Building Student Ownership and Responsibility: Examining Student Outcomes from a Research-Practice Partnership

Marisa Cannata, Vanderbilt University
Christopher Redding, University of Florida
Tuan D. Nguyen, Vanderbilt University

A version of this paper was presented at the annual meeting of the Association of Education Finance and Policy in Denver, CO, March 17-19, 2016.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305E100030. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Building Student Ownership and Responsibility: Examining Student Outcomes from a Research-Practice Partnership

Abstract

This paper is situated at the intersection of two trends in education research: a growing emphasis on the importance of co-cognitive traits and emergence of research-practice partnerships to more effectively scale effective practices. Our partnership focused on building student ownership and responsibility for their learning, which means creating school-wide practices that foster a culture of learning and engagement among students. We report evidence showing that participating in this improvement process was related to small but desirable improvements in grades and course failure rates, while the impact on absences and disciplinary infractions were not statistically significant. We also use qualitative data about the quality of implementation to understand how school-level adaptations may be related to observed outcomes.

Keywords: Student ownership and responsibility; non-cognitive traits; course failure; absences; research-practice partnership

Building Student Ownership and Responsibility: Examining Student Outcomes from a Research-Practice Partnership

Despite decades of ambitious high school reform, substantial evidence demonstrates reforms are inconsistently implemented and struggle to impact student learning (Datnow, Hubband, & Mehan, 2002; Dragoset et al., 2017; Gross, Booker, & Goldhaber, 2009; Mazzeo, Fleischman, Heppen, & Jahangir, 2016), although progress has been made in graduation rates (DePaoli et al., 2015). In response, the last several years has seen a proliferation of new approaches to achieving school improvement at scale, such as improvement science and design-based implementation research (Cohen-Vogel, Cannata, Rutledge, & Socol, 2016; Bryk, Gomez, Grunow, & LeMahieu, 2015; Fishman, Penuel, Allen, & Cheng, 2013). While these methods may differ in specifics, they share an assumption that improvement at scale comes not from replicating a proven program, but by practitioners and researchers working together with iterative, continuous improvement approaches to design and implement on issues of school improvement (Bryk et al., 2015; Cohen-Vogel et al., 2015; Fishman et al., 2013).

These new approaches to scale reflect a shift in scholarship on scaling up educational innovations that emphasizes adaptive integration and a focus on integrating new effective practices into existing systems rather than strict fidelity to the original design (Cannata & Rutledge, 2017; Russell et al., 2015). While several innovations developed through research-practice partnerships have been tested in rigorous efficacy studies with desirable outcomes (Booth et al., 2015; Snow, Lawrence, & White, 2009), there is less evidence on the impact of innovations implemented using these new continuous improvement approaches to scale. One example is the Community College Pathways initiative led by the Carnegie Foundation for the

Advancement of Teaching, which improved course completion rates (Sowers & Yamada, 2015). Much of the research on the continuous improvement approaches to scaling up have focused on how educators engage in the process of continuous improvement and the work of launching improvement networks (Cannata, Cohen-Vogel, & Sorum, 2017; Hannan et al., 2015; Russell et al., 2017).

While the lack of empirical support for these models can be attributed, in part, to their infancy, it is also due to the diversity of approaches and challenges in evaluating the outcomes of a reform model that promotes iteration as a key feature. Given the increasing federal and private investment in such approaches, we argue for the need for evidence of how such models may improve students' outcomes. In this paper, we report evidence of student outcomes from a multi-year partnership within one large, urban district. In this partnership, three schools co-developed practices to improve student ownership and responsibility as part of a continuous improvement process. We evaluate evidence of student ownership by assessing changes in grades, course failures, discipline, and attendance. We adopt a mixed methods framework to describe both evidence of student outcomes and the implementation approaches that may have shaped these outcomes. We seek to answer two research questions:

(1) To what extent did the innovation reduce students' disciplinary infractions and the number of failed courses and improve student grades and attendance?

(2) How does the adaptive integration of each implementation component explain observed outcomes?

We begin with a review of the literature, describing the emergence and aims of continuous improvement approaches in education. We then outline research on co-cognitive traits and their relationship with various student outcomes. After describing the approach used to design, adapt,

and implement the innovation in the partner district, we describe the data used for this study as well as the quantitative and qualitative methods used to answer our research questions. We then present our results, describing how school teams adapted the innovation components to their school context and staff perceptions of student outcomes. We then provide quantitative evidence on four student outcomes: attendance, grades, course passing, and discipline.

Implementation through Continuous Improvement

New approaches to scaling up are designed to overcome shortcomings in the translation of research into practice that occur as district leaders lack the time or skill to utilize research or believe it does not address their needs (Coburn, Honig, & Stein, 2009). Calls for research-practice partnerships seek to more adequately address the contextual factors that shape implementation and scale up (Coburn, Penuel, & Geil, 2013). Under the umbrella of research-practice partnerships, several models have been put forward, each with differing roles for researchers and practitioners (Coburn et al., 2013). For instance, the Strategic Education Research Partnership's model gathers educational professionals across a school district hierarchy to come together, identify an ongoing problem, and iteratively develop programs based on researchers' suggested solutions and program developers' facilitation (Donovan, Snow, & Daro, 2013). Improvement science, such as Networked Improvement Communities, seek to engage educators in collecting data for short-cycle improvement efforts and build up to larger scale change through continuous improvement (LeMahieu, Grunow, Baker, Nordstrum, & Gomez, 2017). Design-based implementation research aims to develop usable practices through the collaboration with educators in a process that shifts their role from implementers to co-designers (Penuel, Fishman, Cheng, & Sabelli, 2011). Common across these approaches is having educators and researchers work together to design, study, and iterate effective practices as

educators adapt them to their specific contexts (Cohen-Vogel et al., 2015).

Using continuous improvement approaches to implementing educational innovations is important because, while there are many innovations that have positive outcomes in rigorous efficacy trials, it is less clear whether these innovations are usable for schools (Coburn & Penuel, 2016). Educational implementation research has long noted that schools adapt innovations to focus on their unique needs, sometimes to ill effect (Datnow & Park, 2009; Dede, 2006; Siskin, 2016). Continuous improvement approaches to scale can bring discipline to the adaptation process as school teams bring evidence of what they have accomplished to share and discuss with others focused on the same problem of practice (Cannata, Cohen-Vogel, & Sorum, 2017; LeMahieu, et al., 2017).

The continuous improvement process includes some form of the Plan, Do, Study, Act (PDSA) cycle, which is a model for organizational improvement that requires identifying the aim of a particular improvement, testing the change idea, and monitoring whether the observed changes led to the intended improvement (Langley et al., 2009). Results from an individual test can lead to either revising and testing the change again, or deciding to scale it into more diverse contexts. An important component to this process is that educators are directly involved in designing the innovations that will be tested in their context. This attention to local context is particularly important for achieving scale as innovations must be able to fit with contexts that vary greatly while coping with change, promoting ownership, building capacity, and enabling effective decision-making (Cohen et al., 2013). At the same time, challenges exist when major decisions about the focus and content of the reform are left to local decision-makers (Cannata & Nguyen, 2015; Cohen et al., 2013; Nunnery, 1998).

While continuous improvement approaches bring potential benefits, assessing the impact

of the innovations developed through them encounters methodological challenges. First, as the continuous improvement process is used to iteratively develop the innovation, there is no way to distinguish the effect of this process from the specific practices that are developed. In other words, the treatment is comprised of both the design process and innovation design. Second, while the innovation design was rooted in a common focus, schools had leeway to customize the innovation design to their unique context, resulting in differences in how the innovation was implemented across these schools. In this study, we adopt a difference-in-differences estimation strategy to compare student outcomes among the innovation schools to the remaining schools in the district. This approach allows us to test and control for pre-existing differences between schools to ensure the estimate of the innovation's impact takes into account any pre-treatment differences (Angrist & Pischke, 2009). To address the concern with school-level adaptations, we examine both the overall effect of participating in this program as well as differences among the innovation schools to examine how each school's adaptations are associated with student outcomes. In the next section, we describe the research base for the innovation designed through this continuous improvement process.

Co-Cognitive Traits and Student Achievement

The innovation developed through this partnership included practices designed to build greater student ownership and responsibility (SOAR), which was identified as what distinguished higher and lower performing high schools in this district (Cannata, Smith, & Taylor Haynes, 2017). The definition of SOAR is grounded in both the specific findings from this district and the broader literature on co-cognitive student behaviors. This focus on changing students' beliefs and mindsets and providing them problem-solving skills to engage in academic work builds on a robust empirical research base on co-cognitive factors (Farrington et al., 2012). With the

increasing interest in co-cognitive factors comes scholarly debate about how to define and label them (Duckworth & Yeager, 2015). We use the term co-cognitive, which stems from research on giftedness, to reflect that these characteristics are not separate from cognition, but "interact with and enhance the cognitive traits ordinarily associated with the development of human abilities" (Renzulli, 2011, pp. 307).

Numerous studies have demonstrated that people's implicit theories about their own personal qualities influence how they respond to challenges and setbacks (Dweck, 2007). Research on students' implicit theories of learning indicates that student mindsets are malleable and related to important student outcomes (Yeager & Walton, 2011; Yeager et al., 2016). For example, Blackwell, Trzesniewski, and Dweck (2007) designed an intervention to change middle school students' mindsets and found that students who learned theories about growth and fixed mindsets saw improvements in their math grades. Growth mindset interventions in middle and high schools have also led to fewer students failing algebra, improved test scores, and college enrollment (Yeager & Walton, 2011; Yeager et al., 2016). More generally, teachers' instructional practices can predict students' co-cognitive traits (Blazar & Kraft, 2017).

The increased focus on co-cognitive traits complements the growing recognition of the importance of outcomes other than student achievement. Recent research has indicated that high school course grades are better predictors of college access, college graduation, and longer-term life outcomes than test scores. GPA, for example, is a consistent predictor of graduating from both high school and college, and a "primary driver of differences by race/ethnicity and gender in educational attainment" (Farrington et al., 2012, p. 3). This may be because grades reflect a multidimensional construct that contain both cognitive and co-cognitive factors (Bowers, 2009; Brookhart et al., 2016). Further, failing a course predicts dropping out of high school (Bowers,

Sprott, & Taff, 2013). Indeed, when schools act on information in early warning indicators to target improvements in course failure rates, students are more likely to graduate and enroll in college (Mac Iver & Messel, 2013).

Absenteeism is another indicator often included in early warning systems, and reducing absenteeism can also improve high school graduation and college enrollment (Faria et al., 2017; Mac Iver & Messel, 2013). Further, students who miss more school than their peers consistently score lower on standardized tests (Ginsburg, Jordan, & Chang, 2014; Gottfried, 2011). Chronic absenteeism, defined as missing at least 10 percent of a school year, exacerbates the achievement gap between lower and higher income students (Balfanz & Byrnes, 2012).

Student disciplinary infractions are an additional measure of students' co-cognitive traits. In the district in this study, student disciplinary infractions cover a range of behaviors, such as bullying, fighting, disrespect to teacher, but is often met with a similar outcome: in-school or out-of-school suspension. This exclusionary tactic results in the immediate loss of academic time in the classroom (Gregory, Skiba, & Noguera, 2010). Such disciplinary action is associated with student grades and achievement test scores (Arcia, 2006; Rausch & Skiba, 2004) and school dropout (Raffaele Mendez, 2003). Further, evidence has consistently shown that underserved racial/ethnic minorities have received such disciplinary action (Rocque, 2010; Skiba et al., 2014), possibly contributing to demographic achievement gaps (Gregory, Skiba, & Noguera, 2010).

Translating Research into Practice: Developing the SOAR Innovation

This partnership began with intensive study of higher and lower performing high schools in the district, and identified SOAR as the key differentiating feature (Cannata, Smith, & Taylor Haynes, 2017). In 2012-13, a district design team examined this research, conducted additional needs analysis related to SOAR and designed an initial SOAR prototype aimed at creating a set

of norms and school-wide practices that foster a culture of learning and engagement among students. Specifically, SOAR practices focused on building a student growth mindset and developing problem-solving skills to improve student engagement. This districtwide team included educators from three high schools identified as the first sites of implementation, known as innovation schools, as well as other district high schools, central office administrators, researchers, and program developers.

With the districtwide team setting the initial focus, each innovation school established a SOAR team in 2013-2014 to pilot practices and make further develop the practices based on their learning. SOAR teams had 6-8 members, who were almost all teachers (one school had an assistant principal on the team). The SOAR team was responsible for leading implementation in their school, often by working with the administration, and both developed SOAR practices using PDSA and provided training and other support for other teachers in the school to enact SOAR practices. By the end of the development year, the core practices of the innovation included 1) teaching about growth mindset, 2) student grade monitoring and goal-setting activities, 3) problem-solving activities that supported students in improving their grades, and 4) a behavioral reflection form designed to get students to reframe problematic behaviors before creating a disciplinary referral. The final SOAR component focused on building a school culture around SOAR. Full implementation began in 2014-15 and continued in 2015-16.

Through this development process, school teams were also given leeway in customizing the common district design to their particular context. While each school design team implemented these common elements of the design, its delivery varied in ways that may shape student outcomes. For instance, at one school, grade monitoring was one of many activities implemented in weekly advisory periods. At the other two schools, advisories were only held

every three weeks. The grade monitoring process itself varied across schools, but generally included having students set goals for grades they wanted in each class, track their progress toward those goals with each progress report, and engage in teacher-led activities designed to help them understand how their academic behaviors (e.g., not turning in assignments) contribute to their grades. Growth mindset lessons were based around ideas brain science research and built on similar interventions (Blackwell et al., 2007). Despite similar content, schools varied in how this content was delivered. Hancock implemented the growth mindset intervention in a single advisory period, Williams implemented it in a series of lessons on the second day of school, and Smith focused on building teacher knowledge of brain science and implicit theories of learning.

Engaging researchers and practitioners in this continuous improvement approach was not without its challenges. This paper focuses on the implementation and outcomes of the innovation that emerged from this development process. Elsewhere, we analyze the dynamics and dilemmas of this research-practitioner partnership (Cannata & Nguyen, 2015; Cannata, Redding, Brown, Joshi, & Rutledge, 2017).

Research Design

Study Sample

We use both qualitative and quantitative data to understand outcomes of the SOAR innovation in one district over two years of implementation. This southwestern district served approximately 80,000 students; the majority were low-income or from traditionally underserved racial/ethnic groups. The innovation schools were selected through a collaborative process with district personnel and school administrators. The selected schools expressed an interest and willingness to participate in this innovative reform model. While a school's value-added performance was not used in the selection of these schools, their school value-added suggests

that they were moderately performing schools in the district.

The qualitative data for this study come from field visits in these three innovation high schools. Two four-day field visits occurred in the first year of implementation (October and April) and one three-day visit in March of the second year of implementation. In each visit, we interviewed the principal, other administrators, members of the SOAR team, and other teachers in the schools, as well as conduct focus groups with students and teachers or support staff. The student focus groups were linked to observations of specific teachers, which were chosen because they taught classes designed for mostly higher or lower performing students in grades 9/10 or 11/12. Due to the small sample of classroom observations, we did not use the observations to measure the degree of implementation, but to ground student focus groups in discussions of specific classroom practices that capture aspects of growth mindsets and problem-solving skills. Table 1 lists the data sources from these field visits. We use the fieldwork data to provide evidence on enactment of the practices and how participants described the outcomes they were achieving as a result of this work. The interviews and focus groups focused their understanding of student ownership and responsibility and specific innovation practices, support for the innovation, and perspective on implementation, including the extent to which they enact SOAR practices. Interview and focus group guides are in the Appendix.

In addition to the fieldwork data, we also take advantage the rich administrative data from the district. We draw on administrative data for all high school students enrolled in the 2010-2011 to 2015-2016 school years, to allow for comparisons between innovation and non-innovation high schools from year to year. The data used for this study includes 91,410 student-year observations. Of these students, approximately 3% are dropped from the analysis due to missing data. The analytic sample includes 33,215 unique student observations.

Measures

Outcome measures were identified collaboratively with district stakeholders. These outcomes include students' grades, passing rates, absences, and disciplinary infractions. Student's grades are the averages of the students' scores for each class. In 2013-2014, this measure ranged from 0 and 100, with an average student grade of 82 (see Table 2). When operationalizing a students' passing rate, we focus on the number of courses a student did not pass throughout the school year. Students were considered to be failing a course if they did not score at least a 70% in a course. As students could be registered for up to nine courses a semester, the maximum value for this variable is 18. Although the modal value for this variable is 0, on average, students did not pass 1 course. The measure for days absent is the number of days a student did not attend in a particular year.[1] Student infractions is a measure of the number of infractions a student received in a particular school year. Infractions include code of conduct violations for behaviors such as cheating, disrespect towards teachers, bullying, fighting, disobeying school rules, dress code violations, or possession of tobacco. Infractions also include more serious offenses such as drug or alcohol use, criminal mischief, assault, arson, felony, possession of a weapon, public lewdness, gang violence, or serious misbehavior.

We also include controls for binary indicators of student race/ethnicity (Black, Hispanic, or other race/ethnicity), free and reduced lunch (FRPL) status, gifted status, and grade level. Additionally, we control for the number of days in which a student was enrolled in a school, indicators of whether or not they withdrew or started after the beginning of the school year, and for the number of courses in which a student is registered throughout the school year. At the school level, we control for student enrollment as well the proportion of Black students,

[1] The data on student absences is only available beginning in the 2011-2012 school year.

percentage of Hispanic students, and students who receive FRPL.

Methods

For this study, we adopt a sequential mixed methods research design (Smith, Cannata, & Taylor Haynes, 2016; Creswell, 2009; Teddlie & Tashakkori, 2006). We first draw on qualitative fieldwork data to determine the quality of implementation and participant understandings of accomplishments. The quantitative analysis is then used to ascertain the extent to which students in the innovation schools benefited from SOAR. Using a sequential approach allowed us to formally investigate hypotheses that emerged from the fieldwork. For example, we heard participants suggest students responded differently to SOAR practices depending on their prior background. This led us to examine differential effects by the lagged outcome variables. We first describe how we analyzed the fieldwork data, before describing the quantitative analysis.

Following each research visit, data were coded using an a priori framework for implementation that focused on facilitating conditions (will, capacity, beliefs, and alignment to context), implementation supports (implementation team dynamics, engagement in continuous improvement, leadership, resources/training), implementation quality which itself involved teacher experiences with implementation (enactment of innovation practices, feedback on practices) and student experiences with implementation (responsiveness, perceived outcomes). The coding team first coded several transcripts independently, and then compared coded transcripts to ensure they were applying codes consistently. Through multiple rounds, the coding framework was revised or clarified. For example, capacity was expanded to differentiate between capacity of teachers to enact SOAR practices, capacity of the implementation team to lead the work, and organizational capacity of the school.

Once the coding team agreed on the final coding scheme, they independently coded all

transcripts. After coding was complete, a researcher prepared detailed memos for each school for each major theme in the coding framework. This process was repeated after each field visit. In Year 2, the coding scheme was further expanded to include antecedents to sustaining and scaling the practices. Memos around these themes at each time point served as the primary documents for investigating the quality of implementation. There were five SOAR practices for which we analyzed the quality of how the schools enacted the innovation practices: teaching growth mindset, goal-setting and grade monitoring practices, problem-solving practices, rewarding positive behavior, and building a school culture around these practices. For each of these practices, four researchers independently read memos on implementation quality and categorized each school as high, medium, or low enactment.

This analytic process used several strategies to address potential threats to the validity of our inferences from the qualitative data, including cross-validation between researchers, triangulation among sources and perspectives, exploration of rival hypotheses, and member checking (Miles & Huberman, 1994; Patton, 2002). For example, we sought out comparisons between perspectives of the SOAR team and perspectives of others in the school, recognizing that overreliance on the SOAR team may reflect elite bias (Miles & Huberman, 1994). Triangulating between the qualitative and quantitative findings also encouraged us to consider rival hypotheses. In particular, initial analyses suggested relatively strong implementation in Hancock and Williams, with Smith struggling and eventually discontinuing participation. Yet, this appeared disconnected from staff perceptions of outcomes, along with the quantitative outcomes indicating that Hancock was the school with least positive outcomes. This led us to focus on the specific routines of practices each school enacted.

For the quantitative analysis, we use a difference-in-differences (DD) analysis to compare

students' prior outcomes to average student outcomes in non-innovation schools. This ordinary least squares (OLS) model can be estimated:

$$y_{ist} = \beta_0 + \beta_1 Innovation_{st} + B_2 Y_{ist-1}\, \beta_3 X_{ist} + \beta_4 S_{st} + \delta_s + \gamma_t + \epsilon_{ist} \quad (1)$$

where Y_{ist} is the outcome for student *i* in school *s* in year *t* and $Innovation_{st}$ is a dummy variable for whether the school implemented the SOAR innovation[2], δ_s is a school fixed effect and γ_t is a year fixed effect. X_{ist} are student-level controls, S_{st} includes time-varying school characteristics, and ϵ_{ist} is an error term. In this model, β_1 can be interpreted as the difference in student outcomes between innovation and non-innovation schools after implementation. To account for repeated observations of student over time, standard errors are clustered at the school level (Bertrand, Duflo, & Mullainathan, 2004).[3]

This initial analysis estimates an average treatment effect for the students in schools that participated in this continuous improvement process. In addition to this overall treatment effect, we examine several heterogeneous treatment effects. These included differences across the three innovation schools, prior student performance, post-treatment year, and among traditionally underserved student subgroups.

In additional analysis, we examine the robustness of the DD research design (Angrist & Pischke, 2009). An assumption of this research design is that innovation and non-innovation

[2] The post-treatment period does not include the year when the SOAR team developed and piloted the practices of the SOAR innovation. This decision is justified for two reasons. First, the piloting that did occur was limited to members of the SOAR team. Second, when a practice was piloted, it tended to only be implemented once or twice, limiting its potential impact on student outcomes. Nevertheless, it is possible that this piloting would lead to pre-treatment differences.

[3] Three of the variables—the number of classes failed, days absent, and disciplinary infractions – are count outcomes and, thus, not normally distributed. In Table A6, we replicate our main analysis using negative binomial regression to deal with this over-dispersed data. The results are consistent across these two modeling strategies in terms of the direction and significance. When the results do change, they are more precise in the negative binomial regression results. We report the OLS results for ease of interpretation.

schools had similar pre-treatment trends in the outcome variables. We test for this assumption in two ways. We first estimate the relationship between innovation school participation and student outcomes not only in the post-treatment period but in all school years in which we have data. These estimates function as a placebo test where any significant differences in the slope between innovation and non-innovation schools prior to treatment would indicates a violation of the parallel trends assumption. Graphically, we also show the predictions from this regression to visually examine the presence of pre-treatment trends, when holding all other variables at their mean. Evidence of a violation of this assumption would indicate pre-treatment differences between innovation and non-innovation schools that could explain why innovation schools have more positive outcomes in the post-treatment period, outside of their participation in the continuous improvement process. In general, we find evidence of parallel trends in terms of student grades and number of failed courses but not attendance and disciplinary infractions. In addition, we find evidence that Williams High School and Smith High School pre-treatment trends generally resembled the non-innovation schools. The evidence of pre-treatment differences at Hancock High School limit our inference of the effect of SOAR at this school. We discuss the full results of this sensitivity analysis below.

Findings

Quality of Implementation

Table 3 summarizes how each school enacted each of the SOAR components. While each school enacted the core components to some extent, Williams and Smith emphasized specific routines, and Hancock emphasized school culture and student relationships. All three schools enacted goal-setting and grade-monitoring, although they varied in the extent to which this routine was emphasized. Over three-quarters of teachers in Williams described grade monitoring

as the core practice of SOAR. Further, all participants in Willliams, including teachers who did not have a second period class and thus did not directly participate, were familiar with the process and its purpose, and all but one teacher reported implementing it every three weeks. The following example typifies how many teachers described the practice:

> With the three weeks progress reports and six weeks progress reports, the kids are charting their own grades, they're seeing where they stand, what they need to accomplish in the class as far as what average they would need for the remaining six weeks in order to pass. Asking them reflective questions on what they can do to improve themselves, how they got to be where they are, and they have to write little statements about that.

While nearly all teachers in Williams implemented the practice, there was variation in the amount of engagement that teachers described having with students during the process. A Williams administrator described this variation by saying that some teachers "just hand out the sheet and say do this... But those teachers who really do engage with their students in conversations about goal setting… then I think they've really gotten a lot out of it." Student comments in Williams are consistent with administrators and teachers in that they indicate nearly all teachers enact the routine, but teachers vary in the depth in which they engage students in it. One student describes her experience,

> It all depends on your teacher. Like I have [teacher] second period, she'll go through with you and like she'll go around the whole class each person, ask why did your grade go down, what can you do to get it up…but then some other teachers, like people say that they just have you fill out the graph and turn them in.

Smith and Hancock also enacted the grade-monitoring routine, with stakeholders reporting a similar variation in teacher engagement as Williams. However, while grade-monitoring was reported by three-quarters of Williams teachers as the major emphasis of SOAR, only about a third of teachers in Smith said the same. The remaining Smith teachers described problem-solving or discipline as SOAR's major focus. While SOAR did include routines in these three areas, only a few Smith stakeholders described them as working together, as one teacher

said, "So they're getting their grades from whatever – whoever their homeroom teacher is, but they're also looking at things like making decisions, planning, how to improve, how to set goals, and how to reach them, how to problem solve." Most stakeholders in Smith indicated that it is unclear how these routines are connected. Even a member of the Smith SOAR team member said their work struggles because it "meant too many different things to too many different people."

In Hancock, grade-monitoring was just one component among many activities that they do in their weekly advisory period, which all stakeholders described as the main SOAR emphasis. Other advisory topics focused on college and career readiness, such as understanding transcripts, financial aid, and college admissions. Teachers in Hancock frequently spoke about the "SOAR curriculum" and how they were "doing the lessons," with little evidence that SOAR was influencing practice outside the advisory period. Over half of Hancock teachers indicated they have not been asked to do anything outside of advisory. For example, one teacher said,

> We have the advisory class, where we implement the SOAR lessons. So we teach the kids the lessons that are supposed to – I guess they're supposed to transfer to classes. But as far as our regular classes on a daily basis, there's nothing that they've asked us to use, you know, no methods in classroom that they ask us to use to promote the SOAR.

When asked if teachers talked about the ideas from advisory in other classes, students in nearly all focus groups said this only occurred if they were also involved in a separate program in the school that also has a college preparation focus. While grade-monitoring in Hancock was just one activity in SOAR, more central to advisory was developing relationships between students and teachers. Most Hancock teachers saw advisory as a way to mentor students. As one Hancock teacher explained it, "This is a great opportunity for a mentorship…Not just the student ownership, but coming in and having a relationship with the child so that they can be some sort of mentor for them." Most students in Hancock said they appreciated the extra support provided by teachers, with one student saying,

> A lot of our teachers, if they see that we're failing, they pull us aside and they like talk to us and like what we can do to make up that grade or something, and then they also like -- they want you to talk to your parents to make sure your parents also know that they need to get onto you about doing your work, and I think that helps a lot of students.

A second SOAR component, growth mindset, was also enacted more strongly in Williams than in Hancock and Smith. In the first week of school at Williams, the SOAR team organized the school to deliver a set of seven lessons about growth mindset throughout the day, which all teachers in Williams reported implementing. From the student perspective, they participated in an all-day learning experience around growth mindset. In focus groups, nearly all students in Williams said that they heard about growth mindset "every class the entire day," and students in one focus group called it a "conspiracy." Beyond the first week in school, about the half the teachers in Williams reported also reported using classroom practices that further fostered a growth mindset, such as allowing students to redo assignments or creating an atmosphere where students feel comfortable making mistakes. For example, one Williams teacher described using mindsets as a way to foster discussion about literature,

> I try to use the growth mindset as a way for the students to identify certain behaviors and certain motivations…I was working with my students on the Crucible, talking about which characters were willing to admit they had made a mistake.

Another teacher stated "just the policies that I have in my classroom, I've already kind of tried to implement everything to get that growth mindset." Students in Williams confirmed that at least some teachers incorporated ideas about growth mindset, with students in all focus groups providing examples of teachers who continuously reinforce ideas of growth mindset. Despite the inconsistent follow up, most teachers, administrators, and SOAR team members in Williams reported that growth mindset ideas were beginning to be part of the school culture. For example, one teacher said, "even the kids are starting to – it's starting to creep into the vocabulary."

In contrast, few stakeholders in Smith and Hancock described growth mindset as a core

component of SOAR. In fact, the growth mindset professional development in Smith was not led by the SOAR team, which perhaps signaled that it was a separate initiative. Further, since the growth mindset component in Smith was solely teacher professional development, students in only one focus group were familiar with the ideas behind growth mindset and brain science. In Hancock, the advisory did include a lesson on growth mindset, although few teachers reported engaging with growth mindset outside of this lesson.

As noted above, about two-thirds of teachers in Smith also described either the problem-solving or behavioral reflection routines as the major emphasis of SOAR. The problem-solving process in Smith, called IPAC (Identify, Plan, Act, Check), was developed after the Smith SOAR team recognized that students were having difficulty acting on the goals they had set. One Smith teacher, for example, described the evolution from goal-setting to problem-solving by saying "it's problem-solving and their ability to solve things on their own. … The program was actually implemented where the students put input last fall and then we came up with steps on how to solve problems." Most Smith teachers reported introducing IPAC to students. From the student perspective, about half the students in Smith focus groups indicated the problem-solving steps were a major push in the school. For example, one student said, "Well practically everyone wrote down their steps as a whole class…and they kind of put, posted them around the school and you see it almost every day." Yet other students did not think the problem-solving routine was a major focus, saying "I felt like they really didn't explain it well" and "it seemed like they just did it because they were told do it." Still, students in a quarter of the focus groups described benefiting from IPAC. For example, one student said it helps because "because they see it and they're like, oh this is a problem that I'm having and it helps them change what they're doing cause now that they see the problem they know how to fix it."

In contrast, when asked about problem-solving practices, nearly all Williams teachers responded by saying something similar to "I can't speak from the SOAR experience of problem-solving, but I can definitely speak from just my classroom instruction." In other words, while nearly all teachers, administrators, and support staff in Williams said they engage students in problem-solving, no stakeholder beyond the SOAR team linked these efforts to SOAR. In Hancock, the problem-solving component was considered one lesson among many in advisory, similar to growth mindset. The problem-solving lesson in Hancock included a video of people getting stuck on an escalator and waiting for someone to fix the escalator, instead of just walking down the short escalator to get to their destination. Despite being a single lesson, this example was the most frequently cited example of SOAR shaping conversations outside of advisory. Multiple Hancock teachers reported referring to the escalator example in their classes, saying, "that escalator analogy, I think [is] important and I think a lot of teachers probably use it."

There was one routine in Smith that most teachers did report implementing frequently: the GROW sheet, which was designed to have students reflect on discipline problems before writing a referral. A Smith administrator described how students frequently fill out GROW sheets, saying "on any given day, almost any period, [I] see at least one student outside of the room working on their GROW sheet. So that means teachers are using them." In Smith focus groups, students confirmed that the GROW sheet is an important routine in the school, saying "that's like a really big thing." One Smith teacher explained how the GROW sheet allows the teacher to have individualized conversations with students and allows students to take responsibility for their actions:

> I definitely think the GROW sheet's working.…I was able to talk to each one of them and say, … you're part of the problem and they're part of the problem but you have to take responsibility for what you're doing and if you have a problem, you need to come to me first and not be screaming things across the room and making comments and – you know,

so – but it gave them a chance to like, get away from that person.

Students in Smith appreciated the GROW sheet as "a chance for students to tell their side of the story" and that it "helps the student and the teacher out…the students won't be able to go to [in-school suspension]…the teacher will talk to the student about what's the problem." Most Smith teachers did report less frequent use of the GROW sheets later in the year, with a sense that they became less necessary over time as students adjusted to behavioral expectations.

In contrast, in both Williams and Hancock, the behavioral reflection routine (known as Think It Out in these schools) was only enacted by a small number of teachers. In Hancock, nearly all teachers were familiar with the routine, but teachers and the SOAR team agreed "only the ninth grade team has implemented with fidelity." In Williams, even the SOAR team admits the behavioral reflection has been minimally engaged by the school, with one member saying, "Some teachers use it…We've not really gone back to talking about the Think It Out forms more subsequently… it has languished."

Despite relatively widespread implementation of the GROW sheet, goal-setting, and IPAC, implementation in Smith was hampered by lack of administrative support, which contributed to reduced coherence of the SOAR activities. Ultimately, this lack of administrative support led to Smith officially ending its participation in the SOAR project in the second half of Year 2. At that time, the principal described his/her resistance to SOAR by saying,

> I was recruited to come here as a turn-around principal … I had a strong focus, I'm an experienced principal. So I already had a vision of what I wanted to do. And so when I came, I already had the vision and this was put on me… So it really goes back to the beginning and it was just this year I just finally said enough is enough.

Several SOAR practices were implemented in the first half of the year, and there was evidence that at least some teachers continued practices such as the GROW sheet and goal-setting after SOAR officially ended, although it is unclear how widespread this was.

Perceptions of Student Outcomes

The qualitative data also provides evidence on how school stakeholders perceived SOAR's impact on student outcomes, which we can triangulate with the school-level student outcome data. In general, teachers and administrators in all three schools felt the innovation had a palpable influence on students' academic engagement and classroom behavior. In particular, staff at all schools indicated the grade monitoring routine of SOAR has helped students be more aware of their grades. A Williams teacher indicated: "The thing that I really like to see is to see the kids talking to each other about [their grade sheet]. I hear more academic conversations than inappropriate ones … so that's nice. That's something I wouldn't have heard last year." Similarly, a Hancock teacher said, "the biggest focus that I've seen this fall…just getting kids to really be aware of their current status, grade-wise, and how to ask questions about their grades. That's been a really huge change, this year."

Teachers in Williams and Smith also suggested that students were not only more aware of their grades, but demonstrated more ownership of their grades by completing assignments and going to tutoring. A teacher in Williams said, "I'm starting to see a little bit, changes in the kids, because they are starting to take more ownership into their learning, and they ask questions that kids in years past wouldn't have asked." This increased awareness of grades stands in contrast to the culture that used to exist in the school, where students did not always link their grades with their class performance. A Williams teacher summarized this change in student attitudes around grades: "It's not just 'they gave me a grade.' Well, now I've earned this grade and I'm trying to make my grade better, and we actually talked about." By shifting the locus of control from teachers to students, students were described as taking more ownership over monitoring their classroom performance. Several Smith teachers we spoke to saw a connection between students

setting goals to improve their grades and a decrease in the number of incomplete and missing assignments. For example, a math teacher said, “I used to struggle to get my failure rate down to 20%. The last six weeks it was at 8%.”

In contrast to the other two schools, fewer Hancock teachers described systematic changes in student academic engagement and classroom behavior. For example, one Hancock teacher described a conversation with a student, “I said, ‘Well, have you talked to your teacher?’ No…So they don't see that there's a solution to that. They just sit – they would rather fail the class than to go talk to the teacher.” Another Hancock teacher said, “At least in conversation, they’ll be like, yeah, I can do this to get better. But in practice, it doesn’t always go through.”

Smith and Williams teachers were also more likely to report improved discipline as a result of SOAR. Teachers in Smith reported that the GROW sheet decreased the number of students they had to send to the principal’s office and they described how the GROW sheet impacted the way their students processed their behaviors. For example, a Smith teacher explained, “I know having something…that’s like a cool down for them, for them to have to reflect instead of just them maybe escalating and just sending them to the principal.” Another Smith teacher described how the GROW sheet was the most effective element of SOAR: “It gives a kid a chance to kind of explain themselves. Okay, what’s your problem? I’ve got a crazy scenario at home. I just got in a fight with my mom before I got to school and that’s why I’m having a bad day.” The majority of the Smith and Williams school staff we spoke to felt that disciplinary problems had gone down. In contrast, staff at Hancock felt that disciplinary issues were not a problem at the school, before or after SOAR began.

In addition to overall student outcomes, stakeholders in all schools noted heterogeneity in the uptake of various SOAR practices across different students. In general, the sentiment was

that the practices were most beneficial for moderately performing students. For example, a Williams teacher said, "The ones who are more middle of the road, you know? Students who need that little extra push…those are the ones who during the grade reporting, they talk to me seriously about why they got certain grades they got." Across schools, teachers reported that the highest performing students did not need the SOAR supports, and that the lowest-performing students were not responding to them. For example, a Hancock teacher said in Year 2,

> You always have that top 10% group and the low 10% group. But that other 80%, they kind of fluctuate. And if you take that 80%, the middle 80%, from four years ago, five years ago, of the seniors… and put them into their classes this year, I don't think they would've made it. … we're a more rigorous campus based on what we're doing every day. And I think SOAR is a big part of that."

Students across schools agreed that the SOAR practices were of differential value to students based on characteristics of the students. For example, one student in Williams explained,

> I mean I think the big problem with it is the people who will accept it are the people who don't necessarily need to accept it. You know, they're already fairly successful in school. … But then the people who really could benefit from you know, thinking I can do better, I can work hard, are the people who don't work hard and aren't willing to work hard. So the problem with it is it's like the benefit is greater, but the acceptance is lower

The heterogeneity of effects that the teachers describe is explored in more depth through the analysis of student-level district administrative data. The next section turns to these data.

Student Outcomes from District Administrative Data

Before turning to our DD results, we report on descriptive differences between innovation and non-innovation schools during the year prior to implementation (Table 2). Compared to the non-innovation schools, students in the innovation schools had slightly better grades, were absent fewer days, and had fewer disciplinary infractions. Fewer innovation school students received free or reduced price lunch or were Black, although more innovation school students were Hispanic. We found no evidence of significant differences in terms of school

characteristics. To further investigate these pre-implementation differences, we disaggregated the results by the innovation schools. Students from Hancock failed more classes than students in non-innovation schools in the district but had lower grades, on average. Students at Williams and Smith had higher grades than students in non-innovation schools and students at Smith also failed fewer courses. Students at Smith and Hancock also were absent less frequently. Compared to the district, Smith and Hancock had fewer Black students but more Hispanic students.

With slight evidence of demographic differences between the innovation and non-innovation schools—particularly at Hancock—we examine pre- and post-treatment trends in Figure 1, comparing the predicted outcomes across all periods while holding all student and school characteristics at their mean. For the number of failed classes and average student grades, these results suggest similar pre-treatment trends between innovation and non-innovation schools. This graphical evidence also suggests that, after implementation, any reduction in the number of failed classes or improvements in student grades would be small in magnitude. Figure 1 also indicates that the innovation schools did not consistently follow the pre-treatment trends for non-innovation schools in the district in terms of student absences or the number of disciplinary infractions. Williams and Smith each saw reductions in student absenteeism prior to implementing the SOAR innovation, while the district and Hancock saw little change over this period. Innovation and non-innovation schools alike saw small increases in student absences in the post-treatment period.

Table 4 reports the DD estimates of the four outcomes: classes failed, average grades, days absent, and the number of disciplinary infractions. The coefficient on innovation school indicates that students in the innovation schools failed slightly fewer classes (0.25) and had higher grades (0.90) compared to students in non-innovation schools in the district. These slight

improvements translate to relatively small effect sizes: a 0.05 standard deviation decrease in the number of failed courses and a 0.04 standard deviation increase in average class grades. We find no evidence of a relationship between participating in this school improvement model and decreases in student absences or disciplinary infractions.

When the results are separated by innovation school (Table 5), we find the improvements in student grades and the number of failed courses are concentrated rather evenly across the innovation schools. The reduction in the number of classes failed ranged from -0.14 and marginally significant at Hancock to -0.30 at Williams. Increases in average student grades ranged from 0.83 at Williams to 1.03 at Hancock. These results are consistent with the qualitative data, which found the grade monitoring and goal setting activities to be a central emphasis of the SOAR innovation, although with somewhat more emphasis in Williams and somewhat less emphasis in Hancock. Student absences decreased substantially in Williams (-1.47) but, the overall effect of the SOAR innovation on absences was offset by an increase in student absences at Hancock (0.66). We found no evidence of decreased disciplinary infractions among any innovation school, including Smith, which was the school that focused the most on the behavioral reflection form. Overall, these results indicate that Williams had the most consistently positive effect on student outcomes, a finding that is substantiated by the qualitative data which showed Williams to have the strongest implementation of both routines and sense of coherence around the routines. In Appendix Tables A1 and A2, we examine whether these differences are driven by the first or second year of implementation, but generally find that the year 1 and year 2 effects are comparable. It is worth noting that the effects in Smith are smaller in the second year of implementation, the year in which they stopped schoolwide implementation midway through the school year.

Student Heterogeneity in the Impact of the SOAR Innovation

The qualitative data suggested that an area of heterogeneity was in relation to student ability levels. Teachers and other school stakeholders suggested the SOAR innovation was positioned to have the greatest effects on moderately performing students. They suggested that moderately performing students were more likely to adopt the innovation practices than lower performing students. Higher performing students, on the other hand, were likely to already use some of the practices that were taught as part of the SOAR innovation. To test this hypothesis, in Table 6, we estimate a series of regressions to predict the effect of the SOAR innovation based on prior student performance. We separate students into groups based on their prior performance for each outcome. As this prior performance variable would be endogenous to treatment in the second year, we limit this analysis to the first year of implementation. In general, we provide little evidence to support the hypothesis that SOAR was most beneficial for moderately performing students. If anything, SOAR was most beneficial for the highest performing students, with the exception of the number of failed courses, where students who had failed the most had the largest improvements. We also find little consistency across schools. For instance, at Williams, SOAR was linked with a 2.48 decrease in the number of days absent at Williams among students who had previously been absent 15 days or more. Yet, at Hancock, this same group was absent 2.79 days more.

We also explored heterogeneous effects by federally identified student subgroups (Table A3). We found no evidence of a relationship between the innovation for Black students in terms of number of failed classes or average grades, although we found marginally significant evidence that they were absent 1.62 days less during implementation. Hispanic and economically disadvantaged students failed fewer courses (-0.29 and -0.27, respectively) and had higher grades

(1.05 and 1.03, respectively).

Sensitivity Analysis

A concern with this DD analysis is that positive outcomes attributed to the innovation designed through the continuous improvement reform model are a result of the innovation schools being selected to participate in this program based on unobserved, time-varying characteristics. For example, schools selected to participate in this improvement process may have unobserved, time-varying characteristics that would make them more likely to improve student outcomes, regardless of their participation in this process. If this scenario were true, we would be worried that the unobserved, time-varying factors that led district stakeholders to select the innovation schools in the first place explain the school improvements rather than actual participation in the improvement process and the implementation of the SOAR innovation. A related concern is that innovation schools could have been selected based on past student outcomes. To the extent to which prior student outcomes were related to any transitory shock, any post-treatment improvements may arise from regression to the mean.

In Figure 1, we provided graphical evidence of this parallel trends assumption, demonstrating that the innovation schools, all followed a similar pre-treatment trend as the district in terms of the number of failed classes and average grades, but not consistently in terms of days absent or the number of disciplinary infractions. That our most consistent results pertained to these course failure and grades provides stronger evidence that the grade monitoring activities developed and implemented as part of this improvement process were linked with meaningful, albeit small, improvements to these outcomes. Tables A4 and A5 further examine differences in pre-treatment trends. We find marginally significant evidence that innovation schools had lower grades in 2013 and fewer disciplinary infractions in 2014. When separated out

by school, we find some evidence of pre-treatment differences in the trends of the outcomes, although the pre-treatment differences are concentrated in the number of days absent and disciplinary infractions. Most notably, Hancock consistently had higher course failure rates, lower grades, a higher absentee rate, and a higher and lower disciplinary infraction rate, depending on the year. This suggests that Hancock differed from the district in ways that could have shaped its uptake of the innovation and the resulting effect on students. We find some similar evidence at Smith, although it is strongest for days absent and disciplinary infractions, outcomes for which we did not find consistent evidence of improvement.

Discussion

Overall, we find that implementation of the SOAR innovation developed through the continuous improvement process saw small, yet desirable and statistically significant improvements in student grades and course passing. There were some differences by schools, which may be explained by the implementation emphasis in each schools. In particular, Hancock focused more on cultural and relationship changes, with less emphasis on teachers changing their routine practice. Relatedly, Hancock staff reported shifts in school culture, but were more hesitant to link SOAR to tangible student behaviors. The data also indicate that Williams had the strongest implementation of both routines and sense of coherence around the routines, and the quantitative evidence suggests the most positive outcomes on students. The importance of both routines and coherence of routines is consistent with research that suggests sustainability requires deep understanding of the larger goals that a focus on adopting highly structured practices may overlook (Rubin, Patrick, & Goldring, 2017). The findings further suggest that the impact of the SOAR innovation differed by prior student outcomes. This suggests that innovations designed to address chronic absenteeism or chronically low grades may not be the same as those that can

support students at the other end of the distribution. This is particularly importance in the context of a research-practice partnership, where there is an increased emphasis in not just knowing whether an innovation works, but for whom it works (Means & Penuel, 2005).

There are several potential limitations to this study. One limitation is that some of the outcome variables could be seen as endogenous. Indeed, teachers were responsible for implementing the innovation and assigning grades, determining which students failed, and writing disciplinary infractions. It is possible that teachers may have artificially raised student grades or failed fewer students. Other evidence from the district suggests this is unlikely. In particular, given the multi-year partnership with the district, we heard about pressures that teachers receive from the district to not fail students that pre-date the SOAR innovation. Indeed, helping teachers maintain academic press in a context that is focused on credit accumulation and increasing graduation rates was one of the findings that led to a focus on student ownership in the first place (Cannata, Smith, & Taylor Haynes, 2017). In this way, teachers across schools indicated that SOAR allowed them to hold students accountable. The endogeneity of outcomes may also be a problem for the number of disciplinary infractions, as the use of the behavioral reflection form before a referral may more directly decrease the number of documented infractions. However, we found no evidence that implementation of SOAR had an impact on the number of infractions.

Another potential limitation is the diffusion of the SOAR innovation beyond the three innovation schools. The SOAR design team did include administrators from two other high schools, each of which indicated some adoption of a few SOAR practices in their schools. While evidence on implementation in these schools is less systematic, the administrators report such practices were diffused on a small scale, such as to teachers in the department that administrator

oversaw. Consequently, our results would then be underestimated if improvements were also in other comparison schools in the district.

A broader challenge comes from evaluating outcomes from the specific SOAR innovation in the context of a research-practice partnership. It may be that the outcomes documented here come less from implementation of specific SOAR practices than from the process of participating in the larger process of continuous improvement and sharing lessons in the districtwide network. There is limited ability to methodologically untangle whether the innovation itself is responsible for changes in student outcomes or the school-based improvement process that led to a contextually based design. In other words, the innovation design and design process are both part of the treatment. As a result, evidence of positive changes in student outcomes would be need to generalized to other sites with caution, given the contextually sensitive reform process. That being said, a distinctive feature of the continuous improvement approach is the linkage between specific design features of the innovation with relevant student outcomes. In other words, this research-practice partnership allowed for the testing of hypotheses of the effects of the SOAR innovation that were generated by district stakeholders.

Despite these limitations, we still believe these findings have practical significance for two main audiences. First, it provides evidence on the first two years of outcomes of an innovation designed to build student co-cognitive traits and thus will be of interest to other researchers and practitioners focused on these traits. This study adds to the growing literature on outcomes from interventions that focus on co-cognitive traits (Blackwell, Trzesniewski, & Dweck, 2007; Cutts et al., 2010; Gunderson et al., 2013; Yeager & Walton, 2011; Yeager et al., 2016). These findings are valuable in the context of recent reforms aimed at improving student outcomes other than test scores, such as grades, attendance, and discipline. Evidence from this

paper indicates that the SOAR innovation developed through the continuous improvement reform model made meaningful improvements in students' grades and decreases to student failure rates. In particular, failing fewer courses has the potential to have long-term implications for credit accumulation and on-track graduation (Bowers, Sprott, & Taff, 2013). That being said, we found no evidence that this innovation improved the grades of the lowest performing students or reduced absenteeism among chronically absent students. These heterogeneous effects highlight the need for future studies of co-cognitive interventions to attend to such differences.

Second, the practices were developed through a collaborative design process focused around a research-practice partnership trying to scale effective practices. This partnership approach is gaining popularity with the assumption that greater attention to the context of implementation will result in more sustained improvements at scale (Means & Penuel, 2005). This study provides evidence that these types of partnerships not only contribute to greater ownership and commitment from participants (Cannata, Rutledge, Redding, Smith, & Rubin, 2017), but can also lead to demonstrated improvements in student outcomes. The school-level differences we observed highlighted the extent to which school design teams aimed to change school culture versus routines of practice. The evidence provided here suggests that the focus on changing routines rather than less tangible changes to culture were an important mechanism in improving student outcomes. This finding is consistent with the conceptual underpinnings of continuous improvement, that aim to more tightly link school inputs with specific student outcomes (Bryk et al., 2015).

References

Angrist, J. D. & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

Arcia, E. (2006). Achievement and enrollment status of suspended students. *Education and Urban Society*, 38, 359–369.

Cannata, M., Cohen-Vogel, L., & Sorum, M. (2017). Partnering for Improvement: Improvement Communities and their Role in Scale Up. *Peabody Journal of Education*.

Cannata, M., & Nguyen, T. (2015). *Consensus versus Clarity: Tensions in Designing for Scale.* Presentation at the annual meeting of the American Educational Research Association in Chicago, April 16-20, 2015.

Cannata, M., & Rutledge, S. A. (2017). New Frontiers in Scaling Up Research. *Peabody Journal of Education*.

Cannata, M., Redding, C., Brown, S., Joshi, E., & Rutledge, S., Joshi, E. (2017). *How Ideas Spread: Establishing a Networked Improvement Community.* Presentation at the Carnegie Foundation Summit on Improvement in Education in San Francisco, CA, March 27-29, 2017 and the annual meeting of the American Educational Research Association in San Antonio, TX, April 27-May 1, 2017.

Cannata, M., Rutledge, S. A., Redding, C., Rubin, M., & Smith, T. M. (2017). *Using Continuous Improvement to Achieve Scale: Implications for Depth, Spread, Ownership, and Sustainability.* Paper presented at the annual meeting of the Association of Education Finance and Policy in Washington, DC, March 16-18, 2017 and the annual meeting of the American Educational Research Association in San Antonio, TX, April 27-May 1, 2017.

Cannata, M., Smith, T.M., & Taylor Haynes, K. (2017). Integrating Academic Press and Support by Increasing Student Ownership and Responsibility. *AERA Open 3*(3): 1-13.

Balfanz, R., & Byrnes, V. (2012). *Chronic Absenteeism: Summarizing What We Know From Nationally Available Data*. Baltimore, MD: Johns Hopkins University Center for Social Organization of Schools.

Bernard, H. R. & Ryan, G. W. (2010). *Analyzing qualitative data: Systematic approaches*. Thousand Oaks, CA: Sage Publications.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics, 119*, 249-275.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, *78*(1), 246–263.

Blazar, D., & Kraft, M. A. (2017). Teacher and Teaching Effects on Students' Attitudes and Behaviors. *Educational Evaluation and Policy Analysis*, *39*(1), 146–170.

Booth, J. L., Cooper, L. A., Donovan, M. S., Huyghe, A., Koedinger, K. R., & Paré-Blagoev, E. J. (2015). Design-Based Research Within the Constraints of Practice: Algebra By Example. *Journal of Education for Students Placed at Risk (JESPAR)*, *20*(1–2), 79–100. doi:10.1080/10824669.2014.986674

Bowers, A. J. (2009). Reconsidering grades as data for decision making: more than just academic knowledge. *Journal of Educational Administration*, *47*(5), 609–629.

Bowers, A. J., Sprott, R., & Taff, S. A. (2013). Do we know who will drop out? a review of the predictors of dropping out of high school: precision, sensitivity, and specificity. *High School Journal*, *96*(2), 77.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., … Welsh, M. E. (2016). A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, *86*(4), 803–848.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to Improve: How America's Schools Can Get Better at Getting Better*. Cambridge, MA: Harvard Education Press.

Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What's the evidence on districts' use of evidence. *The Role of Research in Educational Improvement*, 67–87.

Coburn, C. E., & Penuel, W. R. (2016). Research–Practice Partnerships in Education Outcomes, Dynamics, and Open Questions. *Educational Researcher*, *45*(1), 48–54.

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-Practice Partnerships: A Strategy for Leveraging Research for Educational Improvement in School Districts*. New York, NY: William T. Grant Foundation.

Cohen, D. K., Peurach, D. J., Glazer, J. L., Gates, K. E., & Goldin, S. (2013). *Improvement by Design: The Promise of Better Schools*. Chicago ; London: University Of Chicago Press.

Cohen-Vogel, L., Cannata, M., Rutledge, S.A., & *Socol, A.R. (2016). A Model of Continuous Improvement in High Schools: A Process for Research, Innovation Design, Implementation, and Scale. *Teachers College Record Yearbook (Yearbook) 118*(13): pp. 1-26.

Cohen-Vogel, L., Tichnor-Wagner, A., Allen, D., Harrison, C., Kainz, K., Socol, A. R., & Wang, Q. (2015). Implementing educational innovations at scale: Transforming researchers into continuous improvement scientists. *Educational Policy*, *29*(1), 257-277.

Creswell, J. W. (2009). *Research design: qualitative, quantitative, and mixed method*

approaches. Thousand Oaks, Calif.: Sage Publications.

Datnow, A., & Park, V. (2009). Conceptualizing policy implementation: Large-scale reform in an era of complexity. In G. Sykes, B. Schneider, & D. N. Plank (Eds.), *Handbook of Education Policy Research* (1 edition, pp. 348–361). Washington, D.C.: Routledge.

DePaoli, J., Fox, J. H., Ingram, E., Maushard, M., Bridgeland, J. M., & Balfanz, R. (2015). *Building a grad nation: Progress and challenges in ending the high school dropout epidemic*. Civic Enterprises and Everyone Graduates Center at the School of Education at Johns Hopkins University.

Dede, C. (2006). Scaling up: Evolving innovations beyond ideal settings to challenging contexts of practice. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 551–566). Cambridge, UK: Cambridge University Press.

Donovan, M. S., Snow, C., & Daro, P. (2013). The SERP approach to problem-solving research, development, and implementation. *National Society for the Study of Education Yearbook, 112*(2), 400–425.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., & Giffin, J. (2017). *School Improvement Grants: Implementation and Effectiveness* (NCEE 2017- 4013). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement Matters Assessing Personal Qualities Other Than Cognitive Ability for Educational Purposes. *Educational Researcher*, *44*(4), 237–251. https://doi.org/10.3102/0013189X15584327

Dweck, C. S. (2007). *Mindset: The New Psychology of Success* (Reprint). Ballantine Books.

Faria, A.-M., Sorenson, N., Heppen, J., Bowdon, J., Taylor, S., Eisner, R., & Foster, S. (2017). *Getting students on track for graduation: Impacts of the Early Warning Intervention and Monitoring Systems after one year* (No. REL 2017-272). Washington, D.C.: National Center for Education Evaluation and Regional Assistance.

Farrington, C., Roderick, M., Allensworth, E., Nagaoka, J., Seneca Keyes, T., Johnson, D., & Beechum, N. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance: A Critical Literature Review*. Chicago: The University of Chicago Consortium on Chicago School Research.

Fishman, B. J., Penuel, W. R., Allen, A.R., & Cheng, B. H. (2013). *Design-based implementation research: theories, methods, and exemplars*. New York, NY: National Society for the Study of Education.

Ginsburg, A., Jordan, P., & Chang, H. (2014). *Absences Add Up: How School Attendance Influences Student Success*. Portland, OH: Attendance Works.

Gottfried, M. A. (2011). The Detrimental Effects of Missing School: Evidence from Urban Siblings. *American Journal of Education*, *117*(2), 147–182. doi:10.1086/657886

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap two sides of the same coin? *Educational Researcher*, *39*(1), 59-68.

Gross, B., Booker, T. K., & Goldhaber, D. (2009). Boosting Student Achievement: The Effect of Comprehensive School Reform on Student Achievement. *Educational Evaluation and Policy Analysis*, *31*(2), 111–126.

Hannan, M., Russell, J. L., Takahashi, S., & Park, S. (2015). Using Improvement Science to Better Support Beginning Teachers: The Case of the Building a Teaching Effectiveness Network. *Journal of Teacher Education*, *66*(5), 494–508.

Langley, G. J. (2009). *The improvement guide: a practical approach to enhancing organizational performance*. San Francisco: Jossey-Bass.

LeMahieu, P. G., Grunow, A., Baker, L., Nordstrum, L. E., & Gomez, L. M. (2017). Networked Improvement Communities: the discipline of improvement science meets the power of networks. *Quality Assurance in Education*. doi:10.1108/QAE-12-2016-0084

Mac Iver, M. A., & Messel, M. (2013). The ABCs of Keeping On Track to Graduation: Research Findings from Baltimore. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 50–67. doi:10.1080/10824669.2013.745207

Mazzeo, C., Fleischman, S., Heppen, J., & Jahangir, T. (2016). Improving High School Success: Searching for Evidence of Promise. *Teachers College Record*, *116*(13), 1–32.

Means, B., & Penuel, W. R. (2005). Scaling Up Technology-Based Educational Innovations. In C. Dede, J. P. Honan, & L. C. Peters (Eds.), *Scaling Up Success : Lessons Learned from Technology-Based Educational Improvement* (1st ed., pp. 176–197). San Francisco, CA: Jossey-Bass.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook. Second edition*. Thousand Oaks, CA: Sage Publications.

Nunnery, J. A. (1998). Reform Ideology and the Locus of Development Problem in Educational Restructuring: Enduring Lessons from Studies of Educational Innovation. *Education and Urban Society*, *30*(3), 277–295.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (Third edition). Thousand Oaks, CA: Sage Publications.

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational*

Researcher, *40*(7), 331–337.

Raffaele Mendez, L. M. (2003). Predictors of suspension and negative school outcomes: A longitudinal investigation. In J. Wald & D. J. Losen (Eds.), *New directions for youth development* (pp. 17–34). San Francisco, CA: Jossey-Bass.

Rausch, M. K., & Skiba, R. (2004). *The academic cost of discipline: The relationship between suspension/expulsion and school achievement.* Center for Evaluation and Education Policy, School of Education, Indiana University.

Renzulli, J. S. (2011). Theories, Actions, and Change: An Academic Journey in Search of Finding and Developing High Potential in Young People. *Gifted Child Quarterly*, *55*(4), 305–308. https://doi.org/10.1177/0016986211421875

Rocque, M. (2010). Office discipline and student behaviors: Does race matter? American Journal of Education, 116(4), 557–581. doi:10.1086/653629

Rubin, M., Patrick, S. K., & Goldring, E. B. (2017). Make it quick or make it last? Dilemmas in the implementation of a school reform. *Peabody Journal of Education*.

Russell, J. L., Bryk, A. S., Dolle, J. R., Gomez, L. M., LeMahieu, P. G., & Grunow, A. (2017). A Framework for the Initiation of Networked Improvement Communities. *Teachers College Record*, *119*(7).

Russell, J. L., Correnti, R., Stein, M. K., Hannan, M., Bill, V., Schwartz, N., … Pratt, N. (2015). Using data for improvement to support implementation at-scale: Adaptive integration in the TN mathematics coaching project. Presented at the "Using Continuous Improvement to Integrate Design, Implementation, and Scale" conference, Nashville, TN.

Siskin, L. S. (2016). Mutual Adaptation in Action. *Teachers College Record*, *118*(13), 1–18.

Smith, T.M., Cannata, M., & Taylor Haynes, K. (2016). Reconciling data from different sources: Practical realities of using mixed methods to identify effective high school practices. *Teachers College Record 118*(7): 1-34.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating Knowledge of Academic Language Among Urban Middle School Students. *Journal of Research on Educational Effectiveness*, *2*(4), 325–344. doi:10.1080/19345740903167042

Sowers, N., & Yamada, H. (2015). *Community College Pathways: 2013-2014 Descriptive Report*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from https://eric.ed.gov/?id=ED554591

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools*, *13*(1), 12–28.

Yeager, D. S., & Walton, G. M. (2011). Social-Psychological Interventions in Education They're Not Magic. *Review of Educational Research*, *81*(2), 267–301.

Yeager, D. S., Walton, G., Brady, S., Akcinar, E. N.*, Paunesku, D., Keane, L., Kamentz, D., ... & Dweck, C. S. (2016). Teaching a lay theory before college narrows achievement gaps at scale. *Proceedings of the National Academy of Sciences of the United States of America, 113,* E3341-E3348.

Tables and Figures

Figure 1. Pre- and Post-Treatment trends of Student Passing Rates, Average Grades, Attendance, and Number of Infractions

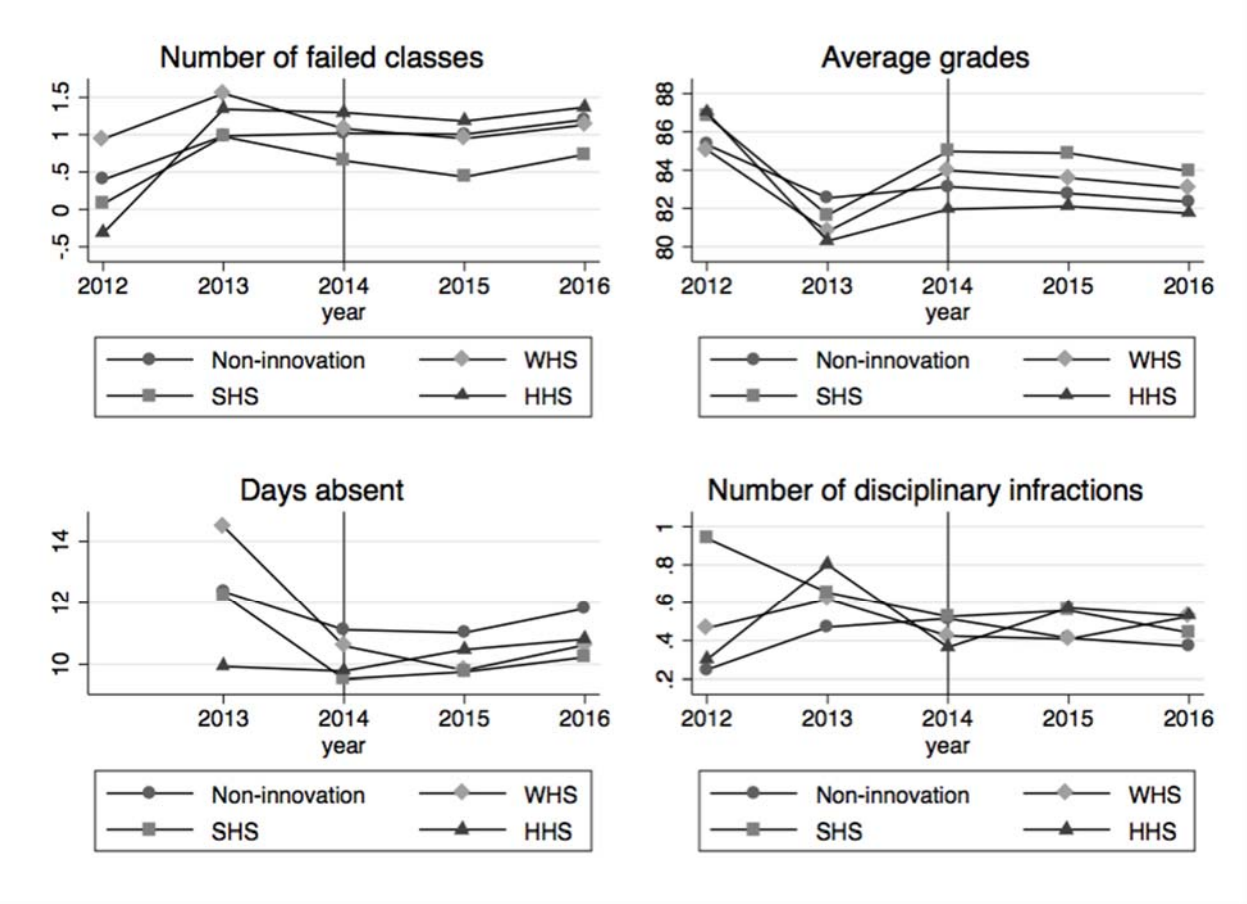

Table 1. Fieldwork Data Sources

	Year 1		Year 2
	October 2014	April 2015	March 2016
Interviews			
Implementation team members	20	22	30
Teacher	71	70	32
Administrator	11	12	3
Focus Group			
Teacher/ Support Staff	11	8	0
Student	12	14	8

Table 2. Descriptive Characteristics Prior to Implementation

	Non-Innovation Schools	Innovation Schools (all)	Williams	Smith	Hancock
Number of failed classes	1.10	1.07	1.25	0.74**	1.35*
Average grade	82.22	83.03***	83.19*	83.84**	81.28*
Days absent	11.36	9.98***	10.39	9.91*	9.32**
Number of disciplinary infractions	0.58	0.45***	0.40	0.58	0.33
Free or Reduced Price Lunch	0.69	0.64***	0.44***	0.74	0.84*
Black student	0.25	0.14***	0.20	0.12*	0.05**
Hispanic student	0.59	0.69***	0.46	0.80*	0.92***
Other race	0.04	0.03***	0.04	0.03	0.01**
Gifted	0.12	0.12	0.17	0.07*	0.11
Days enrolled	169.42	169.70*	170.74	168.73	169.45
Withdrew	0.13	0.13	0.16	0.11	0.13
Late start	0.10	0.10	0.10	0.09	0.09
Number of Courses	13.21	12.85***	12.99	12.89	12.52**
Fraction of Black students	0.25	0.14	0.21	0.12	0.05*
Fraction of Hispanic students	0.58	0.68	0.45	0.79*	0.92**
Fraction FRPL	0.62	0.59	0.40**	0.68	0.79**
School size	1766.30	1740.51	2010.00	1859.00	1016.00
Observations	14406	4439	1798	1695	946

Notes. t-test of significant differences accounts for school-level clustering. Descriptive statistics reported for 2013-2014 school year. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Table 3. Summary of Implementation by School

	Williams	Hancock	Smith
Goal-setting and grade-monitoring	Main emphasis of SOAR with widespread use of grade monitoring routine every three weeks	Grade-monitoring routine one of several lessons delivered through advisory	Some teachers considered this the main priority of SOAR
Growth mindset	Strong initial implementation on first day, but not sustained follow up	Single isolated lesson	Training for teachers, but no formal introduction to students
Problem-solving	Beyond the SOAR team, no awareness of problem-solving as component of SOAR	SOAR team defined problem-solving as a set of lessons focused on college and career readiness, which was the main priority of SOAR	Schoolwide problem-solving process, called IPAC; Some teachers considered this the main priority of SOAR
Behavioral reflection	Little emphasis in the school, minimal engagement by teachers	Only used in 9th grade	GROW sheet a main emphasis of SOAR for many teachers, although less frequent use over time
Schoolwide culture	Cultural emphasis was on student responsibility and grade-monitoring routine	Cultural emphasis was on building student-teacher relationships and mentoring	No consistent cultural emphasis in the school
Other	SOAR activities mostly implemented in extended second period every three weeks	SOAR activities mostly constrained to newly created advisory period which met every week	SOAR activities mostly implemented in extended second period every three weeks; Administration ended SOAR participation midway in Year 2

Table 4. Estimates from School Fixed Effects Model of the Effect of the Innovation on Student Passing Rates, Average Grades, Attendance, and Number of Infractions

	Number of failed classes	Average grades	Days absent	Number of disciplinary infractions
Innovation school	-0.25**	0.90**	-0.57	0.08
	(0.07)	(0.28)	(0.56)	(0.08)
Free or Reduced Price Lunch	0.04	-0.00	0.02	-0.01
	(0.03)	(0.14)	(0.18)	(0.02)
Black student	0.05	-0.68**	-0.34**	0.24***
	(0.05)	(0.22)	(0.11)	(0.05)
Hispanic student	0.12*	-0.84**	-0.27+	-0.02
	(0.05)	(0.26)	(0.13)	(0.02)
Other race	-0.08	0.24	-0.58**	-0.10*
	(0.05)	(0.20)	(0.18)	(0.04)
Gifted	-0.27***	1.36***	-0.71***	-0.08***
	(0.03)	(0.13)	(0.13)	(0.01)
Days enrolled	-0.00**	0.03***	0.09***	0.00+
	(0.00)	(0.01)	(0.01)	(0.00)
Withdrew	0.51***	-2.30***	3.73***	0.44***
	(0.07)	(0.20)	(0.46)	(0.05)
Late start	-0.19*	1.24***	0.80*	0.31***
	(0.08)	(0.25)	(0.33)	(0.06)
Number of Courses	0.07***	0.14*	-0.47***	-0.02
	(0.01)	(0.05)	(0.09)	(0.01)
10th grade	-0.42***	0.87*	0.68**	-0.13**
	(0.08)	(0.34)	(0.20)	(0.04)
11th grade	-0.55***	1.56***	0.56*	-0.26***
	(0.10)	(0.32)	(0.23)	(0.05)
12th grade	-0.94***	3.21***	1.53***	-0.32***
	(0.08)	(0.32)	(0.35)	(0.07)
Lagged dependent variable	0.47***	0.70***	0.73***	0.41***
	(0.02)	(0.01)	(0.03)	(0.02)
Constant	-0.01	19.29***	-6.71***	0.01
	(0.16)	(1.37)	(0.92)	(0.19)
Total Observations	58817	58811	60456	62408
Unique Observations	32474	32471	32710	33215
R-squared	0.28	0.55	0.42	0.28

Notes. Models include school and year fixed effects. Robust standard errors clustered at the school level in parentheses. * p<0.05; ** p<0.01; *** p<0.001.

Table 5. Estimates from School Fixed Effects Model of the Effect of the Innovation on Student Passing Rates, Average Grades, Attendance, and Number of Infractions, by Innovation School

	Number of failed classes	Average grades	Days absent	Number of disciplinary infractions
Williams HS	-0.30***	0.83*	-1.47***	0.10
	(0.07)	(0.28)	(0.30)	(0.07)
Smith HS	-0.25**	0.92**	-0.26	0.03
	(0.07)	(0.27)	(0.30)	(0.07)
Hancock HS	-0.14+	1.03**	0.66+	0.14
	(0.08)	(0.28)	(0.31)	(0.08)
Total Observations	58817	58811	60456	62408
Unique Observations	32474	32471	32710	33215
R^2	0.28	0.55	0.42	0.28

Notes. Models control for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Robust standard errors clustered at the school level in parentheses. + $p<0.10$; * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Table 6. Estimates from School Fixed Effects Model of the Effect of the Innovation on Student Passing Rates, Average Grades, Attendance, and Number of Infractions, by Lagged Student Outcomes

	Innovation school	Williams HS	Smith HS	Hancock HS
Number of failed courses (lagged)				
0 course	-0.18***	-0.16***	-0.22***	-0.12***
1 failed course	-0.38***	-0.42***	-0.46***	-0.09*
2 failed courses	-0.49*	-0.76***	-0.20**	-0.37***
3 or 4 failed courses	-0.39***	-0.39***	-0.55***	-0.13*
5 or more failed courses	-0.53***	-0.68***	-0.54***	-0.13+
Average grade (lagged)				
Below 60	-0.23	4.80***	-1.37	-7.83***
60-69	0.70	1.31***	-1.61***	1.34***
70-79	0.65	1.78***	-0.21+	0.01
80-89	0.64*	0.09	1.05***	0.83***
90-100	0.98*	0.47***	1.53***	1.47***
Days absent (lagged)				
0-3 days	-0.49+	-0.99***	-0.10**	-0.26***
4-7 days	-0.63	-1.47***	-0.11**	0.18**
8-14 days	-1.22	-2.82***	-0.67***	1.23***
15 days or more	-0.35	-2.48***	0.02	2.79***
Number of infractions (lagged)				
0 infractions	-0.06***	-0.06***	-0.05***	-0.07***
1 infraction	-0.09	-0.22***	-0.17***	0.34***
2 or more infractions	0.20	-0.13*	0.22**	0.73***

Notes. Models control for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Standard errors clustered at the school level in parentheses. + p<0.10; * p<0.05; * p<0.01; * p<0.001.

Building Student Ownership and Responsibility: Examining Student Outcomes from a Research-Practice Partnership

ONLINE APPENDIX

Additional Tables and Figures

Table A1. Estimates from School Fixed Effect Model of the Effect of the Innovation on Student Passing Rates, Average Grades, Attendance, and Number of Infractions

	Number of failed classes	Average grades	Days absent	Number of disciplinary infractions
Innovation school (1st treatment year)	-0.28**	1.06**	-0.53	0.06
	(0.08)	(0.28)	(0.66)	(0.09)
Innovation school (2nd treatment year)	-0.22*	0.75*	-0.61	0.10
	(0.09)	(0.33)	(0.52)	(0.10)
Observations	58817	58811	60456	62408
R^2	0.28	0.55	0.42	0.28

Notes. Models control for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Robust standard errors clustered at the school level in parentheses. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Table A2. Estimates from School Fixed Effect Model of the Effect of the Innovation on Student Passing Rates, Average Grades, Attendance, and Number of Infractions, by Innovation School

	Number of failed classes	Average grades	Days absent	Number of disciplinary infractions
Williams HS (Yr 1)	-0.32**	1.00**	-1.65***	0.00
	(0.08)	(0.28)	(0.31)	(0.08)
Williams HS (Yr 2)	-0.27**	0.66+	-1.31**	0.18*
	(0.08)	(0.33)	(0.37)	(0.08)
Smith HS (Yr 1)	-0.31**	1.15***	-0.07	0.07
	(0.07)	(0.27)	(0.30)	(0.08)
Smith HS (Yr 2)	-0.20*	0.71*	-0.42	-0.00
	(0.08)	(0.32)	(0.37)	(0.08)
Hancock HS (Yr 1)	-0.14	0.99**	0.89*	0.14
	(0.09)	(0.30)	(0.32)	(0.08)
Hancock HS (Yr 2)	-0.15	1.07**	0.45	0.13
	(0.09)	(0.34)	(0.40)	(0.09)
Observations	58817	58811	60456	62408
R^2	0.28	0.55	0.42	0.28

Notes. Models control for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Robust standard errors clustered at the school level in parentheses. * $p<0.05$; ** $p<0.01$; *** $p<0.001$.

Table A3. Estimates from Difference-in-Difference Model of the Effect of the Innovation on Student Passing Rates, Average Grades, Attendance, and Number of Infractions, by Student Subgroups

	Number of failed classes			Average grades			Days absent			Number of disciplinary infractions		
Innovation school	-0.24	-0.29**	-0.27**	0.57	1.05**	1.03**	-1.62+	-0.22	-0.60	0.25	0.06	0.07
	(0.14)	(0.08)	(0.07)	(0.51)	(0.29)	(0.30)	(0.75)	(0.51)	(0.55)	(0.22)	(0.06)	(0.08)
Black student	x			x			x			x		
Hispanic student		x			x			x			x	
Free or Reduced Price Lunch			x			x			x			x
Observations	12118	37419	40273	12117	37414	40272	12679	38387	42050	13125	39618	43004
R-squared	0.28	0.28	0.27	0.51	0.53	0.51	0.39	0.43	0.42	0.30	0.24	0.28

Notes. Models control for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Robust standard errors clustered at the school level in parentheses. + p<0.10; * p<0.05; * p<0.01; * p<0.001.

A4. Parallel Trends Analysis

	Number of failed classes	Average grades	Days absent	Number of disciplinary infractions
Innovation school	-0.16	1.17	0.14	0.34+
	(0.49)	(1.27)	(1.18)	(0.17)
2013	1.33*	-4.93*		0.23
	(0.55)	(2.25)		(0.26)
2014	1.26*	-4.02+	-1.47+	0.28
	(0.48)	(1.96)	(0.81)	(0.27)
2015	1.22*	-4.29*	-1.62+	0.17
	(0.49)	(1.89)	(0.85)	(0.18)
2016	1.49*	-4.96*	-0.67	0.13
	(0.50)	(1.95)	(0.41)	(0.23)
Innovation school*2013	0.36	-2.33+		-0.14
	(0.41)	(1.30)		(0.18)
Innovation school*2014	0.01	-0.08	-1.40	-0.41*
	(0.49)	(1.47)	(0.83)	(0.17)
Innovation school*2015	-0.11	0.05	-1.30	-0.26
	(0.49)	(1.36)	(1.13)	(0.15)
Innovation school*2016	-0.03	-0.33	-1.46	-0.22
	(0.49)	(1.31)	(1.06)	(0.20)
Observations	58815	58809	60454	62406
R-squared	0.27	0.54	0.41	0.27

Notes. Models controls for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Robust standard errors clustered at the school level in parentheses. + $p<0.10$; * $p<0.05$; * $p<0.01$; * $p<0.001$.

Table A5. Parallel Trends, by Innovation School

	(1)	(2)	(3)	(4)
	Number of failed classes	Average grades	Days absent	Number of disciplinary infractions
Williams HS	0.55*	-0.26	2.11*	0.22*
	(0.21)	(0.95)	(0.81)	(0.07)
Smith HS	-0.32+	1.51**	-0.12	0.69***
	(0.15)	(0.41)	(0.78)	(0.09)
Hancock HS	-0.71*	1.69+	-2.44	0.05
	(0.25)	(0.81)	(1.52)	(0.16)
2013	0.59	-2.79		0.22
	(0.53)	(2.70)		(0.31)
2014	0.62	-2.19	-1.25	0.27
	(0.46)	(2.24)	(0.94)	(0.31)
2015	0.61	-2.55	-1.35	0.17
	(0.46)	(2.10)	(1.01)	(0.22)
2016	0.80+	-2.99	-0.56	0.12
	(0.44)	(2.19)	(0.50)	(0.28)
Williams HS*2013	0.02	-1.48		-0.07
	(0.28)	(1.42)		(0.12)
Williams HS*2014	-0.49+	1.11	-2.64***	-0.31
	(0.26)	(1.26)	(0.33)	(0.20)
Williams HS*2015	-0.60*	1.07	-3.32***	-0.22*
	(0.26)	(1.10)	(0.59)	(0.09)
Williams HS*2016	-0.62*	0.97	-3.30***	-0.06
	(0.21)	(0.98)	(0.51)	(0.12)
Smith HS*2013	0.31+	-2.40**		-0.51***
	(0.16)	(0.70)		(0.06)
Smith HS*2014	-0.05	0.34	-1.47***	-0.68***
	(0.13)	(0.59)	(0.27)	(0.11)
Smith HS*2015	-0.25	0.59	-1.15**	-0.55***
	(0.20)	(0.80)	(0.37)	(0.09)
Smith HS*2016	-0.14	0.10	-1.46+	-0.62***
	(0.20)	(0.84)	(0.68)	(0.10)
Hancock HS*2013	1.07***	-3.93***		0.28*
	(0.20)	(0.82)		(0.12)
Hancock HS*2014	0.99***	-2.87**	1.10*	-0.20+
	(0.20)	(0.79)	(0.36)	(0.11)
Hancock HS*2015	0.89**	-2.37*	1.89**	0.11

	(0.27)	(1.04)	(0.46)	(0.15)
Hancock HS*2016	0.88**	-2.28*	1.44**	0.11
	(0.26)	(0.99)	(0.36)	(0.15)
Observations	58815	58809	60454	62406
R-squared	0.28	0.54	0.42	0.27

Notes. Models controls for lagged dependent variable, FRPL, student race/ethnicity (Black, Hispanic, other race), gifted status, days enrolled, number of courses, and indicators if the student started school after the beginning of the school year or withdrew before the end of the year. Models include grade fixed effects. Robust standard errors clustered at the school level in parentheses. + p<0.10; * p<0.05; * p<0.01; * p<0.001.

Table A6. Estimates from Negative Binomial Regression Model of the Effect of the Innovation on Student Passing Rates, Attendance, and Number of Infractions

	Number of failed classes	Days absent	Number of disciplinary infractions
Innovation school	0.75***	0.95	0.96
	(0.07)	(0.05)	(0.12)
Free or Reduced Price Lunch	1.09+	1.00	1.05
	(0.06)	(0.02)	(0.05)
Black student	1.19	0.97	1.93***
	(0.13)	(0.02)	(0.19)
Hispanic student	1.35*	0.96**	1.09
	(0.16)	(0.01)	(0.08)
Other race	0.93	0.90***	0.79
	(0.09)	(0.02)	(0.14)
Gifted	0.66***	0.89***	0.57***
	(0.04)	(0.01)	(0.04)
Days enrolled	1.00	1.01***	1.01***
	(0.00)	(0.00)	(0.00)
Withdrew	1.65***	1.32***	2.05***
	(0.09)	(0.04)	(0.08)
Late start	0.99	1.08***	1.46***
	(0.04)	(0.02)	(0.07)
Number of Courses	1.03*	0.97***	0.94***
	(0.02)	(0.01)	(0.02)
10th grade	0.76***	1.10***	0.80**
	(0.04)	(0.02)	(0.07)
11th grade	0.67***	1.11***	0.53***
	(0.05)	(0.03)	(0.04)
12th grade	0.32***	1.25***	0.42***
	(0.02)	(0.04)	(0.04)
Lagged dependent variable	1.35***	1.06***	1.47***
	(0.02)	(0.00)	(0.05)
Observations	58817	60456	62408

Notes. Estimates reported as incidence risk ratios. Models include school and year fixed effects. Robust standard errors clustered at the school level in parentheses. + p<0.10; * p<0.05; ** p<0.01; *** p<0.001.

Interview and Focus Group Guides

Administrator Interview

Introduction

First, I would like to begin by asking some general questions about you and your school.

[ASK ASSISTANT PRINCIPAL]

1) Could you briefly describe your role at this school?

2) What are the major priorities for the school?

3) We know that upholding high expectations for students is always a focus, how are those expectations maintained and enforced here?

Knowledge of the Innovation to Increase Student Ownership

4) In your own words, please describe what it means for students to take ownership and responsibility of their own learning.

5) What's your knowledge of the specific practices or activities that teachers and other school staff have adopted that support with the program's goal of having students take ownership and responsibility for their own learning?

6) Tell me about your understanding of the practices that have been implemented to increase students' growth mindsets?

[PROBE] Have you seen any growth mindset practices when you're doing walkthroughs or classroom observations? If so, please describe.

7) Tell me about your understanding of the practices that have been implemented related to a common problem solving process?

[PROBE] Have you seen any problem solving practices when you're doing walkthroughs or classroom observations?

8) Are there elements of the innovation that you think have been particularly useful?

9) Do you think that the innovation to increase student ownership and responsibility is aligned with your vision for the school? Why or why not?

10) To what extent are students responsive to these activities?

Perspectives on Implementation

11) Some teachers were introduced to this innovation during professional development at the

beginning of the year, how do you think that went?

12) How do you think the implementation team is working as a team?

13) Were you involved in formulating the team's implementation plan? Did you have an opportunity to give input?

14) Are there ways in which you've shifted school resources such as money, time, or materials to support implementation? Can you describe these shifts in resources?

15) We know districts often have multiple priorities and new programs, how do you see the innovation around increasing student ownerships and responsibility interacting these priorities?

16) Everybody knows in this world of education, initiatives and programs come and go. Do you think this innovation on student ownership and responsibility has staying power? Why or why not?

I just have a couple final questions before we conclude:

17) From your perspective, what needs to happen for successful implementation. What's not currently in place that needs to be addressed?

18) Our goal is to understand how this innovation is working on your schools, is there anything that we haven't talked about that's important for us to know.

SOAR Team Member Interview

Introduction

First, I would like to begin by asking some general questions about you and your school.

1) What is your role in the work as part of the innovation to increase student ownership and responsibility?

2) Do you think that the innovation to increase student ownership and responsibility is aligned with the school's vision? Why or why not?

3) What are some of the practices and structures that your school is undertaking to increase student ownership? What are the goals of each activity?

> [PROBE] Have you done anything related to problem-solving? Growth mindset? (IF NOT MENTIONED)

4) Do you believe the practices that your school is undertaking will be effective in reaching the goals of the innovation? Why or why not?

5) What have the challenges been in implementing these practices?

Perspectives on Implementation

6) How successful do you believe the initial PD was in preparing your staff to implement these practices?

[PROBE] Can you tell me about some other specific supports that have been put in place since then to support teachers with implementation?

[PROBE] Do teachers and other school staff know what they are expected to do?

7) Do you feel like you have received adequate training and support to lead implementation at your school?

Perspectives of your work on the SIDT

We now have a series of questions related to your work on the SIDT.

8) Tell me how the SIDT works as a team.

9) We are also curious about how the SIDT work on PDSA (Plan, Do, Study, Act) is going. Can you tell me a bit about that?

[PROBE] Have you learned anything from the data collection part of PDSA?

[PROBE] Have you made any changes to the innovation?

[PROBE] Is there anything you would change about the PDSA process at this point?

I just have one final question before we conclude:

10) How are teachers responding to the new innovation?

11) Our goal is to understand how this innovation is working on your schools, is there anything that we haven't talked about that's important for us to know.

Teacher Interview Guide

Introduction

First, I would like to begin by asking some general questions about you and your school.

1. Could you briefly describe your position at this school?

2. In your own words, please describe what it means for students to take ownership and responsibility of their own learning?

3. Can you describe the goals of the innovation designed to increase student ownership and responsibility?

Perspectives on Implementation

Next, I would like to get your feedback on some of the training and resources you may have been given to help with implementation.

4) Did you attend any professional development or training for this innovation?

> [PROBE] How has professional development been helpful in supporting you to implement these practices?
>
> [PROBE] What other resources or additional training do you need to implement the practices of the innovation?

5) What practices have you been asked to implement in your classroom that support the program's goal of having students take ownership and responsibility for their own learning?

> [PROBE] Has your school made extra time in the schedule for teachers or students to engage in these practices?
>
> [PROBE] If you could give feedback to the implementation team, what do you need to supports your implementation efforts?

Understanding of the Innovation

6) Can you describe your understanding of a growth mindset?

> [PROBE] What student behaviors indicate a growth mindset?
>
> [PROBE] What do you do to try and develop growth mindsets in your students?
>
> [PROBE] How do you think developing growth mindsets is related to helping students take greater ownership and responsibility for their learning?

7) How would you describe a fixed mindset?

> [PROBE] What student behaviors indicate a fixed mindsets?
>
> [PROBE] What's your strategy for addressing students when they exhibit fixed mindsets?

8) Is there a school based problem-solving process you were asked to teach your students?

[PROBE] Have you used this technique with your students?

[PROBE] How do you think developing problem solving skills is related to helping students take greater ownership and responsibility for their learning?

Student Behaviors related to the innovation to promote student ownership and responsibility

I know want to find out a bit about how your students are responding to the practices meant to support their development of greater ownership over their own learning.

9) When students do come in for extra help, what do you think their motivation is?

10) We know teachers spend a lot of time encouraging students. When you give positive feedback, what types of things do you highlight?

11) How do you approach allowing students to revise their work?

12) How do your students respond when you give them a really difficult problem?

[PROBE] How do you support students when facing a really challenging problem?

I just have a few final questions before we conclude:

13) How are students responding to this innovation?

14) Overall, how would you characterize your support for the innovation?

15) Our goal is to understand how this innovation is working on your schools, is there anything that we haven't talked about that's important for us to know?

Student Focus Group Interview Guide

Introduction

We wanted to start this group with a couple scenarios. We want to think about what you would *and* should do in each of these scenarios. We also hope to find out about how adults in your school may help you in these situations.

1. Imagine that you have missed a class for the last three days and you have a test coming up next week. What would you do?

   [PROBE] What should you do?

   [PROBE] What would the adults in your school do to help you?

2. Imagine that you are having a hard time with some new material in a class. You think this material is really difficult and aren't sure whether you understand it. What would you do?

   [PROBE] What should you do?

   [PROBE] What would the adults in your school do to help you?

We have a few questions about an innovation in your school that encourages *students to take ownership and responsibility of their learning*.

3. Have you heard about growth and fixed mindsets in any of your classes?
   [PROBE] What does that mean to you? Can you give examples?

4. Have you heard about how the brain learns in any of your classes?

   [PROBE] What did you take away from learning about the brain?

5. Some people say that you are born smart. Other people say that if you study hard you can get smarter. What do you think about that?

6. Have you learned how to use a problem solving process in any of your classes?

   [PROBE] Can you tell us about that?

   [PROBE] Can you give an example of a time when you have used this problem solving process?

We're really interested in your experience with _____ [teacher from the observed class]. We are hoping to learn more about practices that _____ [teacher's name] has used to try and develop student responsibility among students in your class.

7. What does your teacher do to help you to take greater ownership or responsibility for your learning?
   [PROBE IF NEEDED] Examples: Does your teacher help you organize your assignments, keep track of your homework, etc?

For Observed Behaviors:	**For Unobserved Behaviors:**
9A. Growth Mindsets: I noticed that your teacher talked about mindsets/how brain learns in class. [Give example.] Why do you think your teacher is doing that? [PROBE] Is it typical for your teacher to talk about growth mindsets?	**9B. Growth Mindsets:** Has your teacher ever talked about mindsets or how the brain learns in class? [Brief explanation if necessary] [IF SO] Is it typical for your teacher to talk about growth mindsets?

	[IF SO] Why do you think your teacher is doing that?
10A. Problem-Solving: I noticed that your teacher asked you to use a problem-solving process in class today. [Give example.] Why do you think your teacher is doing that? [PROBE] How often does your class use this problem-solving process?	**10B. Problem-Solving:** Has teacher ever taught you about a problem-solving process? [IF SO] Is it typical for your class to use this problem-solving process? [IF SO] Why do you think your teacher is doing that?
11A. Mistakes: I noticed that your teacher talked about mistakes during class today. [Give example.] Why do you think your teacher is doing that? [PROBE] Is it typical for your teacher to talk about mistakes?	**11B. Mistakes:** Has your teacher ever talked about how you can learn from mistakes? [IF SO] Is it typical for your teacher to talk about learning from mistakes? [IF SO] Why do you think your teacher is doing that?
12A. Praise: I noticed that your teacher used a lot of specific language when giving feedback to students about their work. [Give example.] Why do you think your teacher is doing that? [PROBE] Is it typical for your teacher to give positive feedback like that?	**12B. Praise:** Tell me about ways that your teacher gives you feedback.
13A. Questions: I noticed that your teacher asked a lot of open-ended questions for students to answer. Why do you think your teacher is doing that? [PROBE] How often does your teacher ask those types of questions? [PROBE] **If wait time observed:** I noticed that your teacher pauses after asking a question. Why do you think your teacher is doing that?	**13B. Questions**: Tell me about the types of questions that your teacher asks in class. For example, does your teacher ask you yes/no questions, give you a list of possible answers, or ask more open-ended questions? [PROBE FOR EACH] Why do you think your teacher asks those types of questions? [PROBE] Does your teacher ever pause after asking a question before getting student answers? If so, why?

We have a couple more questions about the work you do at your school.

14. Do your teachers give you an opportunity to revise your work?

    [IF SO] Do you revise your work when given the opportunity?

    [IF SO] Why do you revise your work?

15. What else could adults in your school do to help you take more responsibility for your learning?

Support Staff Focus Group Guide

First, I would like to find out some general information about you and your school. Let's go around and introduce ourselves.

1. Describe the school culture.

    [PROBE] What is the culture among students?

    [PROBE] What is the culture among teachers and administrators?

2. Would you say that teachers, administrators, and other professional staff have a shared vision for this school?

3. What are some current priorities and/or major initiatives happening in your school?
4. What experience do you have implementing new programs or initiatives at your school?

    [PROBE] We know that schools are implementing new practices or programs all the time. How successful is your school in implementing new ideas, practices, or programs?

Now, we have some questions about the your understanding of the goals of the innovation that addresses how the school can help increase student responsibility and ownership over their learning.

5. What do you know about the innovation designed to increase student ownership and responsibility?
6. How did you learn about this innovation?

7. Part of this work is developing a culture of growth mindsets in this school. What do you think about that?

    [PROBE] Are there any aspects of your own practice that have been influenced by the growth mindset work?

    [PROBE] Are there any other ways you support the development of a growth mindsets culture in your school?

8. Another part of this work is teaching students about problem-solving. What do you think about that?

    [PROBE] Are there any aspects of your own practice that have been influenced by the problem-solving work?

    [PROBE] Are there any other ways you support the development of a problem solving culture in your school?

9. Do you believe this focus on student ownership and responsibility will help students succeed?

10. Earlier you described the current priorities and initiatives in your school. How does this innovation around student responsibility and ownership fit into the bigger picture at your school?

11. Earlier you talked about your prior experience implementing new programs/initiatives. In thinking about this focus on student ownership and responsibility, what is not currently being done that you think needs to happen for it to be successful?

I just have one final question before we conclude:

10) Our goal is to understand how this innovation is working on your schools, is there anything that we haven't talked about that's important for us to know.