Chapter 6

Supporting Conditions for Collaborative Data-Intensive Improvement

As the room filled up, those new to the gathering saw a modern conference room that served as both a workspace and a cafeteria. After nearly two days of work, three teams, each made up of researchers and practitioners from a Charter Management Organization (CMO) based in the Bay Area, scrambled to put the finishing touches on their presentations. This day was preceded by nearly two and half years of work. The path connecting the early ideas of a partnership between researchers and practitioners around the use of data-intensive research methods was not without twists and turns. But at the heart of the partnership, from very early on, was a sense of trust and mutual benefit that helped members of the partnership lean into the twists and turns that accompanied using new forms of data to improve teaching and learning in classrooms throughout the CMO.

The teams preparing their final presentations were formed on the first day of the two-day event, which the partners called a "data sprint." The sprint was an opportunity for the partners to come together to jointly analyze data from digital learning environments and administrative data systems. Three teams were organized around three driving questions: What is the relationship between external and internal measures of student learning and achievement? What is the relationship between students' activity in a digital learning environment and external measures of student achievement? How many distinct learning behaviors can be measured using digital learning environment data? The sprint was structured as a two-day event in order to allow sufficient time to analyze digital learning system data and co-develop follow-up actions, such as new instructional routines that teachers could later implement in their classrooms. The first of three teams began its presentation on the degree to which the CMO's own assessments correlated with external measures of students' college readiness, such as standardized test scores. Nearly every component of the presentation, as well as the accompanying analysis, had been a collaborative undertaking, from the merging of data files and transformation of variables to the interpretation of results.

At the end of the presentation, the first team laid out the next steps they would take based on what they had learned from their analyses. Using simple scatterplots with carefully chosen reference lines, they had identified a group of students who were behind on both internal and external measures of college readiness. The second and third teams, similarly, identified previously unnoticed patterns in students' use of the CMO's digital learning environment. For example, they found that students who tended to follow up a poor performance on an assessment by accessing an available learning resource did better in the course overall than students who tended to follow up a poor performance by retrying the assessment. These presentations would launch months of work examining the patterns identified during the sprint across multiple grades, content areas, and over wider timescales.

Getting to a point where researchers and practitioners were able to jointly engage in data-intensive research did not happen overnight. New knowledge, skills, and dispositions needed to be developed by both researchers and practitioners. In this chapter, we describe our approach for engaging in data-intensive research–practice partnerships (RPPs). Across multiple partnerships, we have engaged in two overarching activities: First, we collaborated with an educational organization around using data-intensive research methods to explore the organization's own data. Second, we explored ways of improving how we worked with each educational organization. These two activities served as the building blocks of a multi-year, multi-project research agenda aimed at developing a repeatable approach for collaborating with practitioners, what we have come to refer to as *collaborative data-intensive improvement* (CDI). We will return to the CMO described in the introduction of this chapter and describe how we set the foundation for the data sprint over multiple years of collaboratively analyzing the CMO's data and working to improve the partnership over time. Before returning to the CMO and introducing a second partnership, we briefly describe the origin of CDI through the lens of prior efforts to use data in schools, building on the overviews provided in the previous chapter.

Origins of Collaborative Data-Intensive Improvement

Working with practitioners to engage in data-intensive research is not a new idea. Groups such as the Youth Data Archive (YDA) at the John W. Gardner Center for Youth and Their Communities regularly work in close collaboration with practitioners to develop research questions, interpret analyses, and co-develop changes (Russell, Jackson, Krumm, & Frank, 2013). The "data for good" movement through organizations like DataKind and Bayes Impact has demonstrated the potential for bringing together data scientists and practitioners from non-profit and non-governmental organizations. And foundations like the Ann E. Casey Foundation have funded efforts to make use of integrated data systems from non-profit and governmental service providers to better understand issues facing youth and families. Thus, there are multiple examples of researchers, data scientists, and practitioners coming together to analyze and take productive action based on analyses of large volumes of data from, in particular, administrative data systems.

As we began thinking about how to use new sources of data with practitioners, we were drawn to the work of organizations like the Harvard Strategic Data Project, DataKind, and YDA (e.g., McLaughlin & London, 2013). Across such standout organizations, though, we saw two gaps that had yet to be addressed when we started thinking about data-intensive RPPs in 2014. First, most if not all partnerships were structured around administrative data systems—we wondered what could be gained by forming partnerships around data from digital learning environments, alone or in combination with data from other sources. Second, the work involved in actually engaging in collaborative data-intensive research had not been detailed within the existing literature—we set out to develop in-depth accounts of how to engage in collaborative data-intensive research to help subsequent partnerships based on our lessons learned. We addressed these two issues by launching a series of partnerships, consulting prior research where it was available, and reflecting on what worked and what did not. Within each partnership, we engaged in a style of inquiry referred to as design-research, which we describe later, and through a series of design-research cycles we set out to identify *supporting conditions* and *key phases* for engaging in CDI.

In education, design research maps closely onto the development of the learning sciences (Bransford, Brown, & Cocking, 2000). The key insight of this approach, going to back to the pioneering work of Ann Brown (1992) and Allan Collins (1992), is that evidence collected under tightly controlled conditions, such as laboratory settings typical of early psychological research, may not in fact generalize well to the day-to-day realities of schools. This insight, which has been wrestled with by many in the field of education, developed into a methodology where researchers use theory to develop interventions that are then tested in real learning environments. Through iterative refinement and collaboration with practitioners, design research involves creating a "humble" theory for why an intervention worked (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). Many learning scientists and methodologists have weighed in on the pros and cons of design research, and it is not an approach without critics (Kelly, 2004; Shavelson, Phillips, Towne, & Feuer, 2003). Yet for some, including us, it offers a way of approaching practical educational problems while working to build theory (Bell, 2004). The key unit of analysis in traditional design research is a learning environment, such as a classroom or after-school learning experience. Through working in real-world environments, the goal is not to isolate individual causes for why an intervention worked or not but rather to attend to, as best as possible, how an intervention affects and is affected by the myriad complexities within each setting (Barab & Squire, 2004). In our work, the key units of analysis are the partnerships themselves and how the work of the partnership translates into changes in the ways that practitioners interact with learners.

At the outset of our efforts to develop data-intensive RPPs, we simultaneously embraced and questioned the assumption that data from digital learning environments would be beneficial to understanding and improving learning environments. As we noted previously, up to this point in the development of the fields of educational data mining and learning analytics, there were only a handful of collaborations between researchers and practitioners using data from digital learning environments (e.g., Krumm, Waddington, Teasley, & Lonn, 2014). Examples of collaborations built around data residing in administrative data systems were more prevalent, and a number of these involved large urban school districts developing early warning systems with university partners (e.g., Allensworth & Easton, 2005; Balfanz, Herzog, & MacIver, 2007). The paucity of meaningful collaborations around digital learning system data was unfortunate given the pitched rhetoric around "big data in education," which carried the implication that data from digital learning environments was somehow a self-activating resource. The process for translating data into new actions that teachers would take in classrooms was either ignored or assumed to be trivial (Piety, 2013). While we embraced the potential value for newly emerging types of data, we questioned how much could truly be accomplished without having educational data scientists move closer to practice and practitioners move closer to the work of educational data science.

There is a long history of attempts to bridge research and practice in education, of which design research is but one example (Lagemann, 2000). And while it is easy to say that researchers should work more closely with practitioners, and vice versa, consistently doing so has proven to be challenging. To inform efforts at bringing researchers and practitioners together, we focused on the focal activity in which researchers and practitioners would be engaged, namely, planning for, collecting, and interpreting data, commonly referred to as data-driven decision making (DDDM; see Chapter 5). Explicit models for engaging in DDDM around standardized tests and benchmark assessments existed (e.g., Boudett, City, & Murnane, 2013; Bambrick-Santoyo, 2010). However, the evidence that decisions were informed by data and that these decisions had ultimately contributed to improved instruction was far from clear (Coburn & Turner, 2011; Hamilton et al., 2009). One reason for the limited number of clear cases may have had much to do with the ultimately weak theory of change associated with most DDDM models, which involves educators—alone—developing and deploying capabilities necessary for setting goals, determining causes for success and failure, implementing new approaches, monitoring effectiveness, and reflecting on the overall DDDM process (Penuel & Shepard, 2016). As many have pointed out, there are multiple potential points of failure in this theory, and many researchers who have studied how teachers actually worked with data found that data rarely "drove" decisions because data were often difficult to access and lagged too far behind the processes that practitioners could affect (Little, 2012). Moreover, the interpretations that teachers made of the data were affected by numerous tangential factors (Turner & Coburn, 2012). The question for us, which ran headlong into assumptions that "bigger" data would lead to "better" decisions, was as follows: How would large, complex datasets and less well understood machine and statistical learning techniques interact with these underlying dynamics of data use in schools?

An essential place to start in answering this question was to bring researchers and practitioners together in a partnership that drew on the respective skills and expertise of each. It is easy to assume that schools can and should take on this work by themselves or that data scientists can overcome the challenges of data use in schools absent knowledge of schools or the active participation of practitioners, but the track record for these assumptions is not great. To understand how complex datasets and new analytical techniques could be used in schools, we identified opportunities for improvement within the existing literature and launched several partnerships organized around translating data from digital learning environments into concrete changes; and in reflecting on the multiple partnerships in which we were working, we set out to develop an overall approach for engaging in CDI.

Two Data-Intensive Research–Practice Partnerships

We opened this chapter with a description of an experience from one of two early CDI partnerships that we launched. The CMO that participated in the data sprint and the thousands of hours of collaborative work that preceded it was Summit Public Schools. Summit is a CMO that operates schools in the Bay Area and in the state of Washington. Summit has been recognized as an innovative CMO based in no small part on its deep integration of technology and focus on providing personalized learning experiences to all its students (Murphy et al., 2014). The second partnership that we describe later involves our work with the Carnegie Foundation for the Advancement of Teaching and the Carnegie Math Pathways. As described in Chapter 5, the Carnegie Foundation for the Advancement of Teaching has become the central organization for advancing the use of improvement science in solving long-standing educational problems and inequities. As part of their field-building efforts, Carnegie launched the Carnegie Math Pathways, which is a unique network of 2- and 4-year colleges and universities focused on solving the developmental math crisis in the United States (Bryk, Gomez, & Grunow, 2010).

Summit Public Schools

In 2003, Summit began as a single high school, Summit Preparatory Charter High School in Redwood City, California. Since then, Summit has grown to 11 schools and a national program referred to as *Summit Learning*. Core to the Summit model of teaching and learning is a focus on personalization and strong relationships between students and teachers combined with giving all students a rigorous, college preparatory curriculum. A typical day at a Summit school is broken up into 90-minute blocks during which students engage in project-based learning in core subject areas. Project-based learning is an instructional method where students gain knowledge, skills, and dispositions through authentic, engaging, and complex problems (Larmer, Mergendoller, & Boss, 2015). In addition to project-based learning blocks, students engage in "personalized learning time" and weekly mentoring sessions. Personalized learning time offers students an opportunity to work on core academic content at their own pace, and mentoring sessions are times when students work one-on-one with a teacher who advocates for them and helps them develop self-directed learning skills.

Every Summit student is provided with a Google Chromebook and access to a customized learning management system (LMS) referred to as the *Summit Learning Platform*. Students use the platform in all of their courses and for a majority of their overall learning activities. For example, students use the platform during personalized learning time to access required assessments and teacher-curated resources, in the form of "playlists." Completing a playlist involves passing a 10-item content assessment that students can take as many times as they need and whenever they feel ready to take the assessment. Students interact with playlists during personalized learning time, which includes two 90-minute blocks throughout the week and for extended periods of time on Fridays, which is also when students interact with their individual mentors. Through the platform, Summit students also can work on elements of projects and can communicate with their teachers about their progress on specific elements of a project. From its founding in 2003, Summit has created a learn fast culture in which all elements of the student learning experience—from mentoring to the *Summit Learning Platform*—are continuously refined over time.

Our partnership with Summit began in the summer of 2014. Prior to the start of the partnership, Andy and a senior leader at Summit had participated in a national conference on the topic of personalized learning. At the conference, Andy and the then Chief Information Officer for Summit Public Schools discussed the multiple research projects in which they had been a part. Summit was an early adopter of multiple technologies and as a CMO they had participated in multiple studies on how they used technology to support teaching and learning. Summit staff lamented the fact that researchers' insights and efforts were often directed toward writing reports, as opposed to helping Summit staff grow and improve. This observation led the two to jot down the basic outline for a partnership organized around the idea of analyzing data from Summit's digital learning environments for the purpose of improving teaching and learning at Summit. A few follow-up phone calls later, the partnership between Andy and Summit had expanded to include Alex Bowers from Teachers College, Columbia University.

In the following sections, we provide a chronological description of our partnership with Summit Public Schools. Throughout the project, the participating researchers met regularly to reflect on the partnership and to clarify lessons learned about the *supporting conditions* for engaging in collaborative data-intensive research. The goal of identifying these conditions was to help subsequent partnerships launch and organize their own work. Design-research cycles, like those described previously, were organized around key events such as initial brainstorming meetings and subsequent meetings where members of the partnership would come together to jointly analyze and interpret data products. From the start of the project, we regularly experimented with how best to bring researchers and practitioners together, and we engaged in multiple data analyses geared toward helping Summit practitioners improve learning opportunities for students.

Setting the Foundation

The process for identifying the first round of research questions that would guide the partnership began with an initial, face-to-face meeting of leaders from Summit Public Schools and members of the research team. At the meeting, we engaged in a round of brainstorming activities where Summit leaders proposed topics and questions that the partnership could explore. Examples from this initial meeting included "identifying and measuring self-directed learning behaviors," "identifying the relationships between micro-momentary choices that students were making and their college-going trajectories," and "identifying specific ways to keep students on-track." Following an initial round of brainstorming, the technology and information teams from Summit outlined the data that were captured and stored by their various systems. This first meeting concluded with a preliminary set of topics for the partnership to pursue and a developing understanding of the data that could be used to explore each topic. Following the initial brainstorming meeting, the partnership blended Summit's research interests with the knowledge and expertise of the research team using the following process: (1) members of the research team wrote brief descriptions for how they could attempt to answer each question that was developed during the initial brainstorming session; (2) practitioners then reflected on the approaches proposed by the research team; and lastly, (3) the research team and Summit leaders came together to evaluate the potential impact and feasibility of answering each question.

The first question that the partnership collaborated on involved understanding patterns in students' attempting and completing content assessments. Summit's 10-item content assessments are quizzes that students are required to complete at the end of each playlist. Different courses require different numbers of content assessments to be completed. A distinctive feature of playlists is that students are given both the freedom to work on whatever playlist they choose and discretion in how they navigate each playlist. A core element of Summit's learning model is that students are provided with opportunities to grow and demonstrate "habits of success," such as self-direction, curiosity, and civic identity. Using data from the 2013–14 academic school year from the LMS used by Summit at the time, we began exploring patterns of how students took and passed content assessments, not only to answer the focused research questions but also as a concrete way to measure students' self-directed learning behaviors. Self-directed learning is closely related to self-regulated learning (e.g., Pintrich, 2004), which refers to the ways learners actively regulate their own cognition, motivation, behaviors, and elements of their environment in order to achieve a goal.

The project officially kicked off in September 2014, and we completed our first joint data interpretation meeting at the end of October. The speed with which the research team was able to complete these first analyses of students' behavior in taking content assessments was important given the partnership's goal of doing research differently and shortening the time between posing a research question and having a potential answer. What made exploring this question possible in such a short time frame was the fact that an entire prior academic year's worth of data had already been collected and stored within Summit's LMS and student information system.

To explore patterns in content assessment taking, the partnership used two initial strategies. First, we specified what content assessment taking should look like so that each student's actual content assessment taking could be compared against that normative standard. Second, the different patterns found in students' actual content assessment taking were examined in relation to outcomes that the partnership valued, such as course grades. For this first analysis, we examined student assessment taking patterns in relation to students' final grades in four core courses: Math, English, Science, and Humanities in ninth grade. For these four courses, we wrangled data from a database that tracked and stored students' content assessment taking in the LMS and from Summit's student information system, which contained students' course grades and standardized test performances.

We arrived at two major takeaways during an early fall meeting to review initial analyses. The first insight was based on the finding that the extent to which students struggled with content assessments varied for different assessments. One way in which we identified the degree to which students struggled was by examining students' scores the first time they attempted a content assessment. For some content assessments, the median student scored 4 out of 10 on his or her first attempt, while for other content assessments, the median student scored a 7 or 8. This and related findings generated questions around what was contributing to low scores (e.g., content assessment difficulty, students' prior knowledge related to the specific content being assessed, or the ways students prepared for taking the content assessment). The implication from these analyses was that effort should be directed, both by Summit practitioners following the meeting and by the research team in the form of new analyses, toward understanding factors contributing to students' struggle with particular assessments.

The second key insight from this same early fall meeting was that teachers should have easier access to cumulative and longitudinal data on students' attempts and completions of content assessments on the Summit platform. Up to this point in the history of the platform, teachers lacked basic information about what students were doing in the system. They could not answer questions such as "How many days were students taking to complete a playlist?" or "What resources were students using most often?" By aggregating a year's worth of data and structuring a conversation around how to interpret the data, the research team was able to demonstrate the potential benefit of providing longitudinal data directly to teachers through new data displays in the platform.

This initial cycle of inquiry would set expectations for the many partnership-driven analyses to follow. Direct engagement between researchers and practitioners helped improve researchers' understanding of the data they had analyzed and provided practitioners with an opportunity to see how learning at their school was playing out at scale and over time. Both kinds of insights helped the partnership brainstorm potential changes to the content assessments themselves and to how teachers worked with students to prepare for content assessments.

After this first cycle of inquiry, the research team added another dataset into the mix—students' use of the specific learning resources in their playlists. Using this additional dataset, the research team was able to examine relationships among students' standardized test performances, their use of playlist resources, and their content assessment taking and course grades. Perhaps unsurprisingly, an early finding was that students with lower incoming standardized test scores were attempting math content assessments more frequently. While lower incoming content knowledge may explain the need for more attempts in order to demonstrate mastery on an assessment, we also observed that students with higher standardized test scores were using the system in different ways than their peers with lower incoming scores. For example, higher-scoring students were using more unique learning resources and looking at those resources prior to taking a content assessment rather than afterward.

Building on Lessons Learned

The second deep-dive meeting between researchers and Summit leaders was held in winter of 2014 with the goal of discussing the analytic findings regarding students' learning resource use, content assessment taking, and course performance. This meeting prompted the partnership to think about how the findings could be communicated directly to Summit teachers. Up to this point, the researchers had been working most directly with the CMO leaders, and the leaders took responsibility for communicating findings and negotiating potential changes with teachers. The partnership targeted an upcoming all-CMO professional development meeting as an opportunity for teachers to learn more about the data analysis findings and to interact directly with the visualizations and other data products that represented those findings. The partnership collectively developed a strategy for using findings from the fall and winter meetings to create datasets that could be integrated into Summit's own data management and visualization tools. The hope was that by having the researchers take care of data wrangling and giving Summit teachers the opportunity to work with the organized datasets using tools they were already familiar with, a large number of teachers could engage with these data in an in-depth way. To help teachers navigate the new, unfamiliar data elements within their familiar systems, the research team briefly presented a description of the meaning of each data element and demonstrated how teachers could use a flowchart developed by the Summit information team. The flowchart was intended to help teachers identify whether or not a playlist was ripe for revision based on how many times students attempted its content assessment and the ways in which students used the playlist's learning resources.

During the first year of the partnership, Summit practitioners and the research team engaged in three cycles of inquiry into Summit's own data based on Summit's research questions. Across multiple meetings, the partnership experimented with how to surface practitioners' research questions, how to present findings, and how to translate findings into follow-up actions. As the partnership moved into the second year, members made a concerted effort at a two-day meeting to reflect on and highlight lessons learned that would inform the second year of working together.

In reflecting on the first year, the partnership members affirmed the importance of having Summit lead the question-generation process and of having researchers support that process by reflecting on questions and the potential impacts and feasibility of addressing them. It was apparent that the speed with which researchers had answered the first question posed by Summit leaders helped build trust with Summit. Another key reflection was the way in which the partnership came to value and take seriously the fact that opportunities to meet and discuss data analyses were *learning events* as opposed to presentations where researchers would present and defend their analysis. Instead, meetings were structured as collaborative opportunities for researchers and practitioners to learn from one another. Concretely, researchers made intentional efforts to not just present findings but to make as explicit as possible the thinking that went into each analysis. For example, we included data products that were built using sample or fake data to help practitioners understand the logic behind an analysis before presenting that product using their own data. Lastly, we organized a specific kind of learning event where researchers provided training to Summit's information team on how to use the R software.

The second year of the project got under way with a return to the original 10 research questions that were generated at the partnership's initial brainstorming meeting. From the original list the partnership selected two questions to focus on: (1) How do Summit's internal metrics relate to external benchmarks for college readiness? (2) Can we characterize what students do in the platform as successful or unsuccessful? To explore how Summit's internal metrics related to external benchmarks, we analyzed relationships among students' course grades and multiple standardized test scores, using data from multiple grade levels and subject areas. We then organized a broader meeting of teachers to explore the degree to which measures collected by Summit correlated with college-readiness indicators from external organizations. Across multiple meetings, we worked hard to help practitioners understand how to interpret specific relationships between internal and external metrics. Alex Bowers, who had recently done an in-depth analysis of the relationship between grades and standardized test scores, directly supported Summit staff during this period as they interpreted the developing findings.

Following these meetings, the researchers in the partnership conducted fewer analyses of the relationships between internal grades and external test scores because Summit staff were easily managing and analyzing these data. For these types of analyses, our roles shifted to helping interpret analyses that were initiated and carried out by Summit staff.

As researchers' involvement in conducting these correlational analyses declined, we began work on the second research question concerning successful and unsuccessful behavioral patterns within the Summit platform.

While data related to student outcomes, such as course grades and standardized test scores, had not required much, if any, feature engineering, data from the platform did. Data from the platform contained millions of observations and required knowledge of learning theory as well as certain technical skills in order to turn those observations into meaningful features that could then be compared against student outcomes. Over time, the pattern of researchers' activity became clearer: When there was little need to engineer features, researchers supported school staff who conducted analyses themselves in thinking about how to interpret findings; when a lot of feature engineering was required, researchers did more of the analytic work as well as supporting data interpretation.

Characterizing successful and unsuccessful student behavior patterns using platform data required significant feature engineering. Starting with data from the then current school year, we used three general approaches to look for patterns. The first approach built off of our analyses the prior year and entailed simply summarizing how each student used the platform in terms of the number of resources used, unique resources used, resources used before taking the first content assessment, number of content assessment attempts, and similar measures of the quantity of various types of activity. The second approach involved identifying strings of events. At a general level, each playlist comprises different types of resources and assessments. We coded each digital learning event as either a resource (R) or as an assessment that was either passed (P) or failed (F). Thus, each playlist that a student worked on could be represented as a string of letters (e.g., RRFRRFP). When combined with the quantitative metrics (e.g., number of unique resources used on a playlist), the event strings characterized the many possible ways that students used the learning platform. The third analytic approach we used involved quantitatively characterizing movements from one event to another. We created metrics, using conditional probabilities, that quantified how likely it is that a student would move from a certain kind of event to another kind of event. For example, if the student has failed a content assessment, how likely is it that the student will go immediately to another assessment and fail it? How likely is it that the student will go from the failed assessment to examining a learning resource? Across each pair of event categories, we identified students' most likely transitions, i.e., their most likely next step.

We presented these different ways of characterizing students' patterns in a spring 2016 meeting with Summit leaders and teachers. Throughout the summer of 2016, we used known grade and achievement score outcomes from the 2015–16 school year, and worked to identify more and less successful learning behavior patterns. Naturally, "it depends" was a common phrase in our discussions. For example, students who came to a playlist with a high level of domain-relevant knowledge tended to use few resources and ultimately needed fewer attempts to pass the required content assessment. To the question "Should students be using more learning resources?" *It depends*. Across multiple analyses, we developed evidence for a variety of patterns that members of the partnership took with them into the data sprint described in the opening of this chapter. In bringing researchers and practitioners together for a concentrated amount of time to explore new questions, we attempted to set up specific tests of change that could be enacted following the sprint—thus moving the partnership toward a better understanding of more or less successful learning behaviors.

Carnegie Foundation for Advancement of Teaching

During the same summer that our partnership with Summit began, Andy and colleagues started working with researchers and staff at the Carnegie Foundation for the Advancement of Teaching. At the time, Carnegie was well into launching and supporting the Carnegie Math Pathways, which is a national effort focused on improving developmental, or remedial, math courses in 2- and 4-year colleges throughout the United States. In some colleges, students can be required to take a developmental mathematics course if they have been identified as not ready for college-level mathematics. These courses can be a significant barrier to college completion; only a small proportion of the students who are required to take these courses pass them and go on to earn the college-level math credits required for many degrees. One study of 57 community colleges found that 80 percent of the students assigned to a sequence of developmental math courses did not successfully complete a transfer-level (i.e., credit-bearing) math course within three years (Bailey, Jeong, & Cho, 2010). The number of lives affected by the developmental math crisis in the United States is staggering, so Carnegie brought together experts in mathematics education with college teams who all wanted to tackle this problem using a new and promising set of approaches referred to as improvement science.

With Carnegie as the hub, they formed a networked improvement community (NIC), which as introduced in the previous chapter is a type of scientific community that is organized around a common aim, guided by a common understanding of the problem it is trying to solve, disciplined by the use of improvement science tools, and deliberately structured to share knowledge across those participating in the network (Bryk, Gomez, Grunow, & LeMahieu, 2015). Members of the Carnegie Math Pathways

NIC designed two different course sequences, or *pathways*, representing alternative, intensified approaches to fulfilling developmental math requirements and earning college credit in either statistics or quantitative reasoning, referred to as Statway and Quantway, respectively.

From the beginning, the Carnegie Math Pathways NIC has been organized around the aim of increasing the percent of students—from 5 to 50—who achieve college math credit within one year of continuous enrollment as compared to other developmental math offerings. Multiple studies demonstrate how this aim has been met and exceeded by colleges participating in the Carnegie Math Pathways (e.g., Van Campen, Sowers, & Strother, 2013; Yamada, 2017; Yamada, Bohannon, & Grunow, 2016; Yamada & Bryk, 2016). As the hub of the NIC, Carnegie worked with various researchers and practitioners to identify key drivers that were seen as necessary for achieving their aim (see Bryk et al., 2015, p. 75). Among these drivers, Carnegie singled out various "noncognitive" factors that have been shown to affect student success (see Duckworth & Yeager, 2015). These factors can include but are not limited to students' beliefs about their ability to learn math, their sense of belonging in school, their perceptions of value for learning math, and the ways in which they set goals, monitor progress toward goals, and reflect on what worked and what did not (i.e., self-regulation skills [Zimmerman, 2002]). These factors coalesced around the idea of improving students' academic tenacity and use of effective learning strategies, what would be referred to throughout the NIC as "productive persistence."

A key instructional resource in both pathways was the use of online learning systems. From the partnership's earliest conversations, Carnegie wanted to explore how the online learning systems were being used and the degree to which data from these systems could be used to measure and support students' productive persistence in Pathways classrooms. Thus, the partnership with Carnegie was chartered as an opportunity to leverage data that was collected by the online learning systems, whereby Carnegie led the coordination with faculty teaching at 2- and 4-year colleges to interpret data products and co-develop change ideas.

Getting Up to Speed

In the summer of 2014, we began working with data from the 2013–14 academic year. The partnership made the early decision to focus solely on analyzing data from Statway based on the ease of extracting data from the online system, which was built on the Online Learning Initiative (OLI) platform. The online system captured each page viewed, when it was viewed, each practice item that was attempted, when it was attempted, and whether an item was answered correctly or not. Along with page views and practice items, the online system also captured time- and item-level data from assessments, referred to as “Checkpoints.” Checkpoints come at the end of “topics” and “modules,” which organized Statway content into meaningful chunks.

For this year of Statway, there were approximately 1,600 students enrolled in over thirty 2- and 4-year colleges. Across reading, practice, and assessment activities, these students generated more than 7,300,000 rows of data. Wrangling and exploring these data involved working closely with the technology team at Carnegie as well as multiple researchers who themselves had spent time wrangling and exploring multiple datasets prior to the start of the partnership. The partnership was able to jump into analyses quickly because of the prior work that had been done by Carnegie researchers. Similar to our later work with Summit, it became important to find ways of adding value as opposed to duplicating capability. As we moved into the fall of 2014, we identified several data wrangling opportunities that could open up new levels of analysis related to the online learning system data, such as units of time (e.g., sessions and days), learning activities, and curricular organizers (e.g., topics and units) that could be used to aggregate the events that students logged, whereby these different levels of analysis could open up new opportunities to measure students’ productive persistence behaviors.

In the fall of 2014, we started to explore variation in how the online system was used across individual Pathways courses. These analyses were intended to quantify the ways in which individual instructors were using the online system at the scale of the entire network. If valid and reliable metrics could be collected on how instructors were using the online system, these measures could be used to help Carnegie staff coach faculty around best practices, understand differences in course outcomes, and provide a measure of the course context on which to better understand students’ productive persistence behaviors. For this initial analysis, we examined when students completed end-of-module Checkpoints. Using the dates that students within a course completed an end-of-module Checkpoint, we created multiple visualizations that represented both within-course variation (i.e., how students within the same course are different from one another) as well as the between-course variation (i.e., how courses are different from one another). The partnership identified that courses where most students followed the intended order of modules had higher proportions of students earning a C or higher than courses where students did not use the online system at all or where students completed end-of-module Checkpoints following a variety of different orders. Figure 6.1 demonstrates one way we visualized the dates on which students completed end-of-module Checkpoints, denoted “CP” in the figure. Each Statway course section is a row, and each time a student submitted a Checkpoint for a given module is represented by a shape. The within- and between-course variation captured

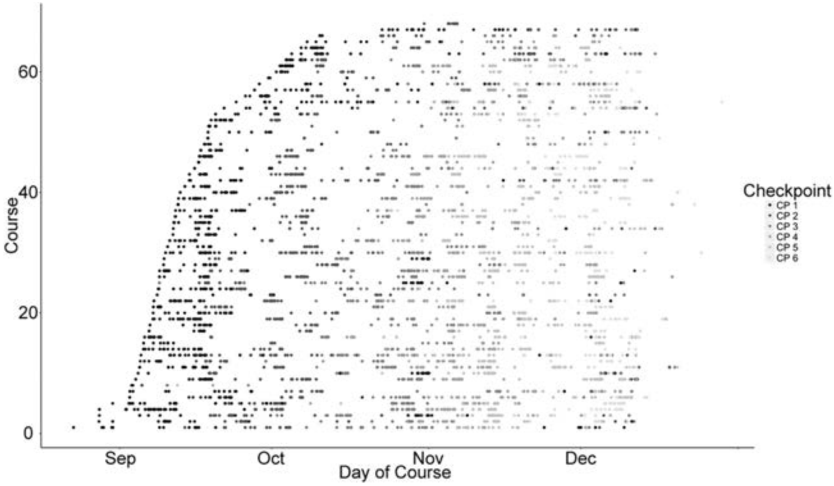

Figure 6.1 Statway End-of-Module Completion

in Figure 6.1 provided a rationale for Carnegie researchers to reach out to faculty throughout the NIC to better understand how and why they were using the system in the ways they were.

We followed up on the course level of analyses by exploring students' use of the online system by focusing on the "session" as the level of analysis. A session was defined by the online environment as the time between logging into the system and logging out, or being timed out, of the system. Just as we looked at within- and between-course patterns for these analyses, we explored within- and between-session patterns for students. In one analysis, we coded each session that a student engaged in as a string of events, which was similar to our approach with Summit's playlist events. For example, we coded page views, practice activities, and Checkpoints as V, P, and C, respectively. For example, a student could log the following strings for two separate sessions: "VVPVVPPC" and "CC." In the first example, the student began the session with a page view, engaged in both page views and practice activities during the middle of the session, and ended the session with a Checkpoint. The second example illustrates a student logging two Checkpoint events in a row. This approach helped in seeing the different ways in which students used the system and in generating features that were used in various unsupervised and supervised learning models. Features included distinctive types of sessions, such as assessment-only sessions (all "Cs"), as well as more robust sessions where students logged Vs, Ps, and Cs within the same session. Along with these different types of sessions, we observed different within-session behaviors, such as the number of activities a student engaged in prior to taking a Checkpoint. Krumm et al. (2016) describe how we later quantified these various within- and between-session features and modeled them in relation to test and grade outcomes. These analyses, as we observed in Chapter 3, were inferential in nature, which helped in providing evidence for the importance of potentially intervening on these behaviors over time. Armed with an understanding of how different Statway courses varied in their use of the online learning system and potential behavioral measures of productive persistence, the partnership began a series of design workshops with faculty participating in the Carnegie Math Pathways NIC.

Design Workshops

Design workshops were geared toward providing faculty with an opportunity to share their knowledge and expertise in interpreting data products, shaping subsequent analyses, and co-developing interventions that they could later implement. In the fall of 2015, we held our first design workshop with faculty mentors, who are a group of faculty who provide support and training to instructors throughout the NIC. In working with faculty mentors, we maintained our emphasis on experimenting with how best to organize meetings between researchers and practitioners around data. The initial workshop was organized around faculty mentors generating prototype data visualizations that they could use as part of their day-to-day teaching. For the partnership, the goal was to take faculty mentors' prototype ideas and translate them into visualizations that would later be deployed in the NIC's LMS.

The workshop was anchored in a brief presentation on a handful of measures from students' use of the system. The brief presentation was followed up with multiple individual and group prototyping activities. In collaboration with Carnegie researchers, we collected prototypes from each activity and later examined them in terms of common themes and the specific data elements they required in order to generate the visualization. We presented data products on the importance of students regularly logging into the LMS; engaging in reading and practice activities; completing Checkpoints; and reading, practicing, and assessing within the same session. Key themes that emerged from faculty mentors' prototypes were the need for more longitudinal representations of students' activity in the system and better alignments between *what* students did in the LMS with *how well* students did on Checkpoints. After analyzing the prototypes that mentors developed, we observed that prototypes often required data elements and relationships among data elements for which there was limited evidentiary support—having evidence to support the importance of the underlying student behavior represented in a visualization was a criterion that we set for the overall design process. Coming out of the first design workshop were a series of prototyped visualizations and new topics and questions to explore.

Building off of the topics from the first workshop, we engaged in a focused set of analyses, which led to a new set of data products that we presented to faculty at a second workshop in the summer of 2016. For this second workshop, we reduced the prototyping elements and increased the number of faculty who participated. Data products at this second workshop addressed the order of students' end-of-module Checkpoint completions (i.e., similar to the fall 2014 analysis), students persisting in the face of challenge (i.e., whether students return to a Checkpoint after a low score), the importance of engaging in reading as well as practice activities, and students completing both topic and end-of-module Checkpoints. As with the first workshop, prototyping activities surfaced multiple questions and follow-up topics. After the second workshop, we took stock of the evidence that was accruing around the importance of students attempting and persisting until successful on the Checkpoints within the first module and how this evidence also resonated with faculty.

Using the evidence related to completing Checkpoints, the partnership began two parallel tasks. First, we began preparing for a third design workshop where we focused on developing change ideas as opposed to prototyping data visualizations. Second, we started working with a large 2-year college in co-designing and testing strategies for helping students complete Checkpoints. Using analyses that had been presented at the second design workshop, Carnegie researchers co-designed three change ideas with participating faculty. The goal of these change ideas was to get students to complete 100 percent of their Checkpoints, i.e., their "homework," for the first two modules. Overall, we referred to this task as the "homework improvement sprint." Participating faculty members were later randomly assigned to test one or more of the change ideas in their classrooms using a planned experimentation approach (Moen, Nolen, & Provost, 2012). The co-developed changes included a work-block session prior to the start of face-to-face class (W), email reminders to students about completing their homework (E), and setting due dates within the LMS (D). Using data from the LMS, we detected positive effects for the percent of completed homework assignments as well as the timeliness of students' completion. Figure 6.2 illustrates the timeliness with which students completed each homework assignment (e.g., [1] CP 1.1.3). These boxplots illustrate how homework completion rates were more timely and less variable over time for faculty who tried out a change idea (i.e., faculty not marked "C" for control or "OTH" for other, non-participating faculty), and subsequent analysis revealed the overall benefit of the different work-block conditions (Meyer, Krumm, & Grunow, 2017).

Importantly, testing the previously mentioned co-developed change ideas was framed within the NIC and as part of the partnership as a learning opportunity, whereby evidence for the effectiveness of the individual

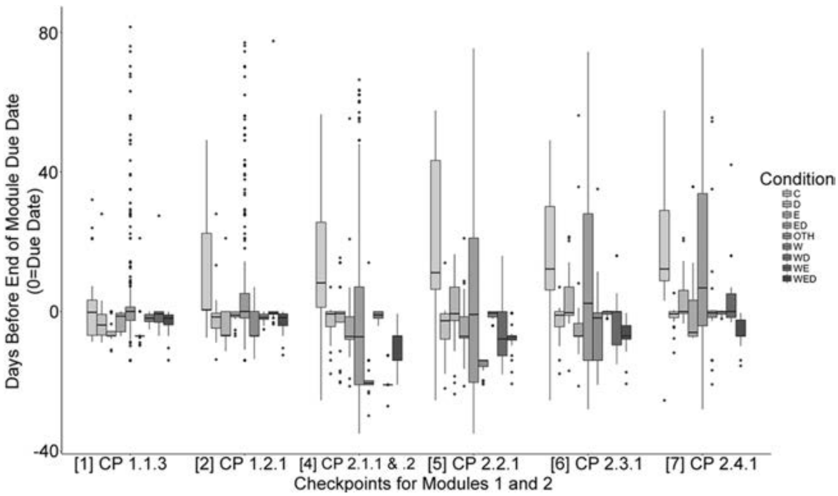

Figure 6.2 Homework Sprint Boxplots

changes would need to be built over time as the change ideas were tried by more faculty and replicated under a variety of conditions. The ability to try out and learn from testing a change idea in multiple, diverse contexts is a core element of NICs in general and the Carnegie Math Pathways NIC in particular. Based on multiple measures generated from LMS data (e.g., the timing of submissions in Figure 6.2), we developed a robust understanding of the many aspects of the changes faculty tested out, which helped in increasing confidence that the change ideas were promising despite the comparatively few faculty who participated.

Supporting Conditions for CDI

The two cases previously cited represent ways in which we have worked to take data-intensive research techniques to the frontlines of teaching and learning. Within each partnership, we tested particular ways of bringing researchers and practitioners together. For example, in Summit, we regularly worked with Summit's leadership teams and identified strategic opportunities to work with and learn from teachers. Similarly, in working with Carnegie, we regularly interacted with researchers and staff who supported the NIC and collectively identified the highest leverage ways of working directly with faculty and faculty mentors. Based on our experiences across these two partnerships, there are three factors that make these partnerships distinct from more traditional ways in which researchers work with practitioners: (1) research questions and topics were based on the needs of practitioners, (2) the primary audience for data products was the partnership, and (3) researchers and practitioners co-developed change ideas.

Across Summit and Carnegie, the questions and topics that were explored by the partnership were based almost entirely on the needs of practitioners or those supporting practitioners. In the case of Summit, these questions came directly from Summit leaders. For Carnegie, initial questions were based on Carnegie's interactions with faculty, and subsequent questions were generated at multiple design workshops. The primary audience for data products across both cases were members of the partnership who jointly interpreted data products, bringing with them their respective knowledge and expertise in identifying follow-on actions. Not only were researchers and practitioners jointly interpreting data products, they were also co-developing potential changes and testing them out in real learning environments. As we reflected on what was unique about these partnerships and in keeping with our design-based approach, we identified four *supporting conditions* that made each partnership successful.

In order for researchers and practitioners to come together to jointly develop data products and change ideas aimed at creating more effective learning environments, four conditions were in place:

1. The partnership between researchers and practitioners was based in **trust**.
2. An **explicit improvement method** organized multiple elements of the partnership's work.
3. **Learning events** provided opportunities for members of the partnership to collaborate and build knowledge.
4. **Common workflows** and accompanying tools supported data-intensive research, improvement activities, and project coordination.

In describing these four conditions, we do not intend to portray them as exhaustive, and it is our hope that other partnerships will use them as well as refine them over time. At a practical level, these conditions, at the very least, are intended to give future partnerships a head start in launching their own work.

Trust

Across the partnerships with Summit and Carnegie, trust was a key component. Some may think trust has little to do with data-intensive research. However, if the goal is to improve educational outcomes, the role of trust between researchers and practitioners is hard to overstate (Penuel & Gallagher, 2017). Bryk and Schneider (2002) define trust as having one's expectations validated in the actions of another. Trust is a multi-dimensional construct based in respect, personal regard, competence, and integrity. Exchanging data, developing data products, and testing out ideas all benefit when both researchers and practitioners trust one another in both word and action.

Respect is experienced in the ways individuals talk to and about one another; respect is also experienced as feeling heard by other members of the partnership. At the start of the Summit partnership, for example, respectful interactions were initiated early on by listening to and building off of practitioners' questions. And across both Summit and Carnegie, respectful interactions also played a role during meetings where the partnership jointly interpreted data products. As researchers in these partnerships, we strove to create meetings—ultimately framed as learning events—where every interpretation was valued and could provide insight into understanding a data product. Acting with integrity, while a seemingly general phrase, manifested in both partnerships as adhering to promises and deadlines. Acting with integrity further involved adhering to specified procedures for working with data, as well as in keeping data analyses and change ideas focused on improving local learning environments.

Bryk and Schneider (2002) further acknowledge the importance of personal regard as a foundational component of trust. Concretely, one way in which we as researchers demonstrated personal regard involved going above and beyond in our roles as researchers; we regularly participated in last-minute presentations and conducted analyses that were not a part of either partnership's formal question development processes. While personal regard can be seen as another person going out of his or her way to help another, competence is about fulfilling one's role within the partnership. For practitioners, competence can entail understanding and describing the various learning environments that the partnership will work to improve as well as accessing and sharing relevant data. For researchers, competence means being able to carry out multiple data analysis tasks as well as being able to organize meetings and events where members of the partnership work to jointly make sense of data products. Key competencies across the two partnerships described previously were data wrangling and feature engineering, which helped each partnership merge datasets and surface new patterns and insights within their data.

In many ways, "collaboration" and "partnership" are empty words until both researchers and practitioners begin validating their words through action. Importantly, trust is a two-way street: Practitioners need to demonstrate their commitment to a partnership through both time and engagement; researchers need to similarly demonstrate their commitment by adjusting their time and schedules to better align with practitioners'. For example, at the start of the partnership, Summit highly valued working at a faster pace than that of typical research projects, and we as researchers demonstrated our ability to work at this pace. Lastly, as we observed throughout multiple meetings, trust can play an important role in jointly interpreting data products and co-developing change ideas. Trust can help mediate potentially unflatteringly outcomes that are brought to light through an analysis and trust can provide a sense of safety in brainstorming potential implications from an analysis.

Explicit Improvement Method

As partnerships get started and begin to organize their work together, it can be useful to have a set of steps to follow and tools to use. Based on our work with Carnegie, we learned firsthand the important ways in which improvement science techniques can help in organizing partnership activities. In particular, strategies for understanding the problem that a partnership will work on as well as developing a theory for how to solve the problem are key steps in almost any improvement project. Across our work with both Summit and Carnegie, tools such as a causal systems analyses and driver diagrams (see Chapter 7) have all helped in shaping data-intensive analyses and co-design work. Following the development of change ideas, explicit improvement methods can be helpful in setting up iterative tests of change.

While there are many improvement methods to choose from, we have regularly made use of the *Model for Improvement* outlined by the Associates in Process Improvement, the Institute for Healthcare Improvement (see Langley et al., 2009), and the Carnegie Foundation for the Advancement of Teaching (Bryk et al., 2015). We have also used tools from the *clinical microsystems* approach developed at Dartmouth College (see Nelson, Batalden, & Godfrey, 2007). Based on our experiences, improvement science techniques play critical roles in shaping what happens before as well as after a data-intensive analysis. For example, in our work with both Summit and Carnegie, we used driver diagrams as a way of identifying key behaviors and outcomes to measure using data from digital learning environments (see Krumm et al., 2016). While improvement methods can help to shape data-intensive analyses, they are also useful in providing approaches for testing potential changes. Improvement routines, such as a Plan-Do-Study-Act (PDSA) cycle, help clarify hypotheses related to an intervention, measurement opportunities, and approaches for making sense of the test. In our work with Carnegie, both PDSA cycles and the planned experimentation methodology (Moen et al., 2012) were used to test change ideas (see Meyer et al., 2017). These approaches provided a common set of tools that members of the partnership could use in carrying out and learning from each test.

Learning Events

A recurring finding from the literature on instructional improvement is that most complex interventions require practitioners to develop new skills and abilities (Cobb & Jackson, 2012). Achieving many of the goals that we set out for each partnership required both researchers and practitioners to develop new skills and abilities. The primary location for this learning occurred in meetings where members of each partnership jointly interpreted data products, co-developed change ideas, or explicitly learned from one another in a more formal setting. At a general level, learning events were structured activities where members of a partnership could develop new understandings by engaging in joint work.

Not every meeting, however, involved collaboratively developing change ideas. For example, with Summit, we organized a formal workshop where we provided direct support on using the statistical software R. Even during meetings where most, if not all, of the meeting was dominated by researchers presenting analyses to partners, very early on we came to view these not as simple information transfers from one group to another but as opportunities to demonstrate to partners the ways in which we approached problems and thought about data. In our work with Summit, this most clearly manifested in a practice where we as researchers would create data products based on mocked-up data to first demonstrate the intuition behind an analysis.

Across both Summit and Carnegie, a key feature of the ways in which we worked with partnership members as well as teachers and faculty, respectively, was continuously playing around with the genre of what it meant to meet and learn from one another. With Summit, the clearest case of this was during the data sprint, whereby in an intensive two-day event our goal was to shorten the time as much as possible from when data were analyzed to the development of explicit change ideas. With Carnegie, we experimented with different approaches for working with faculty and faculty mentors through design workshops. Each one of these latter events had a clear instructional goal and was organized accordingly.

Common Workflows

A key component of data-intensive research involves analyzing, interpreting, and deriving implications from complex datasets. For these activities to take place, partnering organizations need to exchange data. In our partnerships with both Summit and Carnegie, we adopted a similar set of tools and routines for working with data from online learning and administrative data systems. First, data were queried from a database and uploaded to a password-protected, auditable, and role-based file transfer system. This system served as the central repository for raw data. Researchers were granted access to particular files; downloaded those files to password-protected and encrypted local computers; and engaged in data analysis using scriptable data analysis software. Eventually, both partnerships adopted the open-source language R and standardized many elements of the workflow using the same R packages. Data cleaning, wrangling, and analysis scripts were shared across researchers and practitioners, which created the opportunity for more reproducible analyses. Importantly, as both partnerships progressed, the added benefit of scripting every step in an analysis was that it created opportunities for practitioners to learn from worked examples and further develop their own data-intensive research skills (Gee, 2010).

Another key workflow across both partnerships involved sharing and storing data products. We experimented with collaborative file sharing services like Google Drive and Dropbox and learned over time the importance of having an intentional system in place for curating data products. In both partnerships, we produced hundreds of separate analyses—each with takeaways that informed a future action to varying degrees. Being able to revisit past work helped make partnership meetings more efficient. As we actively worked to develop better knowledge management approaches, we used improvement methods on ourselves and our own workflows. For example, our aim was to script 100 percent of a workflow and to be able to trace 100 percent of data products to a driving question. Opening up data transfer, sharing, and analysis to the tools and routines of improvement science helped us identify key opportunities for improvement and make data-intensive research activities more efficient and effective.

Conclusion

In this chapter, we described two cases of CDI and outlined four conditions that we viewed as helping to sustain each partnership over time and ultimately turn raw data from digital learning environments into new insights and change ideas. At the outset, we oriented CDI within the broader traditions of data-driven decision making, educational data mining, and learning analytics. At the intersection of these multiple traditions, we saw clear gaps in that few partnerships existed around using data from digital learning environments and few provided detailed depictions of how to engage in collaborative data-intensive research regardless of the data source. We follow elements of the cases described in this chapter into our discussion for the ways in which CDI projects can be organized and executed across five phases in the next chapter.

References

Allensworth, E., & Easton, J. (2005). *The on-track indicator as a predictor of high school graduation*. Chicago: Consortium on Chicago School Research.

Bailey, T., Jeong, D. W., & Cho, S.-W. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, *29*(2), 255–270. doi:10.1016/j.econedurev.2009.09.002

Balfanz, R., Herzog, L., & MacIver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223–235.

Bambrick-Santoyo, P. (2010). *Driven by data: A practical guide to improve instruction*. San Francisco, CA: Jossey-Bass.

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of the Learning Sciences*, *13*(1), 1–14.

Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist*, *39*(4), 243–253.

Boudett, K. P., City, E. A., & Murnane, R. J. (2013). *Data wise: Revised and expanded edition: A step-by-step guide to using assessment results to improve teaching and learning* (Revised and Expanded ed.). Cambridge, MA: Harvard Education Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences*, *2*(2), 141–178.

Bryk, A. S., Gomez, L. M., & Grunow, A. (2010). *Getting ideas into action: Building networked improvement communities in education*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching, Essay. Retrieved from www.carnegiefoundation.org/spotlight/webinar-bryk-gomez-building-networked-improvement-communities-in-education

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.

Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. New York, NY: Russell Sage.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.

Cobb, P., & Jackson, K. (2012). Analyzing educational policies: A learning design perspective. *Journal of the Learning Sciences*, *21*(4), 487–521.

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement*, *9*, 173–206.

Collins, A. (1992). Toward a design science of education. In E. Scanlon & T. O'Shea (Eds.), *New directions in educational technology* (pp. 15–22). Berlin: Springer.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters. *Educational Researcher*, *44*(4), 237–251.

Gee, J. P. (2010). *New digital media and learning as an emerging area and "worked examples" as oneway forward*. Cambridge, MA: MIT Press.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making*. (NCEE 2009–4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *The Journal of the Learning Sciences*, *13*(1), 113–128.
Krumm, A. E., Beattie, R., Takahashi, S., D'Angelo, C., Feng, M., & Cheng, B. (2016). Practical measurement and productive persistence: Strategies for using digital learning system data to drive improvement. *Journal of Learning Analytics*, *3*(2), 116–138.
Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). Using learning analytics to support academic advising in undergraduate engineering education. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 103–119). New York: Springer.
Lagemann, E. C. (2000). *An elusive science: The troubling history of education research*. Chicago: University of Chicago Press.
Langley, G. J., Moen, R., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance*. New York, NY: Jossey-Bass.
Larmer, J., Mergendoller, J., & Boss, S. (2015). *Setting the standard for project based learning: A proven approach to rigorous classroom instruction*. Alexandria, VA: Association for Supervision & Curriculum Development.
Little, J. W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education*, *118*(2), 143–166.
McLaughlin, M. W., & London, R. A. (2013). *From data to action: A community approach to improving youth outcomes*. Cambridge, MA: Harvard Education Press.
Meyer, A., Krumm, A. E., & Grunow, A. (2017, April). Are these changes an improvement? Using data to inform the improvement of homework practices. Paper presented at the *Annual Meeting of the American Education Research Association*. San Antonio, TX.
Moen, R. D., Nolen, T. W., & Provost, L. P. (2012). *Quality improvement through planned experimentation* (3rd ed.). New York, NY: McGraw-Hill Education.
Murphy, R., Snow, E., Mislevy, J., Gallagher, L., Krumm, A. E., & Wei, X. (2014). *Blended learning report*. Menlo Park, CA: SRI Education.
Nelson, E. C., Batalden, P. B., & Godfrey, M. M. (2007). *Quality by design: A clinical microsystems approach*. San Francisco: Jossey-Bass.
Penuel, W. R., & Gallagher, D. (2017). *Creating research-practice partnerships in education*. Cambridge, MA: Harvard Education Press.
Penuel, W. R., & Shepard, L. A. (2016). Assessment and teaching. In D. H. Gitomer & C. A. Bell (Eds.), *Handbook of research on teaching* (pp. 787–850). Washington, DC: AERA.
Piety, P. J. (2013). *Assessing the educational data movement*. New York, NY: Teachers College Press.
Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407.
Russell, J., Jackson, K., Krumm, A. E., & Frank, K. (2013). Theories and research methodologies for design-based implementation research: Examples from four cases. In B. J. Fishman, W. R. Penuel, A.-R. Allen, & B. H. Cheng (Eds.), *Design based implementation research: Theories, methods, and exemplars: National*

society for the study of education yearbook (Vol. 112, Issue 2, pp. 157–191). New York: Teachers College Press.

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher*, *32*(1), 25–28.

Turner, E. O., & Coburn, C. E. (2012). Interventions to promote data use: An introduction. *Teachers College Record*, *114*(11), 1–13.

Van Campen, J., Sowers, N., & Strother, S. (2013). *Community college pathways: 2012–2013 descriptive report*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Yamada, H. (2017). *Do effects of Quantway® persist in the following year? A multilevel propensity score approach to assessing student college mathematics achievement*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Yamada, H., Bohannon, A., & Grunow, A. (2016). *Assessing the effectiveness of Quantway®: A multilevel model with propensity score matching*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.

Yamada, H., & Bryk, A. S. (2016). Assessing the first two years' effectiveness of Statway®: A multilevel model with propensity score matching. *Community College Review*, *44*, 179–204.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*, 64–70.