Authority, and Power emerged as being significant predictors of student achievement. In essence, we can think of each model as having a unique improvement thumbprint that captures its relative successes and challenges with regards to its key components and implementation strategy (see Fig. 2). This thumbprint can easily be used diagnostically to identify areas on which a school improvement model has room for growth or to highlight areas of strength and creates an accessible jumping off point for discussions about what various school improvement models look like.

Joe Bower (2015) writes: “the root word for assessment is *assidere* which literally means ‘to sit beside.’ Assessment is not a spreadsheet—it’s a conversation.” The assessment of school turnaround efforts in the District presented in this study does not intend to make sweeping claims about school improvement or offer up a formula for success. Instead, this study offers a critical, thoughtful, and data-driven framework to serve as a starting point for important conversations about what school improvement looks like and how we can best facilitate it. It is our hope that future research will leverage this framework in assessing models of not just turnaround efforts but all school improvement efforts. By understanding not only what a school is doing (essential supports) but also how it is doing it (policy attributes), we believe researchers and practitioners will be better equipped to identify levers for improving all schools.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- American Institutes for Research. (2011). *School turnaround pocket guide*. Retrieved from <http://www.air.org/resource/school-turnaround-pocket-guide>
- Ayers, J., Owen, I., Partee, G., & Chang, T. (2012). *No child left behind waivers: Promising ideas from second round applications*. Center for American Progress.
- Babayak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine*, 66(3), 411–421.
- Berends, M. (2000). Teacher-reported effects of new American school designs: Exploring relationships to teacher background and school context. *Educational Evaluation and Policy Analysis*, 22(1), 65–82.
- Bower, J. (2015). *Assessment and measurement are not the same thing*. Retrieved from <http://www.joeboer.org/2015/12/assessment-and-measurement-are-not-same.html>
- Breiman, L., & Freedman, D. (1983). How many variables should be entered in a regression equation? *Journal of the American Statistical Association*, 78(381), 131–136.
- Brick, J. M., & Kalton, G. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5(3), 215–238.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. The University of Chicago Press.
- Calkins, A., Guenther, W., Belfiore, G., & Lash, D. (2007). *The turnaround challenge*. Mass Insight Education and Research Institute.
- Chen, H. T. (1990). *Theory-driven evaluations*. Sage.
- Chen, H. T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, 13(1), 75–83.
- Cohen, D. K., & Ball, D. L. (1999). *Instruction, capacity, and improvement*. Consortium for Policy Research in Education, University of Pennsylvania, Graduate School of Education.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98.
- Datnow, A., & Stringfield, S. (2000). Working together for reliable school reform. *Journal of Education for Students Placed at Risk (JESPAR)*, 5(1–2), 183–204.
- de la Torre, M., Allensworth, E., Jagesic, S., Sebastian, J., Salmonowicz, M., Meyers, C., & Gerde-man, R. D. (2012). *Turning around low-performing schools in Chicago*. Consortium for Chicago School Research, University of Chicago.
- Dee, T. (2012). School Turnarounds: Evidence from the 2009 Stimulus. NBER Working Paper No. 17990, April 2012, under revision.
- Desimone, L. M. (2012). Complementary methods for policy research. In G. Sykes, B. Schneider, D. Plank, & T. Ford (Eds.), *Handbook of Education Policy research* (pp. 179–191). Routledge.
- Desimone, L. (2002). How can comprehensive school reform models be successfully implemented? *Review of Educational Research*, 72(3), 433–479.
- Desimone, L., & LeFloch, K. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, 26(1), 1–22.
- Edgerton, A. K., & Desimone, L. M. (2019). Mind the gaps: Differences in how teachers, principals, and districts experience college-and career-readiness policies. *American Journal of Education*, 125(4), 593–619.
- Firestone, W. A., & Herriott, R. E. (1982). Prescriptions for effective elementary schools don't fit secondary schools. *Educational Leadership*, 40(3), 51–53.
- Fowler, F. J. (2014). *Survey Research Methods (Fifth Edition)*. Los Angeles, CA: Sage Publications, Inc.


- Gliem, R. R., & Gliem, J. A. (2003). *Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales*. Midwest Research-to-Practice Conference in Adult, Continuing, and Community Education. Retrieved from <https://scholarworks.iupui.edu/handle/1805/344>
- Hallinger, P., & Murphy, J. F. (1986). The social context of effective schools. *American Journal of Education*, 94(3), 328–355.
- Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring program effects for specialized populations: Does special education raise achievement for students with disabilities? *Review of Economics and Statistics*, 84(November), 584–599.
- Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, 62, 302–320.
- Herman, R., Dawson, P., Dee, T., Greene, J., Maynard, R., & Redding, S. (2008). *Turning around chronically low-performing schools*. Institute of Education Sciences: National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/practice_guides/Turnaround_pg_04181.pdf
- Hill, K. L. (2016). *Assessing school turnarounds: Using an integrative framework to identify levers for success*. University of Pennsylvania.
- Hill, K., Desimone, L., Wolford, T., & Reitano, A. (2017). Assessing school turnarounds: Using an integrative framework to identify levers for success. In C. Meyers & M. Darwin (Eds.), *Contemporary perspectives on school turnaround and reform: enduring myths that inhibit school turnaround*. Information Age Publishing.
- Jacob, B. A., Courant, P. N., & Ludwig, J. (2003). Getting inside accountability: Lessons from Chicago [with Comments]. In *Brookings-wharton papers on urban affairs* (pp. 41–81). Retrieved from <http://www.jstor.org/stable/25067395>
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14–26.
- Kutash, J., Nico, E., Gorin, E., Tallant, K., & Rahmatullah, S. (2010). *School turnaround: A brief overview of the landscape and key issues*. FSG Social Impact Advisors.
- Lee, V. E., Bryk, A. S., & Smith, J. B. (1993). The organization of effective secondary schools. *Review of Research in Education*, 19, 171–267.
- Leigh, J. P. (1988). Assessing the importance of an independent variable in multiple regression: Is stepwise unwise? *Journal of Clinical Epidemiology*, 41(7), 669–677.
- Liu, P. (2017). Transforming turnaround schools in China: A review. *School Effectiveness and School Improvement*, 28(1), 74–101.
- Mark, M. M. (1990). From program theory to tests of program theory. *New Directions for Program Evaluation*, 47, 37–51.
- Mayer, D. P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, 21(1), 29–45.
- McCauley, C. (2018). A Systems approach to rapid school improvement. *State Education Standard*, 18(2), 6.
- McCoach, D. B., O'Connell, A. A., & Levitt, H. (2006). Ability grouping across kindergarten using an early childhood longitudinal study. *The Journal of Educational Research*, 99(6), 339–346.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171–178.
- Mertler, C. A. (2003). Patterns of response and nonresponse from teachers to traditional and web surveys. *Practical Assessment, Research & Evaluation*, 8(22). Retrieved September 7, 2014 from <http://PAREonline.net/getvn.asp?v=8&n=22>
- Meyers, C. V., & Smylie, M. A. (2017). Five myths of school turnaround policy and practice. *Leadership and Policy in Schools*, 16(3), 502–523.
- Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook*. Sage Publications Inc.
- Murnane, R. J. (1975). *The Impact of school resources on the learning of inner city children*. Ballinger.
- Murphy, J. F., & Bleiberg, J. F. (2019). Why school turnaround failed: Lethal problems. *School turnaround policies and practices in the US* (pp. 75–117). Cham: Springer.
- Murphy, J. (2010). Turning around failing organizations: Insights for educational leaders. *Journal of Educational Change*, 11(2), 157–176.
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Peck, C., & Reitzug, U. C. (2013). School turnaround fever: The paradoxes of a historical practice promoted as a new reform. *Urban Education*, XX(X), 1–31.

- Peck, C. M., & Reitzug, U. C. (2018). My progress comes from the kids: Portraits of four teachers in an urban turnaround school. *Urban Education*. <https://doi.org/10.1177/0042085918772623>
- Peurach, D. J., & Neumerski, C. M. (2015). Mixing metaphors: Building infrastructure for large scale school turnaround. *Journal of Educational Change*, 16(4), 379–420.
- Player, D., & Katz, V. (2016). Assessing school turnaround: Evidence from Ohio. *The Elementary School Journal*, 116(4), 675–698.
- Porter, A. C., Floden, R., Freeman, D., Schmidt, W., & Schulle, J. (1986). *Research series no. 179: Content determinants*. The Institute for Research on Teaching, Michigan State University. Retrieved from <http://education.msu.edu/irt/PDFs/ResearchSeries/rs179.pdf>
- Porter, A. C. (1994). National standards and school improvement in the 1990s: Issues and promise. *American Journal of Education*, 102, 421–449.
- Porter, A. C. (1989). External standards and good teaching: The pros and cons of telling teachers what to do. *Educational Evaluation and Policy Analysis*, 11(4), 343–356.
- Purkey, S. C., & Smith, M. S. (1983). Effective schools: A review. *The Elementary School Journal*, 83, 427–452.
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2001). *Teachers, schools, and academic achievement. Working paper 6691*. National Bureau of Economic Research.
- Rosenholtz, S. (1985). Effective schools: Interpreting the evidence. *American Journal of Education*, 93(3), 352–388.
- Sebring, P. B., Allensworth, E., Bryk, A. S., Easton, J. Q., & Luppescu, S. (2006). *The essential supports for school improvement*. Consortium on Chicago School Research at the University of Chicago Chicago, IL. Retrieved from http://www.uquebec.ca/observgo/fichiers/60612_gse-2.pdf
- Sheehan, K. B. (2001). E-mail survey response rates: A review. *Journal of Computer-Mediated Communication*. <https://doi.org/10.1111/j.1083-6101.2001.tb00117.x>
- Smith, N. L. (1990). Using path analysis to develop and evaluate program theory and impact. *New Directions for Program Evaluation*, 47, 53–57.
- Song, M., & Herman, R. (2010). *A practical guide on designing and conducting impact studies in education: Lessons learned from the what works clearinghouse (Phase I)*. American Institutes for Research.
- Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, 15(3), 250.
- Stratos, K., Reitano, A., Wolford, T., & Miller, M. K. (2014). Renaissance schools initiative: Report on school visits, 2013–2014. The School District of Philadelphia: The Office of Research and Evaluation.
- Streiner, D. L. (2013). *A guide for the statistically perplexed: Selected readings for clinical researchers*. University of Toronto Press.
- Streiner, D. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80, 99–103.
- Strunk, K. O., Marsh, J. A., Hashim, A. K., & Bush-Mecenas, S. (2016a). Innovation and a return to the status quo: A mixed-method study of school reconstitution. *Educational Evaluation and Policy Analysis*, 38(3), 549–577.
- Strunk, K. O., Marsh, J. A., Hashim, A. K., Bush-Mecenas, S., & Weinstein, T. (2016b). The impact of turnaround reform on student outcomes: Evidence and insights from the Los Angeles Unified School District. *Education Finance and Policy*, 11(3), 251–282.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 25(1), 1.
- Stuart, E. A., & Rubin, D. B. (2007). Best practices in quasi-experimental designs: Matching methods for causal inference. *Best practices in quantitative methods* (pp. 155–176). Sage Publications Inc.
- Tannenbaum, C., Boyle, A., Graczewski, C., James-Burdumy, S., Dragoset, L., & Hallgren, K. (2015). *State capacity to support school turnaround (NCEE 2015–412)*. Institute of Education Sciences.
- The School District of Philadelphia. (2015). Action Plan 3.0. Retrieved from www.philasd.org/actionplan
- The Center on School Turnaround. (2017). *Four domains for rapid school improvement: A systems framework [The Center for School Turnaround at WestEd]*. WestEd.
- Thompson, B. (1995). *Stepwise regression and stepwise discriminant analysis need not apply*. Retrieved from <http://eric.ed.gov/?id=ED382635>
- Trujillo, T., & Renée, M. (2015). Irrational exuberance for market-based reform: How federal turnaround policies thwart democratic schooling. *Teachers College Record*, 117, 1–34.

- U.S. Department of Education. (2003). *No child left behind: A parents guide*. Retrieved from: <http://www2.ed.gov/parents/academic/involve/nclbguide/parentsguide.pdf>
- U.S. Department of Education. (2009). *American recovery and reinvestment act: Strategic planning*. Retrieved from <http://www2.ed.gov/policy/gen/leg/recovery/presentation/arra-strategic-planning.ppt>
- U.S. Department of Education. (2009). *Applications now available for \$3.5 billion in title i school improvement grants to turn around nation's lowest achieving public schools*. Retrieved from <http://www.ed.gov/news/press-releases/applications-now-available-35-billion-title-i-school-improvement-grants-turn-aro>.
- VanGronigen, B. A., & Meyers, C. V. (2019). How state education agencies are administering school turnaround efforts: 15 years after no child left behind. *Educational Policy*, 33(3), 423–452.
- Weiss, R. S. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York, NY: The Free Press.
- What Works Clearinghouse. (2008). Turning around chronically low-performing schools. institute of education sciences: National center for education evaluation and regional assistance. Retrieved from http://www.ies.ed.gov/opac.acc.msme.edu/ncee/wwc/pdf/practice_guides/Turnaround_pg_04181.pdf
- Zavadsky, H. (2012). *School turnarounds: The essential role of districts*. Harvard Education Press.
- Zimmer, R., Henry, G. T., & Kho, A. (2017). The effects of school turnaround in Tennessee's achievement school district and innovation zones. *Educational Evaluation and Policy Analysis*, 39(4), 670–696.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Kirsten Lee Hill¹  · Laura Desimone² · Tonya Wolford³ · Adrienne Reitano³ · Andrew Porter⁴

Laura Desimone
lauramd@udel.edu

Tonya Wolford
twolford@philasd.org

Adrienne Reitano
areitano@philasd.org

Andrew Porter
andyp@upenn.edu

¹ Kirsten Lee Hill Consulting, 310 Broad Street, Apt 608, Charleston, SC 29401, USA

² College of Education and Human Development, University of Delaware, 106 Alison Hall West, Newark, DE 19716, USA

³ The School District of Philadelphia Office of Research and Evaluation, 440 N Broad St, Philadelphia, PA 19130, USA

⁴ University of Pennsylvania, 3700 Walnut St, Philadelphia, PA 19104, USA