This chapter highlights the ways in which learning analytics can be used to better understand and improve learning environments, instruction, and assessment (Siemens & Long, 2011). As a set of approaches for engaging in educational research, learning analytics and educational data mining represent relatively new modes of inquiry. The growth of these approaches maps closely to the availability of new forms of data being collected and stored in digital learning environments, administrative data systems, as well as sensors and recording devices. Moreover, the growth of these fields maps closely onto what the National Science Foundation refers to as “data-intensive research,” which encompasses more than learning analytics and educational data mining to include a broad range of social and physical sciences. As new forms of data have emerged (i.e., transaction level data from digital learning environments as well as digital forms of audio, video, and text) and been collected at ever increasing scales, there has been an explosion of efforts to make use of these data for the purposes of research. By and large, most early work beginning in the mid-2000s was directed at exploring research questions that were tractable within highly structured, well-designed digital learning environments like intelligent tutoring systems (ITS; e.g., Koedinger, Anderson, Hadley, & Mark, 1997; VanLehn et al., 2005). The tight alignment between the learning tasks students were expected to engage in and the data that were collected in these environments made them ideal for exploring not just the outcomes of learning but the various ways in which students engaged in learning activities. A basic insight from these early researchers continues to fuel research and efforts to improve instruction—data on students’ learning processes is as useful and sometimes more so than data on students’ learning outcomes.

In this chapter, we expand upon this insight and highlight the ways in which data from digital learning environments, administrative data systems, and sensors as well as recording devices can be used to support instruction in real classrooms by reporting on students’ learning activities through various data products (e.g., dashboards). We do so across four cases that represent varying degrees of proximity to instruction. By highlighting these varying degrees of proximity, we intend to demonstrate the multiple ways in which learning analytics can be used to support instruction. Cases 1 and 2 describe efforts to use learning analytics to support instruction.
through partnerships that bring researchers closer to practice and practitioners closer to the work of analytics. Case 3 describes how process data from digital learning environments can be used to develop better assessments of learning that can be used to organize better learning opportunities for students. Case 4 describes how providing practitioners with access to carefully designed data products and dashboards can help them in making more timely and targeted decisions. The data produced in each of these cases are shared with stakeholders in different ways including online report systems or dashboards (Corrin, this volume).

Overview of Cases

Case 1 with Summit Public Schools and case 2 with the Carnegie Math Pathways are based in an approach to using analytics that Krumm and colleagues refer to as collaborative data-intensive improvement (CDI; Krumm, Means, & Bienkowski, 2018). Collaborative data-intensive improvement is an approach that combines tools and routines from improvement science, data-driven decision-making, as well as learning analytics and educational data mining. The overarching goal of this approach is to provide a structured way for researchers and practitioners to work together around identifying a question to pursue, analyzing complex datasets, developing change ideas, and testing change ideas in local learning environments. Across multiple partnerships, Krumm and colleagues (2018) identified a series of phases and supporting conditions for using data from digital learning environments and administrative data systems to improve learning environments and support instruction. Phase I of a collaborative data-intensive improvement project involves setting up a partnership, which includes identifying participants and jointly defining the aim of the partnership (Bryk, Gomez, Grunow, & LeMahieu, 2015). The second phase entails developing a practical theory for how the partnership will reach its aim (Bennett & Provost, 2015; Yeager, Bryk, Muhich, Hausman, & Morales, 2013). Phase III centers on data wrangling, exploration, and modeling (Wickham & Grolemund, 2017). Phase IV builds on insights from data-intensive analyses in the form of co-developed change ideas, and lastly, Phase V is where members of a partnership iteratively refine change ideas in real classrooms over time. Cases 1 and 2 describe the partnerships from which many of these phases were identified (Krumm, 2017).

Case 3 is situated in the context of introductory programming and computational thinking (CT), a new skill that seeing rapid adoption at all levels of school curricula as part of nationwide efforts to support “Computer Science for All” (The White House, 2016). There is a growing need to measure students’ learning of computational thinking in the context of the complex problem-solving processes inherent in programming, and also support all learners through this process of learning computational problem solving. Given that there are few examples of using learning analytics to measure students’ learning in open-ended programming environments that are popularly used in K-12 classroom, Grover and colleagues push into the emerging realm of computational psychometrics (von Davier, 2017) for detection of student behavior for formative assessment (Black & William, 2009; Heritage & Popham, 2013). They explored how principled, top-down, approaches of measuring complex skills can be combined with bottom-up, data-driven learning analytics approaches for better interpretation (of data logs from such programming environments), and consequently better measurement of computational thinking practices and programming processes (Zapata-Rivera, Liu, Chen, Hao, & von Davier, 2016). Based on learnings from analyzing data logs from ~300 students using the Alice programming environment, they developed a framework (Grover et al., 2017) that formalizes a process where a hypothesis-driven approach informed by Evidence-Centered Design effectively complements data-driven learning analytics in interpreting students’ programming process and assessing computational thinking in block-based programming environments. The framework is shared here, as well as a brief description of the application of the framework on an ongoing research project.
Case 4 is based in recent instructional reforms that address the importance of administrator and teacher making use of student assessment data to inform decisions about curriculum and instruction (Means, Padilla, DeBarger, & Bakia, 2009) and thus making instructional practices more effective (Mandinach & Gummer, 2016). The advances in technology and its popularity in schools have made it easier to collect student performance data. Learning analysts build dashboards and a variety of reports to incorporate information from such data, together with other possible sources of data, and present them to teachers. Although there have been many different types of dashboards built, few studies have shown evidence that teachers make use of information presented on the dashboards and adjust instructions accordingly. Case 4 describes an online homework support tool that was implemented in 44 schools for two years during a large-scale efficacy trial. Data collected during the study suggested that teachers implementing the intervention made substantial shifts in their approach to homework review and instructional practice more broadly.

Case 1: Data-Intensive Research-Practice Partnership

The partnership with Summit Public Schools (Summit) began in the fall of 2014 with the goal of developing a research-practice partnership around data collected and stored in multiple online learning systems used throughout the charter management organization. The partnership included researchers and practitioners from multiple organizational levels at Summit. The partnership built on the ideas of (a) learning directly from practitioners about the problems they experience in their day-to-day work; (b) jointly analyzing and interpreting data generated by students in digital learning environments to solve practitioner-identified problems; and (c) co-developing ideas for changes informed by multiple data-intensive analyses.

Around the same time as the fields of learning analytics and educational data mining were coalescing, new partnership models for engaging in educational research were emerging under the banner of research-practice partnerships (e.g., Coburn, Penuel, & Geil, 2013). Newer forms of data combined with newly developing models of research served as the primary building blocks for the partnership. The research goals for the project included (a) using Summit’s increasingly diverse and sizable datasets to answer their own research questions, and through engaging in these analysis activities, (b) develop a generalizable set of tools and routines for engaging in collaborative data-intensive research that other partnerships could use. To accomplish these research goals, we (i.e., Krumm and colleagues) used a design-based research approach (e.g., Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). A central feature of design-research is that it represents a mode of inquiry that seeks to build theory through directly intervening on learning environments (e.g., Barab & Squire, 2004; Bell, 2004). To inform our design, development, and intervention activities, we used theory and prior research from data-driven decision-making (e.g., Boudett, City, & Murnane, 2013), research-practice partnerships (e.g., Coburn, Penuel, & Geil, 2013), and learning analytics as well as educational data mining (e.g., Baker & Siemens, 2014).

Students in Summit consistently interact with digital learning environments in all grades and subject areas. This level of interaction results in a large volume of structured data on both what students are doing as they engage in learning tasks and how well they perform on those tasks. Summit believes that every student is capable of being college and career ready and that personalized learning opportunities can help students build necessary knowledge, habits, and skills. To accomplish this, Summit developed a whole-school approach where students engage in (a) project-based learning, (b) personalized learning time, and (c) one-on-one mentoring with teachers. Across these three learning opportunities, which span all grades and subjects, students interact with a common learning management system called the Summit Learning Platform (SLP). The
platform houses teacher-curated digital learning resources, such as online videos, and two types of assessments based on “playlists.” One of the two types of assessments referred to as a diagnostic assessment is used to identify gaps in students’ knowledge and a summative assessment, referred to as a content assessment, pulls randomly selected items from a large item-bank and is used to identify whether students have achieved mastery for a focal content area. Students can access resources and take assessments at their own discretion and as many times as necessary. Students spend 30% of their instructional time engaged in self-directed learning as they work to complete multiple summative assessments for a course and spend the remaining percentage of instructional time (i.e., 70%) engaged in project-based learning, which is an approach to instruction where students gain knowledge, skills, and productive dispositions by developing authentic products that are organized around broad and motivating driving questions (Larmer, Mergendoller, & Boss, 2015).

A challenge for any partnership is developing a focus for the partnership’s work (Penuel & Gallagher, 2017). To set the research direction for the partnership, we engaged in a multi-meeting, iterative process of having practitioners from Summit brainstorm topics and questions and having researchers reflect back and react to each question. Based on this process, the first data-intensive analyses that the partnership engaged in addressed the ways in which students attempted and completed content assessments. As noted previously, students have discretion in terms of when they attempt content assessments and how they prepare for them. Using data from students’ interaction with the Summit Learning Platform, we initially examined relationships among students’ standardized test performances on the NWEA MAP, their use of teacher-curated resources, and their content assessment taking in relation to course grades. Based on these analyses and within math courses, we observed that students who had lower incoming MAP math test scores tended to attempt content assessments more frequently. Beyond this, we also observed that students with higher MAP math scores used the Summit Learning Platform in different ways than their peers with lower incoming scores. For example, students with higher incoming test scores, on average, used more unique learning resources. For example, on one math playlist called “linear functions,” students in the lowest quintile on the MAP Math test used approximately 17 unique teacher-curated resources whereas students in the highest quintile used approximately 21. These same students also, as compared to their peers, used more resources prior to taking their first content assessment and overall attempted content assessments many fewer times. Overall, these early analyses hinted at the potential for using data from the platform to inform instruction—it provided a window into the processes that students were engaging in that held potential for explaining students’ eventual performances on individual playlists and for the course overall.

A key element of the overall partnership was working with practitioners at multiple levels of Summit—from organizational leaders to teachers—to make sense of data generated by the platform. Over time, we came to view opportunities to work directly with practitioners as learning events (Cobb & Jackson, 2012). These learning events proved to be the primary locations for going from a data product developed by researchers to a set of implications that would kick off the development of concrete instructional change ideas. At a general level, learning events involved structured activities where members of the partnership developed new understandings by engaging in joint work. Types of learning events included simple meetings where members of the partnership jointly interpreted data products, but learning events also included structured co-design session (Penuel, Roschelle, & Shechtman, 2007), opportunities where teachers could work first hand with data products that researchers developed, and workshops where researchers provided explicit instruction on data analysis software. Learning events provided opportunities for both researchers and practitioners to use their respective expertise to make the data useful for instructional improvement.
At a multi-day learning event referred to as a data sprint, Summit staff worked directly with data and engaged in data wrangling, exploration, and modeling tasks in collaboration with researchers. One data product that the partnership built upon out of this event involved an analysis that identified students who scored low on an assessment and followed it up with another assessment—and often another low score. This cycle of repeated, negative assessment taking was thought to stall students’ progress and lead to, in some cases, students falling further behind their peers. Using math courses once again, the partnership operationalized these patterns as conditional probabilities (i.e., conditional on a student not succeeding on an assessment, what is he or she likely to do next based on prior use of the platform?), and then scaled these analyses to include all grades and courses taught at Summit. These follow up analyses revealed that patterns referred to as adverse transitions were correlated with poorer performances across a range of courses, and also that these patterns declined in frequency over time, which demonstrated that students gradually stopped making these transitions.

Along with the types of transitions that students made following a low score on an assessment, we also explored students’ use of learning resources across playlists. Recall that each playlist in the Summit Learning Platform is comprised of both assessments and resources, and that students are expected to use resources in order to help them pass assessments. We used an unsupervised machine learning approach referred to as hierarchical cluster analyses, combined with a heat map visualization, to explore patterns in students’ resource use (e.g., Bowers, 2010). Much as with the conditional probability analyses coming out of the data sprint, an important next step following the resource-use heat map analyses was scaling the visualization to include all grades and subjects. Key to making analyses useful to practitioners and avoiding over generalizing a finding, we explored within-courses patterns of resource use in order to control for variations in content and the developmental differences of students across grades. Taking an analysis to scale meant examining whether a pattern identified in a handful of courses appeared in other courses. The ability to run analyses on one course and then on all courses proved to be an important value that the research team brought to the overall partnership.

Our partnership with Summit highlighted the ways in which researchers and practitioners can come together to jointly analyze and take action on data from digital learning environments. The design-based nature of the project, which was organized around bringing researchers closer to practice and practitioners closer to research, surfaced multiple factors associated with data-intensive partnerships and helped in clarifying the multiple steps that can be involved in using large, complex datasets to improve instruction.

Case 2: Measuring Productive Persistence to Help Faculty and Students

The second of two cases described in this chapter that was central to the development of collaborative data-intensive improvement as an approach phases and conditions was with the Carnegie Foundation for the Advancement of Teaching (Carnegie) and the Carnegie Math Pathways. At the start of our work together, Carnegie was well into launching and supporting the Pathways, which is a national effort focused on improving developmental, or remedial, math courses in two- and four-year colleges. At many colleges, these courses are significant barriers to a student’s college completion (Bailey, Jeong & Cho, 2010). To help more students get past the hurdle of developmental mathematics, Carnegie brought together researchers and practitioners around the tools and routines of improvement science (Langley, Moen, Nolan, Nolan, Norman, & Provost, 2009). As the “hub” of a developing group researchers and community colleges, Carnegie formed a networked improvement community (NIC) in an effort to accelerate learning and improvement among network members (Bryk, Gomez, Grunow, & LeMahieu, 2015). Members of the Carnegie Math Pathways NIC designed two different course sequences geared toward
helping students fulfill their developmental math requirements as well as earn college credit in either statistics (i.e., “Statway”) or quantitative reasoning (i.e., “Quantway”).

The success of both Statway and Quantway are well documented (e.g., Yamada, 2017; Yamada, Bohannon, & Grunow, 2016; Yamada & Bryk, 2016). Key to the success of the Carnegie Math Pathways NIC is a systemic approach supported by the use of improvement tools and routines. One component of Carnegie’s systemic approach is a focus on “noncognitive” factors that affect student success (see Yeager & Walton, 2011; Zimmerman, 2002). Many of these factors center on students persisting through failure and using good learning strategies, which Carnegie defines as “productive persistence.” At the beginning of our partnership, Carnegie wanted to explore how data from various online learning systems used in the Pathways could be leveraged in measuring and supporting students’ academic tenacity and use of effective learning strategies (see Krumm et al., 2016).

We began working with data from Statway’s online learning system at the time, which was the Online Learning Initiative (OLI) platform. The platform collected information on each page that a student viewed as part of the Statway curriculum, when the page was viewed as well as information on a variety of assessments housed within the system. Through the platform, Statway provided students with practice assessments that students could use to test their own knowledge embedded within the material that they were reading. Each item that was attempted on an assessment, when it was attempted, and whether an item was answered correctly or not were collected and stored by the Online Learning Initiative platform. Along with page-views and practice assessments, the platform also captured time- and item-level data from assessments referred to as “Checkpoints,” which are quiz-like assessments that come at the end of “topics” and “modules” that make up the Statway curriculum.

In the fall of 2014, we started to explore the ways in which the online system was used across individual Pathways courses. One of the benefits of looking at data stemming from the Online Learning Initiative platform was that these data were collected unobtrusively at the scale of entire Pathways NIC. These data were unobtrusive in that they were gathered directly from students as they engaged in learning activities based on what was programmed to be captured by the online learning system. While large volumes of data could be collected, that did not mean that all of it would prove to be useful for understanding students’ learning behaviors, strategies, or outcomes. One step involved in identifying useful data involved becoming familiar with students’ experience of using the Online Learning Initiative platform, such as the ways in which students could read pages, practice material, and take formal assessments—along with the ways in which these data were collected and stored by the system.

One the first exploratory analyses that we conducted focused on the dates with which students submitted Checkpoints. We were interested in identifying how much variation there was among students within a course for when they turned in a Checkpoint as well as between courses for when, on average or modally, students completed a Checkpoint. These analyses revealed that approximately half of all Statway courses had a modal pattern that followed the intended order of Checkpoints and that individuals as well as courses that followed the intended order tended to perform better in terms of end-of-course grades. We followed up course-level analyses by further 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. We explored patterns in what students did before and after a low score on a Checkpoint (i.e., a score below 60%). A key component of productive persistence is persisting with a task after experiencing challenge or difficulty. Low scores on Checkpoints offered a unique opportunity to measure these behaviors (Krumm et al., 2016). We also explored the number of sessions a student logged per week, the number of days between each session, and the types of sessions that students logged, such as assessment only
sessions where students only worked on Checkpoints or robust sessions where students engaged in reading, practicing, and assessment activities within the same session. All of these different operationalizations helped in understanding the ways in which productive persistence played out and could be measured using data from the OLI platform.

One way in which we sought to expand the use of these measures was to put them in front of Statway faculty to better understand how well they captured students’ learning strategies and behaviors as well as whether they could be used to help faculty more effectively intervene with students. In working directly with faculty, we organized design workshops that were geared toward jointly interpreting data products that the research team provided and co-developing data products and follow up actions, such as change ideas that faculty could implement using a data product. Design workshops were structured activities where researchers would present evidence on students’ use of the online learning system and instructors would co-develop additional data products, follow-actions, or both. Over time, the partnership viewed design workshops as the location where data became actionable. Despite the sophistication of any analysis, no data product proved to be actionable in and of itself; each data product required an explicit action to be developed.

One of our first workshops was organized around developing instrumental data products related to students’ productive persistence within the Online Learning Initiative platform. Outcomes from this first workshop included finding new ways to operationalize students’ engagement with online learning materials over time and creating data products that captured what students did alongside how well they did. For a second design workshop, evidence for the importance of attempting and succeeding at Checkpoints had been building across multiple analyses, and the data products and change ideas that were developed during this workshop led to a focused improvement project related to students completing Checkpoints. An initial improvement sprint following the workshop led to demonstrable increases in students completing end-of-module Checkpoints (Meyer, Krumm, & Grunow, 2017).

Across multiple iterations, the design workshops themselves as well as the data products and change ideas that were produced to support them proved to be valuable for both researchers and practitioners. For researchers, they offered venues for learning from faculty on what they found meaningful and whether certain patterns that were identified had face validity. For practitioners, they offered an efficient touch-point for engaging in data-intensive research activities. While they offered efficiencies for practitioners, they required significant pre-work on the part of the research team both in terms of data analysis and in organizing the workshops themselves. Following up with practitioners after a workshop was key to the overall success of the workshop. Overall, these workshops were a potent strategy for translating findings from data-intensive analyses into changes in instructional practices.

Case 3: Learning Analytics for Supporting Novice Programmers

The Context of Introductory Programming in K-12 Classrooms

Policy and educational leaders see computer science (CS) and computational thinking (CT) skills (Grover & Pea, 2013, 2018; Wing, 2006) as necessary for all citizens, not only computer scientists, with a view to building a strong STEM pipeline. Such problem-solving skills are seen as necessary to succeed and innovate in a world infused with—and lives shaped by—computing and digital devices.

Most K-12 computer science courses teach programming to support learning of computational thinking practices such as logical and algorithmic thinking, decomposing problems, debugging, and use of computational thinking concepts to create solutions that can be executed by a computer. However, programming has historically been difficult for novices to learn (e.g., Pea & Kurland, 1984;
Applying Learning Analytics to Support Instruction

This is because programming is a complex activity that involves understanding a problem as a computational task, mapping a design for the program, drawing on problems previously programmed that have a similar structure, instantiating abstract program patterns, coding the program, and then testing and debugging (Pea & Kurland, 1984). It involves not only issues of syntax of the programming environment but also the semantics of putting together computational solutions as well as strategies and pragmatics such as testing and debugging the code.

These problems persist for novices despite the emergence of block-based programming environments that provide a visual programming interface that makes it easy for novices to get started with creating programs and animations without worrying about issues of programming syntax. However, these environments do not currently aid in formative assessment of the use of computational thinking practices and disciplinary concepts of computing to aid the learning process in the context of programming. Examining programming process using learning analytics (LA) gives a more complete picture (Baker & Siemens, 2014). Being able to support and scaffold this process requires us to have the ability to detect and recognize actions (single or sequences of multiple actions taken together) as evidence in support for or against the use of computational thinking. Thus, students’ actions need to be interpreted as they work so that formative feedback can be provided to steer learning.

Recent learning analytics work in the context of programming has included analyzing students’ steps to a solution using data from digital environments such as number of actions in students’ programs and number of successful and unsuccessful program compilations (Blikstein et al., 2014). The use of clustering techniques (Bouchet, Harley, Trevors, & Azevedo, 2013) has led to identifying various programmer behavior profiles, and unsupervised methods have been used to derive program-state patterns and state transitions to predict success outcomes (Berland, Martin, Benton, Petrick Smith, & Davis, 2013). Most of these techniques have involved looking for patterns in data largely from the “bottom-up” (Winne & Baker, 2013).

New hybrid or blended LA have begun to assess students’ learning processes in digital learning environments for science & math that combine top-down and bottom-up approaches to better understand students’ knowledge and skills. Examples include Gobert, Sao Pedro, Raziuddin, and Baker (2013), Shute and Ventura (2013), and Zapata-Rivera, Liu, et al. (2016). Many of these by combining bottom-up LA with Evidence Centered Design (ECD; Mislevy, Almond, & Lukas, 2003), a principled approach to guide assessment design for top-down, hypothesis-driven generation of a priori patterns about learner actions. Evidence Centered Design focuses on three related models: student (what are targeted cognitive constructs?), task (what activities allow students to demonstrate cognitive constructs?), and evidence (what data provide evidence of cognitive constructs?). It helps connect important constructs that we want to measure with observable behaviors (including patterns of learner actions). Also, importantly, evidence is obtained by deliberately putting students in situations or tasks that will elicit the needed evidence. Once semantically meaningful patterns are defined a priori, data mining and learning analytics techniques can be used to analyze the patterns further.

We present a theoretical framework that researchers can use to design measurement systems for programming environments for research or application. We are using this framework currently as part of a broader effort to study and detect patterns of learner behavior during programming, as a first step toward being able to provide feedback to the learner and instructor about student learning in real-time.

**Exploratory Work as a Backdrop to the Evolution of a Framework**

We analyzed a dataset from an assessment task designed and used in prior research (Werner, Denner, Campe, & Kawamoto, 2012). 118 females and 202 males aged 10 to 14 years completed
the 30-minute task which involved modifying existing code in the Alice programming environment (Dann, Cooper, & Pausch, 2009). Students’ computer programs and Alice data logs were collected, and the programs were scored manually using a rubric for computational thinking including algorithmic thinking and abstraction. We applied Evidence Centered Design to “reverse engineer” this task into specific computational thinking concepts and skills and give evidence of what those might look like in log files. We also compared action sequences between students who scored high and low (relative to the median) to determine commonality of sequences for each group. We found sequences that were significantly more common among students with high grades and one sequence that occurred significantly more frequently for students with low grades. Our analysis showed positive correlations among higher grades, number of code edit actions, and number of testing events.

Through our exploratory work, we gained insights into interpreting student actions from logs. However, we also discovered that tasks need to be complex enough to yield rich process data logs as students apply more strategic computational thinking skills for better coverage of focal constructs. Measuring learning through automated means requires evidence of appropriate as well as repeated use of constructs (Koedinger, Corbett, & Perfetti, 2012). Lastly, it became apparent that without additional measures for ground-truthing or mapping the sequences back to specific instances in students’ programming progressions that we have evidence for, one cannot validly interpret such sequences.

**A Framework for Blending Hypothesis- and Data-Driven Learning Analytics**

Building on the learning from our exploratory work, we designed a framework, or process, that employs Evidence Centered Design in its typical forward-design application, beginning from important focal knowledge and skills, and proceeding to task implementation. This approach yields an overall methodology for combining top-down Evidence Centered Design-like approaches to assessment development and delivery with bottom-up, data-driven LA approaches.

The framework (Figure 9.1) describes an iterative process that begins with identifying important computational thinking concepts and practices that we would like to measure. Careful design of tasks put students in situations that evoke behaviors to provide potential observables of these concepts and practices. Detailed analysis of program code from different solutions reveals students’ use of constructs (correct or otherwise) and varied approaches to solutions. Similarly, analyzing data from screen recording and/or in-person “over-the-shoulder” observations reveals aspects of students’ actions that are never seen in the final program. These can reveal student misunderstanding of concepts even if the final solution seemingly demonstrates appropriate usage of constructs. Combined qualitative analyses of the program solutions along with data-driven examination of programming process of a designed task together provide a deeper understanding of students’ actions than is possible from data-driven analytics alone, including potential code sequences that map to practices that are identified through data logs. These *a priori* patterns lay the foundation for detectors for these patterns and provide a richer interpretation of student process in programming environments.

**Applying the Framework**

Guided by the framework, we applied Evidence Centered Design for the design of programming tasks to generate richer process data to observe repeated use of constructs and computational thinking practices. Two such tasks were piloted in two high school introductory computer science classrooms with 27 and 28 students. Data included final Alice files and log data for all
students and screen recordings for six students. Analysis of logs revealed similar issues that students struggled with in both tasks, for example, hard-wired vs. general solutions, improper termination conditions, decisions pertaining to parallel vs. sequential execution, effective solution decomposition, and (in-)appropriate random number use.

Analyses of screen captures from the six students using a “process over product” lens to assess computational thinking practices suggested that some students demonstrated abstraction, modularization, and testing in parts while others did not. Such observations will serve as useful patterns to search for in students’ log data as evidence for computational thinking skills. In addition, we noticed certain phases during students’ programming process when a student was unable to progress. Such situations can easily lead to frustration and loss of engagement and can thus serve as good candidates for potential patterns to be detected as students work on their assessment tasks. Detection of such flailing behaviors in real-time can help a teacher identify when to help a student.

Task piloting and analysis led to more refined tasks that were then used in three high school classrooms in the Western US with a total of close to 100 students. Data collection also included
screen recordings and interviews with three students in each classroom. The screen recordings are being used to validate program snapshots created from log data, in addition to aiding student recall of process during one-on-one interviews with each student. These interviews are also being used to ascertain the nature and timing of help that students may have liked to support their work. This will help us understand the nature of formative feedback and supports that can scaffold learning for students during programming.

Case 4: Learning Analytics Enabled Formative Assessment and Changes in Teacher’s Instructional Practices

Homework is already required in schools and a meaningful amount of instructional time is allocated to homework (Fairman, Porter, & Fisher, 2015; Loveless, 2014). But it is also controversial and perceived as needing improvement (Kohn, 2006; Bennett & Kalish, 2006; Trautwein & Koller, 2003). Online homework tools can provide immediate feedback to students and real-time information for teachers to monitor student progress. In this section, we focus on how teachers’ homework review practices change when they have access to data on student homework performance and the role of such formative assessment data for informing teachers’ instructional decisions and adaptations.

Formative Assessment and Data Use in School

The concept of formative assessment has received much attention in K-12 research and practitioner communities (Black & William, 1998a, 1998b; Boston, 2002; Heritage & Popham, 2013; Roediger & Karpicke, 2006). Researchers and practitioners characterize formative assessment as a process that uses student data to inform adaptive changes in instruction (Bennett, 2011; Brookhart, 2007; Guskey, 2007; Heritage & Popham, 2013). The growing interest in formative assessment is, in part, an outcome of the general dissatisfaction with the quality of information obtained from summative assessments that generally do not provide sufficiently fine-grained or timely feedback on student learning (McMillan, 2007; Wiliam, 2016). Research documents modest to medium effect sizes of formative assessment on student learning (Black & Wiliam, 2009; Brookhart, 2007; Guskey, 2007; Hattie, 2009; Kingston & Nash, 2012; Shavelson, 2008; Speece, Molloy, & Case, 2003; Thum, Tarasawa, Hegedus, Yun, & Bowe, 2015) for a variety of different modes, grade levels, content areas, and cultural settings. Frequent use of formative assessments can improve achievement, particularly when the results are used to adjust instruction (Bergan, Sladeczek, Schwarz, & Smith, 1991; Speece, Molloy, & Case, 2003).

In recent years, administrator and teacher use of student assessment data to inform instructional decisions has been at the forefront of instructional reforms (Means, Padilla, DeBarger, & Bakia, 2009). Advocates of these reforms emphasize that teaching should be responsive to student needs and assessment data is essential to enable teachers to adjust instruction to better support individual learners. The expectation is that teachers skilled in data use will develop more effective classroom and instructional practices (Mandinach & Gummer, 2016).

The ASSISTments Online Homework Support Tool

ASSISTments is a web-based platform that provides support to students as they solve mathematics problems and provides detailed student-level and class-level formative assessment data to teachers to help inform adjustments in classroom instruction and pacing (Heffernan & Heffernan, 2014). Prior small-scaled studies showing the promise of ASSISTments have been synthesized (Rittle-Johnson & Jordan, 2016). In a recent efficacy trial funded by the IES, SRI
recruited 46 middle schools from Maine, including 87 teachers and over 2,800 seventh-grade students. In the study schools were randomly assigned to the treatment or control condition for two years. Schools assigned to the control condition continued with their homework practices as they normally would. For the schools assigned to the treatment condition, in the first year, teachers received professional development and practiced using ASSISTments with their seventh-grade classes. In the second year, these teachers continued to use ASSISTments with a new cohort of seventh-grade students, who were the student-level study population. The TerraNova Common Core mathematics assessment was administered to students to measure end-of-year outcomes. Using a hierarchical linear model (HLM), we analyzed student outcomes by condition. The adjusted mean scores on the TerraNova were 8.84 points higher in the treatment condition, and this result was statistically significant (effect size $g = .18$, $p = .007$) (Roschelle, Feng, Murphy, & Mason, 2016). According to published technical norms (CTB/McGraw-Hill, 2012) that relate TerraNova scale scores to grade level equivalents, the degree of improvement corresponds to what would be expected from .5 to one years of additional learning time.

A key component of ASSISTments is the easy-to-use online reports. The system analyses use log of all students and generates Item Report that shows data for each student on every problem and each math skill covered in the assignment, which questions and/or skills were particularly challenging, and what the common wrong answers were. The data allows teachers to make real-time, informed decisions about what to teach next, and it is ideally used to guide homework review. Figure 9.2 shows an item report with the results of six items. Teachers can see the percent correct per problem and use that data to identify weaknesses for the class. The common wrong answers support cognitive diagnosis of misconceptions. Problem numbers appear across the top row, class-level results appear in the next rows, and individual student results appear on anonymized rows below. Each cell shows what the student entered first. The cell will be yellow if the student had to be shown the answer.

Research has found that teachers do not typically know how to use data to inform instruction (Mandinach & Gummar, 2016; Means et al., 2009) or they could make errors when trying to make sense of score report results (Zapata-Rivera, Zwick, & Vezzu, 2016). Pape et al. (2013), however, found that when both professional development and formative assessment technology are provided, teachers can learn more about their students and adapt instruction with resulting improvements in student outcomes. In the ASSISTments model, teachers received a total of five days of professional development. The professional development entails discussions of the foundational instructional and learning theories behind ASSISTments as well as practical issues associated with the use of ASSISTments. Teachers learn how to use the system and how to interpret ASSISTments reports. They also receive advice on practical instructional strategies for responding to students with different needs. The later sessions focused on helping teachers sharpen their ability to adjust instruction in response to information in the reports and refine their class routines.

Impact of Using ASSISTments Reports on Teacher Practices

During the Maine efficacy study, extensive data was collected through system use log, interviews with teachers and school principals, teacher observations, surveys, and logs, and system use data generated within ASSISTments. All the data focus on the implementation process leading to outcomes. We analyzed and triangulated data from different sources to see whether teachers used data from ASSISTments reports as a part of the implementation of the intervention and whether teacher’s instructional practices changed given the availability of the formative assessment data.

Based on the teacher use log data, we measured the proportion of ASSISTments reports that a teacher opened at least once. Opening a report is an important indicator of whether
Percent correct per problem. This data identifies class weakness helping you drive your instruction.

<table>
<thead>
<tr>
<th>Student/Problem</th>
<th>Average</th>
<th>PRABFGH</th>
<th>PRAFBG</th>
<th>PRAFBG</th>
<th>PRAFBG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem Average</td>
<td>72%</td>
<td>82%</td>
<td>42%</td>
<td>81%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Common Wrong Answers will present clues as to why students answer incorrectly.

<table>
<thead>
<tr>
<th>Common Wrong Answer(s)</th>
<th>A. 50% [feedback]</th>
<th>C. 42% [feedback]</th>
<th>A. 34% [feedback]</th>
<th>D. 23% [feedback]</th>
<th>B. 76% [feedback]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Problem 2</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Problem 3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Problem 4</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Problem 5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Problem 6</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

This means 76% of the incorrect answers were B. The percent will be in **highlight** if it is over 50%.

Figure 9.2 An item report for teachers showing the results of six items.

(Neil & Cristina Heffernan)
Applying Learning Analytics to Support Instruction

the teacher is using ASSISTments to review student work and is a precursor to using ASSISTments to adapt instruction. Across classrooms, the median for report-opening was 64%, which is above the expected opening rate (50%).

In the instructional logs and surveys, teachers were asked whether they reviewed all homework problems, they asked students which problems to review, or they reviewed selected problems based on student’s performance (aka. data-driven targeted review). We found that the intervention had statistically significant effects on homework review practices. When the continuous variable (based on teacher logs) was used as the outcome measure, the effect size was 1.23, \( p = 0.005 \) and for the dichotomous variable (based on the teacher survey), the odds-ratio was 45.8, \( p = 0.001 \). Among all 38 treatment teachers who responded to the survey, 37 reported doing targeted homework review. While in the control condition, 12 of the 36 teachers reported that they didn’t do targeted homework review.

Analysis of interview transcripts and classroom observations data also provided convergent evidence of shifts in teaching practices. These shifts centered around three areas:

1. Targeted in-class review of homework problems and concepts based on needs of students. Compared to those in the control condition, treatment teachers were more likely to focus their homework review (\( p < .01 \)) to cover fewer number of homework problems but in more depth. Teachers stated that the item report provided a starting point for their instructional planning; they reviewed the item report to quickly identify problems where a majority of students struggled and the common wrong answers and purposefully select which concepts they needed to review during the class. In contrary, control teachers relied on students’ willingness to ask for help on certain problems, random or sequential selection of homework problems for review, recitation of correct answers of all homework problems but not demonstrating or discussing solution procedures, or projection of answers for students to self-correct.

2. Use of data from homework to initiate and motivate homework discussion. Presenting reports engages students directly with data on the homework results and reduces students’ reluctance to ask for help on problems, as they could see that other students struggled with some of the same problems. This helped to create a safe classroom environment where students were more willing to speak up and engaged in the discussion of homework.

3. Use of homework data to inform instructional decisions during subsequent lessons. Treatment teachers acknowledged that they used the data from ASSISTments to inform instructional decisions broadly. These decisions included: instructional pacing, what concepts they needed to address during subsequent lessons, and/or which students to provide more instructional support. Treatment teachers viewed the ASSISTments reports as a valuable resource for understanding how students performed on the homework generally but more specifically how well students understood or struggled with certain concepts and procedures.

**Conclusion and Discussion**

In this chapter, we presented four cases that demonstrate how learning analytics can be used to improve learning environments across different grade levels and subjects. While data from digital learning environments, administrative data systems, as well as sensors and recording devices can be used to support instructional improvement, it is important to recognize that these improvements are as much about the supporting work of researchers and practitioners as they are about the data themselves—data is not a self-activating resource as it requires teams
of individuals to interpret, derive implications, and develop change ideas. Across the four cases described in this chapter, researchers and practitioners working in collaboration as well as the use of approaches such as Evidence Centered Design can provide structures and activities for translating data into an instructional change.

Key to the types of data addressed across the four cases is that they originated from processes that preceded a valued outcome, such as an end-of-course grade or standardized test performance. While these data can be collected from activities over time, they still need to be reliable, valid measures of those processes. One challenge to creating valid measures is that the technology from which the data are being collected may not collect all of the relevant data (Krumm, Means, & Bienkowski, 2018). A great deal of work and energy can go into analyzing these data, all the while critical instructional activities are occurring outside of the learning system. Working directly with practitioners can help in better understanding the instructional context in which technologies are used as well as in making more informed interpretations of the events that are captured by a technology. Moreover, approaches such as Evidence Centered Design provide a framework for interpreting available data and in developing an evidence-based argument around what processes are being measured.

Assessment in online learning has been studied for a number of years, but only recently, researchers have begun promoting and advocating the use of learning analytics for assessing academic progress, predicting future performance, and spotting potential problematic issues (Johnson, Smith, Willis, Levine, & Haywood, 2011, p. 28). The Gordon Commission (2013) recommends “separate responsibility for the use of data drawn from rich descriptions of these transactions for administrative and for student development purposes. Teachers would be enabled to interpret these data diagnostically and prescriptively” (p. 15). When using learning analytics for assessment, researchers are urged to differentiate assessment of learning (e.g. summative assessment) versus assessment for learning (e.g. formative assessment, diagnostic assessment). When the purpose of the assessment differs, the design of the learning task’s focal knowledge, its features and timing (e.g. when learning is still happening vs. when learning has completed), its alignment with learning standards, potential observations, and inferences from the tasks shall be adjusted accordingly. Learning analytics can be a powerful tool for formative assessment, and for instructors to take corrective measures and monitor progress. Data collected through learning environments tends to be rich, multi-dimensional, longitudinal, embedded, and importantly—inexpensive. Such data can provide opportunities for assessing learners at a much finer-grained scale than a traditional exam; we can not only score an answer entered by a learner right or wrong, but also look at characteristics of how learners answer the question, such as how long it took them to answer, or whether their mouse hovered over a wrong answer for a while, to gauge the level of performance and confidence. On the other hand, such data can also be noisy as compared to data collected from more controlled testing environments. While there are promising applications as shown in case 3, strong evidences are warranted with regard to reliability and validity of the measures produced by learning analytics (Tannenbaum, this volume).

Acknowledgements

The research reported in this chapter was supported by the Institute of Educational Sciences, US Department of Education (R305A120125), National Science Foundation (SMA-1338487, DRL-1444621, and DRL-1418332), the William and Flora Hewlett Foundation, and SRI International. The opinions expressed are those of the authors and do not represent views of the funders. We would like to acknowledge the contributions of Marie Bienkowski, Barbara Means, Jeremy Roschelle, Neil Heffernan, Janet Fairman, Satabdi Basu, John Stamper, Michael Eagle,
Applying Learning Analytics to Support Instruction

and Nicholas Diana in the work presented here. We would also like to thank Jill Denner and Linda Werner for their Alice dataset that informed the research.

Note

1 We noticed that the odds-ratio for the dichotomous mediator is very big. This was possibly due to the lack of variability in the mediator for the treatment condition and a big contrast between conditions.

References


