Data-wary, Value-driven:

Teacher Attitudes, Efficacy, and Online Access for Data-Based Decision Making

Michael Aaron Hawn

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Science

COLUMBIA UNIVERSITY

2019

© 2019

Michael Aaron Hawn

All rights reserved

ABSTRACT

Data-wary, Value-driven: Teacher Attitudes, Efficacy, and Online Access for Data-Based Decision Making

Michael Aaron Hawn

How do teachers use online student assessment data? School districts invest increasing resources in online systems for reporting and analyzing student assessment data, yet few studies describe the direct use of such applications or explore how these systems relate to teachers' professional roles, data use attitudes, or data use efficacies. This dissertation applies learning analytics methods for log file analysis and visual data analytics to explore the extensive variation in teachers' online data use behaviors and attitudes over six months in one urban secondary school. Descriptive statistics and visualizations of online usage over time demonstrate strong connections between teachers' online behavior and common organizational factors, such as school level (middle vs. high school), content area, and required training. Correlational evidence suggests that data use self-efficacies have stronger relationships to online use than general data use attitudes. Hierarchical cluster analysis heatmaps are used to identify novel subgroups of teacher online data use behaviors and attitudes. These exploratory findings are used to generate data use dashboards for school-based leadership and an expanded determinant framework for the adoption of online assessment systems. Combining data-intensive methods with theoretical frameworks for self-efficacy, technology acceptance, and use diffusion, this dissertation aims to describe the rich variation in teachers' online data use and attitudes, as well as productively inform the practice and study of data-based decision making in schools.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Acknowledgements

I would like to thank my advisor, Ryan Baker, for his insight, clarity, support, and encouragement over many years. His genuine excitement and openness to learning were clear from the first time I walked into his office and continue to humble and inspire today. His mentorship has been honest and invaluable, and his guidance through the world of learning analytics has consistently challenged, inspired, and entertained. As a teacher and practitioner, I am especially grateful for his appreciation of the challenges that educators face every day and his insistence in keeping the focus on helping students learn.

To the members of my dissertation committee, my sincerest appreciation for your time, feedback, and support. I cannot say thank you enough to Alex Bowers for his extensive mentorship and guidance through this dissertation process, far beyond the call of duty. His relentless enthusiasm, generous feedback, and openness to sharing tools and ideas betters those around him. Dragan Gašević provided key direction for framing these studies in theory and skyped in at ridiculously early hours from out-of-sync time zones. Gary Natriello generated enough ideas to keep me busy for years and constantly kept the important questions in view. Carolyn Riehl understood the story I was trying to tell better than I did and grounded my exploration in the real context of teachers working hard to better understand their students. I am sincerely grateful for their suggestions, critique, and enthusiastic discussion. While it is possible that other PhD candidates have such supportive, authentic, fun, and inspiring committees, I think it unlikely.

I would also like to express appreciation to all my professors and colleagues at Teachers College and the Baker EDM lab. To Deanna Kuhn for your mentorship and for bringing me to

Columbia. To Mia Almeda, Miggy Andres, Ama Awotwi, Rainikka Corprew, Jaclyn Ocumpaugh, Luc Paquette, Sweet San Pedro, Stefan Slater, Colleen Uscianowski, and Yang Jiang for your feedback, tutelage, patience, and senses of humor.

Thanks to my parents and sister for putting learning and the richness of understanding at the center of our family. To my in-laws for countless afternoons and weekends watching the grandkids and for not asking how much longer. Gratitude and love to Caillin, Charlie, and Ada for their understanding and patience with a busy parent and for the joys they give me without even trying. And to Kate, for endless patience, unwavering support. Your fierce and loyal cultivation makes our family a better, greener place.

Lastly, to the many teachers and leaders in education I have had the privilege to work with. Your motivations are the best, your challenges are real, your persistence is inspiring, and your impact in the world is greater than you know.

To Kate

Over the last two decades, increased public accountability for schools and the expansion of large-scale, networked student data systems have encouraged a set of school management and instructional strategies grouped under the broad heading of data-based decision making (DBDM) (Halverson, 2014). The No Child Left Behind Act of 2004 formalized a structure for that accountability and provided focus for the expansion of district-wide student data systems (Marsh, Pane, and Hamilton, 2006). This alignment of a government mandate (public accountability) and technological tools (large-scale data systems) provided motivation and capability to capture, combine, and distribute an unprecedented amount of data on students (Bowers, Shoho, and Barnett, 2014).

Administrators and teachers are now inundated by data from student information systems, assessment systems, behavior tracking systems, and instructional software, as well as by streams from more traditional data sources, such as teacher observation and grading (Mandinach and Jackson, 2012). Broadly speaking, DBDM initiatives attempt to leverage these multiple streams of data in order to improve educational decision-making processes and outcomes. The 2014 Learning Analytics Workgroup, organized by Stanford University, addressed the following core educational challenge: "the growth of data in education surpasses the capacity to make sense of it and to employ insights derivable from the data to guide practices" (Pea, 2014, p. 2). This challenge of data is frequently investigated at large scale, across districts or states (Pea, 2014). However, the same challenge exists at the local level, as schools and teachers attempt to engineer interfaces for converting a flood of data into reasoned and effective action.

In the face of increased accountability for student outcomes, these improved, data-based decisions—made at any level from the classroom to the district—have been singled out by multiple stakeholders as a silver bullet for educational improvement, resulting in myriad management and instructional initiatives, as well as a wide range of commercial assessment products. School leaders and teachers have worked extensively to leverage data for accountability purposes and to shift schools from decision-making cultures based on professional intuition to cultures based on collaborative discourse around evidence (Bowers et al., 2014). Of course, this shift in epistemology presents a variety of challenges to schools, in everything from an increased need for data-literacy among teachers (Mandinach and Gummer, 2013) to heightened expectations for student performance.

Unfortunately, the expectation of data use in the classroom has far outstripped the evidence or even the investigation of its impact. Researchers have largely focused on the potential of data usage, with much less attention given to exploring and testing the subsequent teacher and student outcomes associated with that usage (Coburn and Turner, 2012). Where researchers have attempted to capture the potential of multiple and dense streams of student data, results have been mixed, and outcomes relating data usage to teacher and student outcomes have been particularly difficult to pin down (Mandinach and Jackson, 2012). At the time of the 2009 IES Practice Guide, the level of evidence for recommending DBDM practices was low, even for such basic recommendations as “make data part of an ongoing cycle of instructional improvement,” “provide supports that foster a data-driven culture within the school,” or, surprisingly, even for a recommendation as fundamental as “develop and maintain a districtwide data system” (Hamilton et al., 2009). Almost a decade after this report, the practice of DBDM, while effectively changing the nature of pedagogical conversation nationwide (Mandinach and Jackson, 2012), has yet to gather evidence sufficient to fully justify its potential for improving student learning.

For those who see the myriad possibilities for using evidence to improve decision-making in schools, this weakness of evidence calls out for further investigation. Has data usage simply had too little time to iterate, improve, and add to student gains? Have studies failed to capture key parts of the DBDM process? Or, does current DBDM practice have limited impact on students for other reasons not yet understood?

If this disconnect between the expectations of DBDM and its outcomes continues, educational leaders will face an eventual stalemate and their own, data-driven decision: should they cut their losses and step back from this promising intersection of evidence, analysis, and instruction? Or, should they dig deeper into the behaviors and attitudes of teachers' practice and broaden their analysis of student outcomes? Given the difficulties in finding clear effects for data-based decision making, more exploratory work is needed to observe and analyze teachers' interactions with data and to place these interactions within the larger web of relevant organizational and teacher factors. Towards this end, I ask three research questions about educators' attitudes towards data and their online use of an assessment system.

Purpose of Study and Research Questions

Applying learning analytics, social-cognitive theory, and technology acceptance models, I examine how teachers use data in one urban secondary school, describing the relationships between their attitudes towards data use and their actual behaviors in an online system for viewing and interacting with student assessment data. This study proceeds from the position that teachers and administrators come to schools with varied and complex experiences, feelings, attitudes, and aptitudes around the use of evidence and particularly around the use of quantitative metrics. What is under investigation here is the intersection/interface between certain types of testing data, teachers' access of that data, and some relevant attitudes that teachers hold. It is hoped that investigating these factors can lead to improved supports for teachers in accessing and using student performance data.

In *Study 1: Exploring Teachers' Online Use of Student Testing Data*, I will examine software usage logs with approaches developed by the learning analytics community to better understand the patterns of system usage practiced by teachers within an online platform for student testing data. I will summarize and visualize usage logs using a variety of descriptive techniques in order to address the question:

(R1**)** To what extent and in what ways do teachers use online data and assessment tools?

Study 2: Connecting Teacher Data Use Attitudes and Efficacy to Online Data Use will explore relationships between teachers' online use of student testing data and their self-reported perspectives on data use, including their general attitudes towards data use, self-efficacy regarding data use, and related aspects of teaching self-efficacy. Study 2 will address these research questions:

(R2) What attitudes do teachers have towards data use?

(R3) To what extent are teachers' data use attitudes, technology acceptance, self-efficacy, and roles related to their online use of student testing data?

While limiting investigation to one urban secondary school hinders the generalizability of these studies, that same limitation comes with the opportunity to describe and explore teachers' online usage and attitudes towards data within a rich, school-level context, often unavailable in larger-scale studies. The goals of this study are both exploratory and applied: to inform future studies by describing the variability within one school and, in the process, to develop practical methods to inform school-based decisions.

Research Context

Study 1 and Study 2 are based on log file data and teacher surveys from an urban secondary school in New York State. In 2015-16, the most recent school year with available data, Progress Secondary School (a pseudonym) served approximately 500 students in the 6th through 12th grade and was divided into a middle and high school with separate principals. The demographic makeup of the student body, divided between middle (grades 6-8) and high school (grades 9-12), is represented in Table 1 below. As indicated, Progress Secondary served a high-need population of students across several demographic areas.

Table 1

Student Demographics at Progress Secondary

	School Level		
	Middle School n= ~250	High School n= ~250	Whole School n= ~500
Student Demographic	%	%	%
Male	~60	~60	~60
Female	~40	~40	~40
African-American	~50	~60	~60
Hispanic	~50	~40	~40
Asian	~0	~0	~0
White	~0	~0	~0
English Language Learner	~10	~0	~10
Economically Disadvantaged	~90	~80	~90

The middle and high schools at Progress Secondary utilized separate assessments and assessment schedules, as well as different methods for organizing class schedules, teacher meetings, and teacher assignments to classes. Some of these relevant differences are described below.

Middle school structure and assessment. Each grade in the middle school was divided into homeroom groupings of students who traveled together to classes for most of the school day. Students attended core classes in double periods, two to four times a week, with each single period lasting approximately fifty minutes. To accommodate the needs of Special Education students, some classes were taught by co-teachers, with Learning Specialists providing additional support to individuals and small groups of students. Students also attended classes in physical education, the arts, and Spanish. Each week several periods were set aside for a structured program in independent reading, as well as for intentionally-grouped remediation programs in math and reading.

Teachers were organized into grade-level teams, which met multiple times a week for planning around whole-class and individual student instruction. For the most part, teachers in the middle school taught within one grade, with the same team of ELA, Math, Science, and Social Studies teachers serving the students of one grade. Several teachers were newly-hired at the beginning of the 2015-16 school year, and several teaching positions in the middle school were held by long-term substitutes throughout the year.

The middle school completed three interim testing sessions during the 2015-16 school year, with one grade-level test in English Language Arts (ELA) and another in math for each session. Each testing session included a different mix of multiple choice and constructed response items drawn from past state tests and item banks. Additionally, the middle school administered weekly progress monitoring quizzes in ELA and a computer adaptive assessment in reading and math three times over the course of the year.

High school structure and assessment. The Progress Secondary High School was organized along more traditional lines, with classes meeting one period (approximately fifty minutes) a day, five days a week. Most high school classes were taught by a single teacher, with some classes co-taught to accommodate the needs of Special Education students. An additional academic advisory meeting was held once a week for high school students. High school teachers were organized into both content-area and grade teams, both of which met weekly for planning purposes.

The high school administered four sessions of interim testing for students enrolled in classes which culminated in a New York State Regents exam. Additional schoolwide assessments included an informal reading inventory administered at the beginning of the school year and the same computer adaptive assessment used by the middle school, but only administered once at the end of the year.

Student data systems. School data at Progress Secondary was housed in separate online systems, which made it difficult for teachers to access some assessment data or to easily combine data from different systems for analysis. The Student Information System (SIS) itself was also extremely siloed, to the degree that basic data categories like discipline records and grades could not be combined for analysis. Additional data was stored and managed in multiple systems, including separate online systems for computer adaptive testing, interim assessments, independent reading assessment, special education data and services, and behavior merits and demerits. Additional student- and class-level data were stored and managed within the school's

Google Apps for Education system, on a variety of ad hoc spreadsheets and documents. Other limitations on the school's presentation of data included lack of usability for teachers, lack of flexibility to visualize data, and lack of a usable system for easily communicating reports to teachers.

Student data and analytics platform. In response to these limitations in data integration, Progress Secondary purchased an online data and assessment platform at the beginning of the 2015-16 school year and implemented the platform both as a warehouse for student data and as a platform for the creation and administration of common assessments, including the viewing and analysis of assessment results by teachers. Multiple types of student data were exported from their separate online systems and imported into the online platform, whose import capabilities, database structure, and stronger reporting capabilities allowed for greater flexibility for administration and staff in the viewing and analysis of aggregate and individual student data. The current studies were conducted in tandem with the implementation of this online data platform during the 2015-16 school year.

Literature Review

In this review of relevant research, I will summarize related issues from the literature of data-based and data-driven decision making (DBDM), suggest areas where further research is needed, and discuss how theories of self-efficacy, technology use, and learning analytics methods address these needs.

DBDM is a complex movement in education often traced back to the No Child Left Behind Act of 2001, though prior to the terminology of "data-driven" or "data-based," similar strands of research focused on the intersection of information, evaluation, and decision-making (Natriello, 1987; Riehl, Pallas, and Natriello, 1991). DBDM has been defined as "systematically analyzing existing data sources within the school, applying outcomes of analyses to innovate teaching, curricula, and school performance, and implementing and evaluating these innovations" (Schildkamp and Kuiper, 2010, p. 482). Though DBDM has become a widespread movement, and a large majority of teachers value data for understanding their students and differentiating instruction, a similar majority are also unsatisfied with the data they have access to and the tools they use to access it (Bill & Melinda Gates Foundation, 2015).

The next several sections will first provide a bird's eye view, by reviewing several important models and frameworks of DBDM, then move into a review of the literature around teachers' attitudes towards and practices with data, and finally wrap up with a section summarizing findings related to DBDM outcomes for both teachers and students. When considering teachers' practices related to data use, I will devote considerable space to discussing studies related to teachers' interactions with online data and assessment.

Models and Frameworks of Teacher Data Usage

In discussing the development of conceptual frameworks for DBDM, I will review theories of change used to connect data use strategies to results, process models (Nilsen, 2015), such as the classic data use cycle (Means, Padilla, and Gallagher, 2010) used to guide the process of translating evidence into action, and finally, some of the initial determinant frameworks created for DBDM. Determinant frameworks, in this sense, attempt to identify and organize the various barriers and enablers to implementation (Nilsen, 2015) of teachers' use of data for instruction. While often dependent on classical theories of individual behavior, such as Social Cognitive Theory (Bandura, 1986) or The Theory of Planned Behavior (Fishbein and Ajzen, 2010), determinant frameworks identify barriers and enablers to adoption, but without positing causality or offering an explanatory mechanism. Such frameworks are commonly used in applied fields, such as medicine, where they have found broad use in implementation projects (Nilsen, 2015). While some efforts have been made in the DBDM literature to organize and test the factors related to teachers' use of data, this dissertation attempts to further that understanding.

DBDM theories of change and process models. Central to understanding and effectively implementing DBDM is articulating a clear theory of change that connects actions such as data collection, data analysis, and instructional decision-making to outcomes for students. As recently as 2016, a special issue of Teaching and Teacher Education, saw several studies struggling to articulate clear logic model for the inputs and outputs of DBDM. As Mandinach and Jimerson (2016) have noted, the lack of a clear logic model can impact classroom and school practice.

Table 2 (below) summarizes some of the models which researchers have used to describe the basic stages of DBDM. Where possible, I have aligned parallel stages between the three models and used arrows to indicate any missing stages. While these models were clearly intended as simple theories of change and not more comprehensive program models, it is still interesting to note their differences, including which elements they include as central to a DBDM process.

Table 2

Comparison of Theories of Change for Data-Based Decision Making

Mandinach and Jimerson (2016)	Poortman and Schildkamp (2016)	Tyler (2013)
Teachers are trained to use data.		Teachers can alter practice based on analysis of student testing data.
		Students are tested to gather performance information.
		Teachers are given test results in a way that fosters meaningful analysis.
	Teachers use data to determine the learning needs of their students.	Teachers access and analyze test data, drawing knowledge from analysis that informs practice.
There will be an impact on classroom practice.	Teachers adapt their instruction according to identified learning needs.	Teachers act on new knowledge and alter classroom practice.
These practices will lead to increased student performance.	These changes lead to increased student learning and achievement	Altered practice has a positive impact on student achievement.

Note. These logic models have been adapted to table form from the text of the articles cited.

As Table 2 illustrates, even between these few basic logic models there are points of difference. While all models end with improved student performance, Poortman and Schildkamp (2016) begin with teachers' use of data and specifically reference the diagnosis of learning needs, while Mandinach and Jimerson begin with teacher training on data use and describe a general “impact on classroom practice” that follows from this training. More elaborately, and more in keeping with the data-use cycles described later in this section, Tyler (2013) outlines a logic model that includes additional steps and assumptions, particularly in regard to which data are used (testing data) and the steps that teachers’ move through to make use of data: receive data, access them, analyze them, connect analysis to practice, and act on their knowledge.

Along with these more succinct theories of change, DBDM researchers and practitioners have laid out several multi-stage process models, models which have both described data use—based on qualitative studies of school data use practice—and prescribed approaches to teachers’ data use for use in professional development.

Mandinach and Jackson (2012) review several major process models of DBDM, including in their review only those models supported by research. These process models tend to promote a cyclical pattern of usage and to collapse the complicated steps in generating usable knowledge from raw data into broad descriptions, such as “analyze data” (Means, Padilla, and Gallagher, 2010) or “collect and prepare student learning data” (Hamilton et al., 2009). The usage framework from Means, Padilla, and Gallagher (2010), developed through examination of national surveys and case studies, is shown in Figure 1 below as one example of a cyclical framework. Other, similar models, while maintaining a cyclic structure, differ in terms of the content and number of stages included as key components of data use.

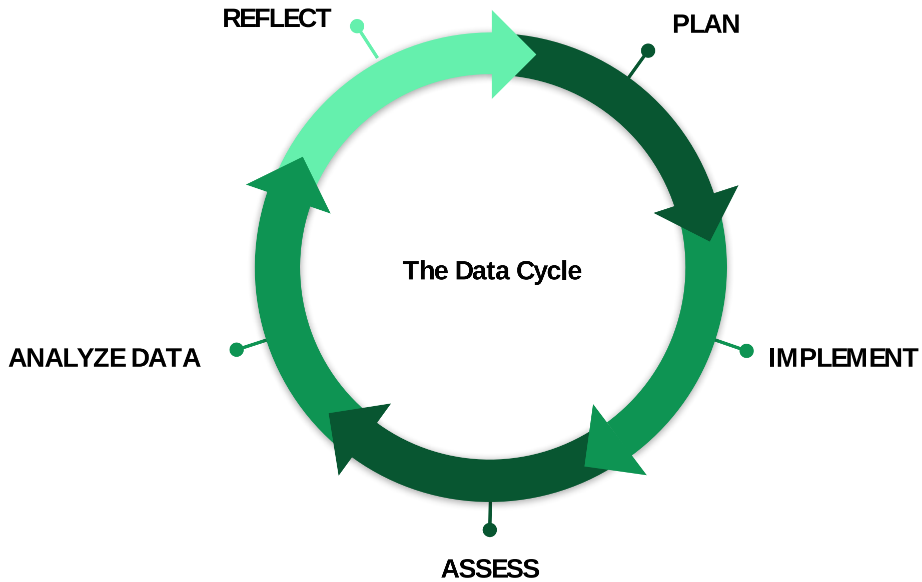


Figure 1. The data cycle, adapted from Means, Padilla, and Gallagher, 2010.

This data cycle from Means et al. (2010) is one of many, and Table 3 below summarizes several data use cycles in table format for easier comparison of their key components. The Plan Do Study Act (PDSA) Cycle, attributed to Walter Shewhart and Edwards Deming (Moen and Norman, 2010) and hugely influential in business improvement process, is included as an industry-related precursor to DBDM cycles.

Table 3

Comparison of Data Use Cycles

Row Label	PDSA Cycle (Moen and Norman, 2010)	Conceptual Framework for school and classroom evaluation (Natriello, 1987)	Conceptual Framework for DDDM (Mandinach et al., 2008)	Abbott's Framework of Improvement and Readiness (Abbott et al., 2008)	IES Practice Guide Cycle (Hamilton et al., 2009)	Data Cycle (Means et al., 2010)
A	Plan: Design, abandon, revise program components	Establish purposes for evaluating students Set criteria and standards for performance	Decide	Plan	Modify instruction to test hypotheses	Plan
B	Do: Implement changes	Assign tasks to students	Implement	Implement		Implement
C	Measure effects of change	Sample information on student performance	Collect and organize data		Collect and Prepare student data	Assess
D	Study: Summarize measurements	Appraise student performance	Summarize and Analyze Information	Evaluate data	Interpret data	Analyze data
E	Report results	Provide feedback to students and others				
F	Act: Decide on changes needed for improvement	Monitor outcomes of student evaluation	Prioritize and Synthesize Knowledge	Internalize data Collaborate Assess	Develop hypotheses to improve student performance	Reflect

As seen above, the stages of the PDSA cycle map onto DBDM cycles, with some of its elements providing useful context for understanding possible omissions in educational data use cycles, particularly the exclusion in DBDM cycles of "reporting of results" (in row E) as a distinct stage. Some of the distinctions in Deming's PDSA cycle between "Acting," "Planning," and "Doing," differ from some later DBDM process models as well. In Deming's model, for example, "Actions" are not actual interventions with students, but instead evidence-based decisions about necessary high-level changes. "Planning" in the PDSA model refers to decisions for change in specific program components, and "Doing" refers to the actual implementation of these decisions.

While extensive common ground exists among conceptual frameworks for DBDM, Table 3 also illustrates that key components of the DBDM process vary between models. Abbot's framework (Abbott et al., 2008), for example, introduces collaboration as a component in row F, while Natriello (1987), writing from an evaluation perspective prior to more recent trends in DBDM literature, incorporates a valuable breakdown of several specific decisions and actions not included in later DBDM models, specifically by identifying stages for establishing the purpose of evaluation and setting criteria and standards for student performance.

As an alternative to strictly cyclical frameworks, Mandinach et al. (2008) present an additional framework based on six district-level case studies. In contrast to other DBDM frameworks, the Mandinach et al. (2008) framework (Figure 2) connects DBDM to information theory on the transformation of data to information and to knowledge (Ackoff, 1989; Nunamaker, Romano, and Briggs, 2001). The model includes stages of movement, from (a) the organization and collection of data to (b) the analysis and summary of that data as information, to (c) the synthesis of information as knowledge. Only after data is converted to knowledge does the model proceed to decision, implementation, and finally, to assessing the impact of the DBDM process.

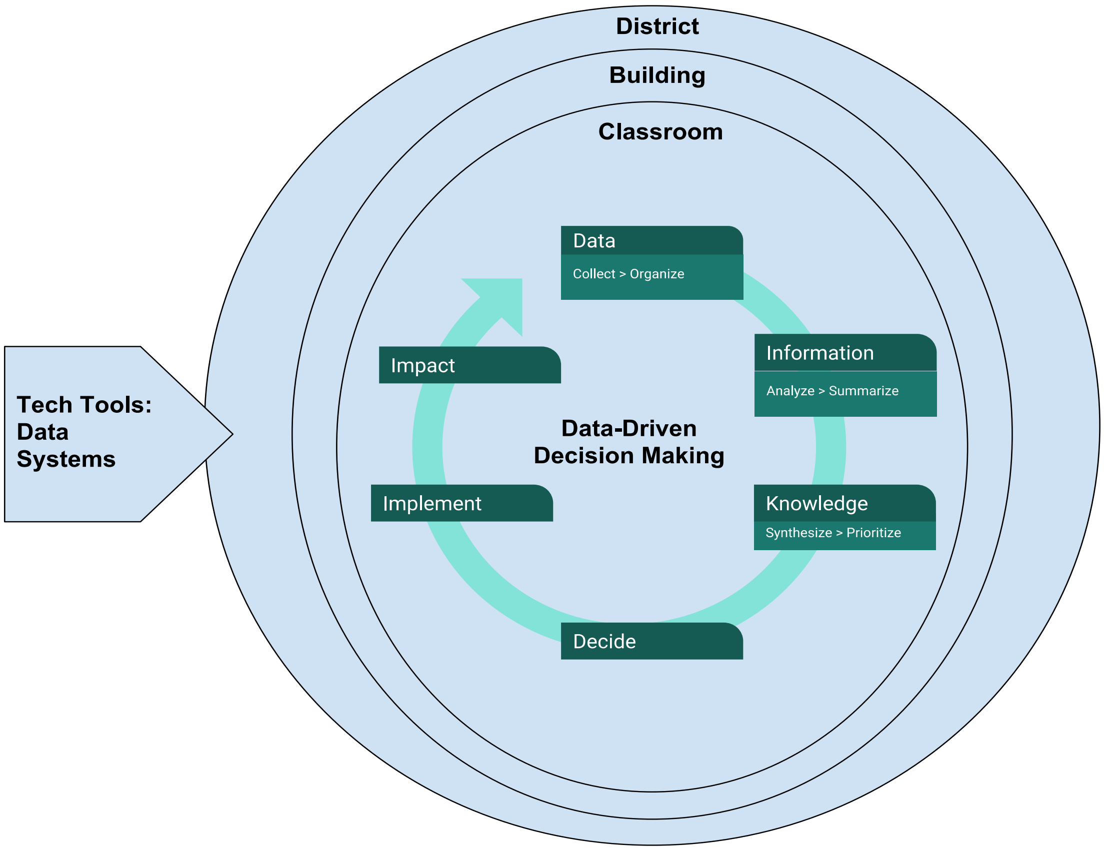


Figure 2. Conceptual framework for data-driven decision making. Adapted from Mandinach and Jackson (2012).

Building on these generalized process models for data use, researchers have begun to flesh out some of the key dynamics that impact adoption across the stages of data use. Marsh (2012, p. 4), for example, builds on Mandinach et al.'s (2008) conceptual framework by proposing a classification of "leverage points" where support, or lack of support, may impact the success of data use implementation.

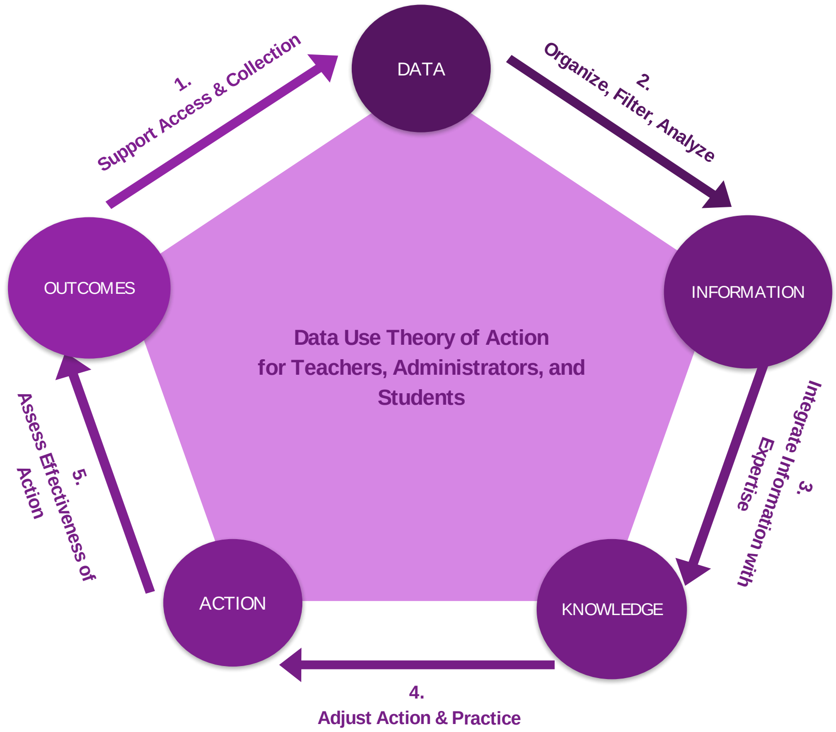


Figure 3. Marsh's data use theory of action. Adapted from Marsh (2012).

Figure 3 illustrates both the overall progression from data to action, as well as numbered points where interventions may fail or flourish. As Marsh points out, school or district actions meant to impact DBDM may target one or more of these leverage points in the form of supports, norms, incentives, or even mandates (2012). While Marsh (2012) uses these intervention points as a schema for categorizing data use interventions, their presence also points to an underlying set of barriers, unaccounted for in previous process models. Some of these critical implementation factors include an organization's capacity for support to teachers, the characteristics of the data in question, the role of leadership, the availability of time for data use, the policy and curriculum context, the role of interpersonal relationships, and the importance of the data user's pre-existing beliefs, skills, and knowledge in regard to specific components of data use.

Marsh (2012) laid the groundwork for later exploration and clarification of which factors represent the difference between success or failure of DBDM practice. A prime example of such factors would be the impact of the varied epistemologies, attitudes, and mental models that teachers bring to their profession and with which any DBDM intervention or process must interact. As Senge suggests, a thorough understanding of personal mental models is critical to an organization's process of alignment and unified movement (2006). While teachers may be receptive in the abstract to data-based initiatives, a range of individual beliefs about the validity of assessment, their trust in administration, and their perceptions of quantification and evaluation may impact any specific cases of implementation.

DBDM determinant frameworks. While data use has been encouraged from the federal down to the school level, multiple reviews and studies have drawn attention to the many obstacles impacting educators' use of student data at the classroom level. Qualitative studies, for example, have highlighted factors that hinder or encourage data use, factors such as time available to use and discuss student data, data literacy, and timely and relevant streams of data (Hoogland et al., 2016). Synthesizing these qualitative efforts, several studies have developed determinant frameworks of the factors most commonly identified to encourage and inhibit DBDM (Datnow and Hubbard, 2015; Hoogland et al., 2016; Prenger and Schildkamp, 2018; Schildkamp and Kuiper, 2010; Schildkamp, Poortman, Luyten, and Ebbeler, 2017).

As an example of this trend, Datnow and Hubbard's (2015) literature review extends the work of Marsh (2012) in identifying several categories of factors that facilitate or inhibit teachers' data use and organizing these determinants along a continuum from individual teacher beliefs to school-level practice to district accountability demands. Table 4 summarizes these factors drawn from Datnow and Hubbard (2015).

Table 4

Factors that Inhibit and Encourage Teachers' Data Use

Organizational Category	Relevant factors	Facilitating Factors	Inhibiting Factors
District Accountability Status	Accountability ranking	Examining schoolwide issues and opportunities for growth	Focusing efforts on students at the cusp of proficiency
School Leadership	Schoolwide Tone, Support, Policies, and Structures for Discourse	Promoting thoughtful use of data to inform action Using organizational learning to understand, not just identify problems	Promoting that data in and of themselves drive action Emphasizing accountability for short-term, compliance monitoring
Organizational Context	Structured Time for Collaboration	Strong instructional communities Grade-level agendas, norms, and expertise Protocols for analysis/reflection Urgency for student progress Multiple sources of data	Implementing multiple instructional initiatives at the same time Limiting teachers' ability to fully integrate data usage into their practice
Teacher Capacity for Data Usage	Training	Competence in data analysis and interpretation Confidence in using data to inform instruction	Lack of training in understanding and using data Lack of training in assessment Lack of confidence in data use and statistics
Teacher Beliefs	Pedagogical orientation (Remesal 2011) Assessment for learning vs. Assessment for accountability	Assessment improves learning and teaching	Assessment makes students accountable Data are punitive and derived from invalid assessments Assessments are unfair to teachers Changing student performance is beyond teacher control

Note. This table summarizes content originally presented in paragraph form in Datnow and Hubbard (2015b).

Another strand of work contributing to determinant frameworks for DBDM has been conducted by Schildkamp and colleagues in the Netherlands. Schildkamp and Kuiper (2010) develop their framework through literature review, as well as interviews and documents collected from six best-practice schools. Their schema identifies school organizational characteristics, data characteristics, data use, and user characteristics as primary categories of factors impacting data use. These factors contain significant overlap with those identified in Datnow and Hubbard (2015), along with some additional categories acknowledging the impact of data characteristics, such as accuracy and timeliness, and of data uses, whether instructional, policy-related, or accountability-focused.

Hoogland et al (2016), through systematic literature review, and Schildkamp et al (2014), through international comparative case studies extend this work of identifying the determinants of data use, while recognizing that different uses of data may entail different sets of determinants. Schildkamp et al (2017) explores these same questions, using a large-scale survey to test the significance of organizational, data, and user characteristics against teachers reported use of data for separate purposes of accountability, school development, and instructional. Figure 4 summarizes the determinant framework investigated in Schildkamp et al. (2017).

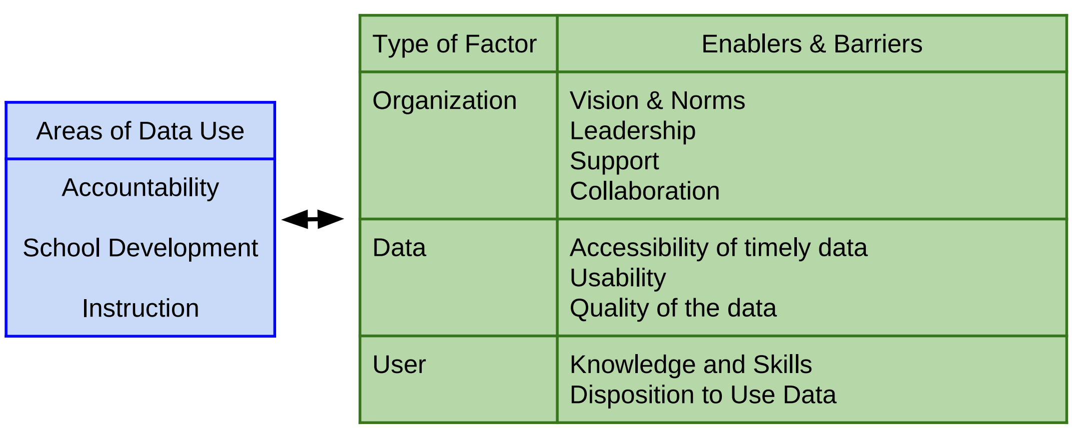


Figure 4. Data use enablers and barriers. Adapted from Schildkamp et al., 2017

In the case of instructional use, Schildkamp et al. (2017) found that a school's organizational characteristics, user characteristics, and collaboration were all significantly related to instructional data use ($p < .001$), with the model explaining 19% of the variance in teachers' reported data use. This 19% of the total variance accounted for 16% of the variance at the teacher-level and 27% at the school level. While a larger proportion of the between-school variance was explained, the variance at that level accounted for only 4% of the total variation (ICC) in instructional data use, while the level 1 variation between teachers (with a large majority unexplained) accounted for 96% of the total variability in instructional data use.

While Schildkamp and colleagues (Schildkamp and Kuiper, 2010; Schildkamp et al., 2014; Hoogland et al., 2016; Schildkamp et al., 2017) have followed a rigorous program for developing schema and testing factors related to data use, the sources of variance in data use between teachers remains largely unexplained. Additional studies, employing hierarchical methods to study teachers' online use of data (Shaw, 2010, Tyler, 2013; Wayman et al., 2017) have found similar proportions of school- vs. teacher-level variance, with differences between teachers accounting for most of the variation in online use of student data. These studies of online data use have struggled to identify teacher-level determinants impacting this variation. Online data use studies will be discussed more fully in an upcoming section on teachers' practices related to data use. The next section, however, will turn to the literature on teachers' attitudes towards data, an area closely tied to better understanding and explaining the large variations in teachers' DBDM practice.

Teacher Attitudes Toward Data Use

Teacher attitudes towards the use of student data have been explored in multiple qualitative and quantitative studies with a variety of instruments. While many of these studies have been case studies of the relationships between teacher attitudes towards data use and teachers' data use behaviors (Datnow, Park, and Kennedy-Lewis, 2012; Hoogland et al., 2016; Jimerson, 2014), others have tackled large-scale survey projects across whole districts (Wayman et al., 2009).

Surveys into teachers' data use attitudes range from theory-driven surveys to pragmatic tools that help schools and districts capture practitioners' perspectives on teacher and staff data use. The Teacher Data Use Survey (TDUS), developed by Wayman, Wilkerson, Cho, Mandinach, and Supovitz (2016) is an example of such a pragmatic tool, along with its forerunner, the Survey of Educator Data Usage (SEDU) (Wayman, Cho, and Shaw, 2009b), both of which have been used in several large-scale investigations of educator data usage. Along with multiple subscales exploring educators' reported frequency of data usage, the SEDU includes subscales on Attitudes Toward Data, Data Use Practice, Computer Data Systems, Data's Effectiveness for Pedagogy, District Vision, Principal Leadership, Support for Data Use, Instructional Resources, and Time to Use Data (Wayman, Snodgrass Rangel, Jimerson, and Cho, 2010).

In terms of overall teacher attitudes towards data use, Wayman et al.(2009a) found that educators reported more positive responses on scales measuring their perceptions of data's overall effectiveness for pedagogy (M=3.47) and a district's vision for data use (M=3.49). Educators were less positive about principal leadership (3.22), their personal attitudes toward data use (M=3.19), computer systems (M=3.05), and the support they received for data use (M=2.99), and they gave their most negative responses when asked to evaluate the amount and quality of time allocated for teacher data use (M=2.34).

Teachers expressed some of the following general attitudes and concerns about data usage (Wayman et al., 2009a). While not all teachers shared these concerns, they were mentioned frequently enough that the authors included them as general patterns in teachers' attitude towards data use:

- Teachers like getting more information about their students. They want to use data.
- Practical challenges in accessing and using data undermine the desire to use data.
- Teachers do not have enough time to use data.
- Some teachers believe that they already know what the data can tell them about their students, making analysis a waste of time and energy.
- Teacher judgement should be trusted over data analysis. When data take precedence over teacher judgement, teachers find this frustrating.
- Teachers do not trust how data is used in the school
- Teachers feel like data use is imposed on them
- Regarding interim assessments, teachers are concerned when students are tested on content that has not yet been taught or when teachers are not integrated into the writing of interim tests.
- Teachers believe that data does not capture how their students are learning.
- Teachers do not believe that students should be thought of as numbers.

Along with these general attitudes towards data usage, teacher attitudes towards data support and computer systems play a critical role in their response to data use. Accessing disparate computer systems is a ubiquitous obstacle to data use, with Wayman et al. (2009a) reporting over 58 separate systems in the district surveyed. Lack of integration among these various systems is seen as a major cause of frustration and a major obstacle to data access and analysis. Some sources of data, particularly special education and English Language Learner data may not even be available to teachers in online systems and require extensive manual copying to be accessible for practical use.

As data use in schools is a complex and multifaceted construct, so are teacher and administrator attitudes towards that use. While there is some evidence that teachers are sympathetic or even enthusiastic about data use in the abstract, there is also evidence that this enthusiasm wanes the closer data use comes to the particulars of a teacher's experience, becoming subject to the constraints of school politics, technology, training, and, most importantly, time (Wayman et al., 2009).

Teacher Data Use Practices

What Data do Educators Use? Teachers access a great variety of student data in their daily practice, everything from excused and unexcused absences, to discipline referrals, student surveys, student work, student grades, and multiple forms of assessment (Mandinach and Jackson, 2012). Student Information Systems (SIS) alone, without the overlay of assessment or other analytics systems, generate large amounts of student data in the form of teachers' records of grades and attendance. Nevertheless, the literature of DBDM often focuses on the use of assessment and testing data, as in Datnow and Hubbard's (2015) review of teachers' use of assessment data, which indicates that the predominant form of assessment consulted in DBDM is interim or benchmark assessments, defined as those assessments that "evaluate students' knowledge and skills in a limited time frame and can be easily aggregated across schools and classrooms" (Datnow and Hubbard, 2015, p. 3). Interim assessments are often given at least three times a year and provide a framework for schools to measure progress towards students' mastery of standards in the context of a state-level testing program. Such assessments have become a common part of the school year and teachers are often expected to analyze the results of interim assessments and use them to plan for future instruction (Datnow and Hubbard, 2015).

The types of data accessed can also depend on its intended use. While interim assessments may play a greater role in planning schoolwide or grade-level interventions, classroom level observations or formative assessments can play a larger part in determining class-level interventions, such as student grouping. Leveled reading inventories or conferencing may also impact classroom groupings or even individual interventions or tracking, where such tools can help determine the appropriate level text for independent reading or online remediation work in math (Park and Datnow, 2017).

The types of data that teachers access and their uses for that data can also vary according to organizational factors, such as school level. Wayman et al. (2010) reports that elementary and secondary teachers differed in their interview responses, with elementary teachers much more likely to discuss the use of data for re-grouping students according to skill or level.

How do Educators Use Data for Instructional Purposes? As discussed earlier, data can serve multiple purposes, aiding analyses related to accountability, school development, and instruction. Research suggests that teachers still struggle with the use of data in the classroom context (Hoogland et al., 2016; Schildkamp and Kuiper, 2010). Unfortunately, while some studies have gathered direct observations on how teachers alter instruction based on assessment data, most have been based on self-report. Hoover and Abrams (2013) found, for example, that 96% of teachers surveyed reported differentiating instruction for remediation, 94% reported re-teaching, and 92% reported changing the pace of instruction. At the same time, however, 64% of teachers reported that school pacing prevented re-teaching. Wayman et al. (2009a) reported a similar inconsistency, where teacher survey responses indicated the practice of reworking instruction to meet the specific needs of students, while interview and focus group responses rarely referred to using data to alter classroom practice.

Even with these difficulties in capturing clear evidence of teacher's actual data use practice, several major themes emerge.

Limited Use. While it is widely believed that access to interim testing results and other student data has the potential to improve student outcomes, this improvement is directly mediated by educators' usage of data to guide the delivery of instruction (Bulkley, Christman, Goertz, and Lawrence, 2010). A survey of teachers in Virginia found that while teachers accessed data from a wide range of assessments, they administered these assessments more frequently than they analyzed them and their analysis was often limited, focusing on average performance (Hoover and Abrams, 2013).

Identifying struggling students. Data use efforts often focus on identifying struggling students and creating plans for remediation (Wayman, 2009a). In many cases, interventions for struggling students may be aimed at so-called "Bubble" students, those immediately below proficiency on state tests (Datnow and Hubbard, 2015).

Re-Teaching: more, but the same. One common approach to data use appears to be to alter the timing of instructional content in response to evidence, while not altering instructional methods. In other words, instructors often revise the topics of lessons, but not the way they teach those topics (Marsh, 2012). Some common patterns of intervention identified by Datnow and Hubbard (2015) include:

- Re-teaching, both on teacher- and school-planned timelines
- Re-teaching based on content or skill for which large number of students are below proficiency

- Re-teaching to lowest performing students (Though assessment to monitor the effects of re-teaching was rare.)

Grouping. Creating short-term or long-term homogenous student groups by proficiency levels in some content/skill area. (Datnow and Hubbard, 2015).

Function follows form. There are many indications that teachers' use of assessment data depends on the types of tasks included in assessments. Open-ended student assessments have fostered more dialogue and collaboration among teachers, while multiple choice assessments have led more often to decisions about student placement (Davidson and Frohbieter, 2011).

Additionally, the usefulness of assessments can be limited by the frequency and purpose of their collection. Formative interim tests, for example, are often administered every nine-weeks, limiting how frequently teachers can use their results to inform instruction. The goals of formative interim testing can also constrain their use: the alignment of such tests to grade-level standards, making grade-wide comparison easier, can severely limit the ability of such tests to capture the performance of students on below grade-level standards.

The investment of teachers in assessment results may also be closely linked to assessments' integration with classroom instruction. In Kerr et al.'s study of data usage (2006), teachers found review of student work and classroom-level assessments more useful for classroom-level planning, while analysis of large scale assessments was seen as more useful for schoolwide planning (Rogosa, 2005; J. Supovitz, 2009).

Typologies of Teacher Data Use. Another approach to understanding teachers' data use is that taken by the Gates foundation in their 2015 study (Bill & Melinda Gates Foundation, 2015). Having extensively surveyed, interviewed, and observed teachers in both traditional and tech-forward schools, the study identifies six subgroups of teacher behavior in regard to the use of student data and technology: Data Mavens, Growth Seekers, Aspirational Users, Scorekeepers, Perceptives, and Traditionalists. Data Mavens and Growth Seekers are both high frequency and high competency users of data, with Growth Seekers focused additionally on using student data as a means of improving their own growth as teachers. Aspirational Users believe in the use of data but can be overwhelmed in its application. Scorekeepers focus on the use of data for testing purpose. Perceptives and Traditionalists both use data less often in their practice, with Perceptives depending more on observation of students and Traditionalists on students' grades. Unfortunately, the Gates foundation methodology does not identify the particular survey constructs or observations that might differentiate between the data use types.

While not strictly related to data use, another recently constructed typology regarding teacher use of technology may be relevant to better understanding teachers' use of data, especially given the integration of online technology with instructional data use. Applying latent class analysis to nationally-generalizable teacher survey data, Graves and Bowers (2018) identify four statistically-significant subgroups of teacher technology use. Dexterous teachers are frequent and versatile users of technology, as opposed to Evaders who limit their use of technology even for basic tasks. Two additional categories tend towards specialized uses of technology: Presenters focus on technology as an aid in lectures or for student presentations, while Assessors implement instructional technology for students' repeated practice of basic skills in math and reading.

Teacher Usage of Online Data Systems. While extensive work in the form of surveys, interviews, and focus groups has attempted to clarify how teachers access student data, one of the most direct ways to better understand that access is to observe the interactions that educators have with the many online systems that collect and distribute student data. While analyzing such data fails to capture how teachers use other sources of data and clearly does not track the uses to which data is put in the classroom, it does provide an opportunity to naturally observe and describe an important subset of teachers' interactions with student data and to explore how other teacher factors may be connected to those interactions. So far, despite a growing demand for online student data and assessment systems, only a very few studies have described and analyzed how they are used by teachers. More such studies that provide rich description of how teachers use such systems on a daily basis may aid their developers in innovating and adapting their products in ways that prioritize and maximize educators' capacity for responsive and effective decision-making with students.

Initial forays into the analysis of teacher interaction with online assessment data were conducted by Wayman et al. (2009a; 2011), Shaw and Wayman (2012), and Shaw (2010) in a series studies of the Mesa Public School (MPS) district, related to the district's implementation of the Acuity formative testing system published by CTB/McGraw. The collection of online log files from teachers' weekly use of Acuity was only one facet of an evaluation which also included educator surveys, focus groups, and user observations. Wayman et al. (2009) focused on three main measures of Acuity use: (a) whether teachers used/didn't use the system; (b) the *prevalence* of use, or the total user actions for each teacher; and (c) the *consistency* of use, or the number of weeks that each teacher used the system. In addition, the Year One Study of Acuity use breaks down teacher usage by sub-functions within the Acuity system, such as Instructional Resources, Reports, and Custom Test functions (Wayman et al., 2009).

Soon after, building on the study of Acuity use in Mesa Public Schools, Tyler (2013) studied teachers' online use of a benchmark testing portal implemented in Cincinnati Public Schools. Tyler attempts to answer questions similar to those of Wayman as to how much and how often teachers used an online system for student data, as well as questions about the use of specific pages and categories of student data. While the earlier studies of Wayman et al. (2009) were able to capture counts of teacher log-ins to the Acuity system, Tyler (2013), working more directly with the log files of teacher access, was able to capture the actual time spent by teachers using the system, calculating both total yearly metrics of usage, as well as the weekly time that teachers spent accessing the system. The teachers under consideration in Tyler (2013) were core subject area teachers of 3^{rd} through 8^{th} graders, the group of teachers most impacted by interim benchmark testing throughout the school year and, therefore, most likely to use the program.

While Wayman et al. (2009) and Tyler (2013) both analyzed systems that provided access primarily to benchmark interim testing data, Gold, Lent, Cole, Kemple, Nathanson, and Brand (2012) analyzed educators' use of a second type of online system, one used to aggregate student data from multiple sources. In such systems, students' overall scores on interim tests may be only one of several types of student data, reported alongside information on student attendance, grades, and discipline. Such systems attempt to address teacher concerns that different types of student data are housed in too many disparate locations. Gold et al. (2012) evaluated one such attempt at data aggregation in their study of the ARIS system implemented in New York City schools in 2008 (Gold et al., 2012a, 2012b). While some types of ARIS data, including biographical, attendance, state test results, and transcripts, differed from those analyzed in Wayman et al. (2009) and Tyler (2013), the basic questions of access remain the same: "to what extent and in what ways" do educators use the data system in question. While following some of the precedents of Tyler (2013) for the analysis of log files, the ARIS evaluation study adds reports of distinct user types within the system, distinguishing between school educators, school principals, data specialists, and inquiry team members, among other user categories, as well as analyzing usage patterns for light, moderate, and heavy users. Heavy users, for instance, made up only 28% of users but accounted for over 80% of time spent in the system (Gold et al., 2012).

Analyzing a similar data warehouse system to Gold et al. (2012), but from the perspective of a different set of users, Drake (2015) applies Latent Class Growth Analysis (LCGA) to principals' usage patterns in accessing an online data warehouse over the course of a year. Using counts of access (as opposed to duration) as the basic unit of analysis, Drake estimates a three-class Latent Class Growth Model using principals' cumulative report access over the course of the year. The relative risk ratios for the latent classes are then calculated for a variety of principal covariates and survey responses. While the log file metrics analyzed were the same as in past studies, the application of LCGA suggests additional methodological possibilities for the analysis of log file data.

The studies described so far have based their analyses on a compact set of log file metrics. These metrics have included counts of access times (Wayman et al., 2009), total time spent in the system, time spent in various subsystems (Tyler, 2013), and the use of user types (Low, Mid, High) to differentiate analysis (Drake, 2015; Gold et al., 2012). These metrics have proven valuable in providing answers to basic questions such as "how?" and "how much?" educators use online data and assessment systems. The ARIS study (Gold et al., 2012), for example, found that only 67% of teachers logged into the system during the 2010-11 school year. For all categories of ARIS users—teachers, administrators, and specialists—the average number of sessions was 21 (with a median of 6), accumulating in an average of 96.7 minutes of use over the course of the year (median use = 32 minutes), with an average session lasting 4.7 minutes (median = 4.2 minutes). More specifically, the average classroom teacher spent 67 minutes using the system over the course of the year, dedicating 19 minutes to viewing individual student profiles, 21 minutes to classroom-level data, 3 minutes to intensive data reports, and 10 minutes to system navigation (Gold et al., 2012a).

Table 5 (below) compares teacher usage metrics across available studies and in doing so draws attention to the lack of viable comparisons of teachers' online use. With these studies analyzing different types of data systems and different populations with different metrics, only a few clear comparisons can be made. For example, the percentage of teachers who used the Acuity system in Mesa Public Schools (Shaw and Wayman, 2012) is comparable to the percentage of teachers accessing the benchmark testing portal in Cincinnati (Tyler, 2013)—93% and 98% respectively. But, beyond that, the metrics used to capture teacher usage of Acuity are not directly comparable to the usage metrics used in Tyler (2013).

In another example, usage metrics in Tyler (2013) and Gold et al. (2012) are more directly comparable, yet these two systems include substantially different types of student data, with the Cincinnati system focused on reporting benchmark assessment data (Tyler, 2013) and the ARIS system in New York City reporting broader information on student grades, attendance, and state testing results (Gold et al., 2012). While the ARIS system appears to have had less usage than the Cincinnati benchmark testing system, their different purposes make any comparisons limited. The inability of the literature to provide even basic descriptive context and comparisons of teacher usage metrics suggests that future studies should attempt to address this lack of description, as well as pay attention to developing and clarifying online usage outcomes for common use.

Table 5

Summary of Teacher Usage Metrics across Studies of Online Data Use

	Shaw and Wayman (2012)			Tyler (2013)	Gold et al. (2012)
Year Data Collected	2008-09	2009-10	2010-11	2008-09	2010-11
% of Teachers Who Used System	70	93	91	98	67
Avg. Annual Sessions	NA	NA	NA	Mean: 33 Median: 28	Mean: 21* Median: 6
Prevalence of Use:					
Avg. Instruct. Actions	40	128	159		
Elementary			171	NA	NA
Junior High Reading			102		
Junior High Math			139		
Consistency:					
Avg. Weeks Used	5	10	9		
Elementary			10	NA	NA
Junior High Reading			4		
Junior High Math			8		
Avg. Hours of Use	NA	NA	NA	Mean: 7 Median: 4	Mean: 0.88
Avg. Weekly Sessions	NA	NA	NA	0.79	NA
Avg. Weekly Duration	NA	NA	NA	10 minutes	NA
Avg. Session Length	NA	NA	NA	NA	4.7 minutes

* Represents all types of users (teachers, administrators, and specialists). All other metrics represent teacher usage alone

Despite the lack of clear comparisons, one common observation across these studies is the skewed distribution of overall use. All three studies (Shaw and Wayman, 2012; Gold et al., 2012; Tyler, 2013) find a strong positive skew, with most users demonstrating low to no usage, while a small group uses the system at much higher rates, projecting a long tail in the positive direction. As a result of these skewed distributions, average usage often exceeds median usage by a substantial amount, creating situations, as in Gold et al. (2012), where 50% of users had six or fewer sessions, yet the average number of sessions per user was 21. With this type of distribution, using the average user as a reference may create an overly optimistic image of use within the system. Similarly, overreliance on metrics calculating the percentage of teachers who access a data system, may paint an overly optimistic picture of actual use, with most teachers accessing the system, but doing so only at a minimum.

School-level vs. teacher-level variability in online data use. With Shaw (2010), Shaw and Wayman (2012), and Tyler (2013) using multi-level models to analyze teachers within schools, some findings are available for understanding the proportion of school-level versus teacher-level variability in online use of benchmark testing systems. For example, Shaw and Wayman (2012) report that in the third-year of Acuity implementation, 23% of the overall variation in teachers' prevalence of use and 35% of their variation in consistency of use was attributable to schools. Adding teachers' years of experience and school's accountability status to the null model failed to explain any school-level variability in teachers' prevalence of use but did reduce between-school variation in consistency from 35% to 29% and explained 27.6% of variance in consistency at the teacher level (Shaw and Wayman, 2012). Shaw (2010), on the other hand, analyzing a subset of data from the larger study of Acuity use found that a much smaller, but still significant 13% of variability (ICC) could be attributed to school-level factors, while 87% of variability in data use was attributable to teacher-level factors. Tyler (2013) found similar results, with only 10% of the unexplained variability in teachers' online access related to between-school variation. These proportions of variance are, in turn, larger than those found in Schildkamp et al. (2017), where school-level variation accounted for only 4% of total variation in instructional data use, as measured by teacher self-report. While results are few, the proportions of school to teacher effects in two data use studies (Shaw 2010, Tyler, 2013) parallels findings of more general school-effects research, where variations at the school level account for 8 to 16% of variance in student achievement, and teacher effects are consistently larger than school effects when both are included in a hierarchical model (Teddlie and Stringfield, 2007). In contrast, two other studies found school effects related to teacher data use to be both lower (Schildkamp et al., 2017) and higher (Shaw and Wayman, 2012) than the range suggested by school effects studies of student achievement. Since qualitative work in data use often highlights the impact of organizational-level determinants of data use (Datnow and Hubbard, 2015b; Schildkamp et al., 2017), more investigation is needed to illuminate how school-level process impacts teachers' actual use of data.

While only Wayman and Shaw (2012) examine the significance of school-level factors in relation to multiple usage outcomes, it is interesting to note that of the two outcomes they analyze—prevalence and consistency—consistency of use (weeks of use/year) was more closely related to school-level factors than was teachers' prevalence of use. This particular finding suggests that different outcome measures of usage may be differentially impacted by teacher- and school-level factors. In this example, where consistency of use was more closely connected to school-level variation, perhaps the impact of a school's meeting and training schedules are more closely aligned to a weekly measure of use, than a measure of total actions or duration, with its greater range and variation. This measure of consistency was, in turn, one of the only usage metrics to produce a significant relationship to student achievement ($p = 0.01$, SD Difference = 0.05), in the form of elementary reading growth (Wayman et al., 2017). With different conceptions of online use implemented in different studies, it is unclear how these differing metrics may have impacted findings of both the determinants and impacts of data use.

As to which school-level factors were significant in explaining variation in teachers' usage, results across available studies are sparse or mixed (see Table 6). Shaw (2010), for example, found that on average a 100-student decrease in enrollment was related to a six-action increase in teacher usage, though this appears to be the only finding related to school size. The significance of schools' Title I and accountability status are mixed across studies, while school level (Elementary vs. Junior High) is consistently significant across studies yet acts in opposite directions. Studies of the Acuity system (Shaw and Wayman, 2012; Wayman et al., 2009, 2011) found teachers' online use significantly higher in the elementary grades (3^{rd} – 6^{th}), while Tyler (2013) found usage significantly higher for teachers of grades 6-8. Even with this switch in direction, the significance of the relationship between school level and online data use is the most well-supported of the factors analyzed. A more contextual understanding of within school processes may bring to light some of the organizational factors complicating school-level analyses.

Along with school-level determinants of teacher data use, the same set of studies examined the teacher-level correlates that might account for variance in online data use. Wayman et al. (2009, 2011) and Shaw and Wayman (2012) relate usage outcomes to teacher demographics, such as years of experience and level of education, as well as to teacher reports of (a) frequency of instructional data use; (b) the general effectiveness of data for pedagogy; and (c) specific attitudes towards the Acuity System in regard to ease of use, dependability, and accuracy. Shaw (2010) examines additional relationships between general teacher efficacy and Acuity use, while Tyler (2013) investigates teacher online usage in relation to students' baseline academic performance and teachers performance on value-added models for student achievement.

Table 6

Summary of School-Level Factors in Studies of Online Teacher Data Use

School-level Factors	Wayman et al. (2009)	Wayman et al. (2011)	Shaw and Wayman (2012)	Shaw (2010)	Tyler (2013)	Gold et al. (2012)
Title I Status	Significant (+)	NA	NA	Not Sig.	NA	NA
% Minority Status	NA	NA	NA	Not Sig.	NA	NA
School Size	NA	NA	NA	Significant (-)	NA	NA
School Level	Significant Elem. Use > Middle	Significant Elem. Use > Middle	Significant Elem. Use > Middle	NA	Significant Middle Use > Elem.	Middle Use > Elem. and High School
Accountability Category	NA	Not Sig.	Significant* (+)	Not Sig.	NA	NA

* Consistency of use (but not prevalence) increased significantly with better Accountability Status (SD Difference from -0.51 to -0.99 with reference to the highest accountability category)

Of these possible teacher-level determinants of online use, some factors were found to be significantly related, though none consistently. Studies by Wayman et al (2009, 2011) and Shaw and Wayman (2012) indicate both significance and non-significance for teachers' years of experience. On the closely related metric of level of education, teachers with at least a Master's degree averaged significantly more online actions (13 more actions) over the course of the year than teachers with only a Bachelor's degree (Wayman et al., 2009). Teachers' self-report on the instructional use of data in the classroom was also significant, with every one-point increase in reported frequency of instructional data use associated with an average increase of 18 system actions. Shaw (2010) found an inverse relationship between general teacher efficacy and Acuity use, with a one standard deviation increase in teacher self-efficacy associated with seven fewer actions in Acuity over the course of the year.

In terms of teachers' content area, Gold et al. (2012) describes Math teachers as having generally higher online usage than other content area teachers. The same study also finds that teachers with schoolwide permissions in the ARIS system showed higher use than teachers with classroom level access only. Table 7 summarizes some of these teacher-level determinants across the multiple studies.

These valuable studies of educators' online data use provide a foundation for studying how and how much teachers access online student data systems. At this point, the varying systems and populations studied, along with different outcome metrics make explicit comparisons of teacher data use difficult. However, as the number of investigations grow, researchers may converge on a framework that facilitates clearer comparison of usage across data use systems. On the one hand these studies indicate that significant proportions of variation in teachers' online access are attributable to differences between schools and teachers. On the other hand, few factors have been consistently identified to account for this variation, particularly the larger amounts of teacher-level variation. With teachers' online use of data systems generally low, identifying the determinants of that use is critical to the future of online data systems and by extension how such systems can be best used to improve student achievement and learning.

Table 7

Summary of Teacher-Level Determinants in Studies of Online Teacher Data Use

	Wayman et al. (2009)	Wayman et al. (2011)	Shaw and Wayman (2012)	Shaw (2010)	Tyler (2013)	Gold et al. (2012)
Teacher Demographics						
Years of Experience	Significant (-)	Not Sig.	Significant (-)	Not Sig.	Not Sig.	NA
Master's Degree	Significant (+)	NA	NA	Not Sig.	Not Sig.	NA
Content Area	NA	NA	NA	NA	NA	Math Use > English, Science, Social Stud.
Value-Added Teacher Score	NA	NA	NA	NA	Not Sig.	NA
System Access	NA	NA	NA	NA	NA	Schoolwide Teacher Use > Classroom Teacher
Incoming Student Achievement	NA	NA	NA	NA	Significant (-)	NA
Survey Scales						
Attitude towards Data System	NA	NA	NA	Not Sig.	NA	NA
Data Effective for Pedagogy	Significant* (-)	NA	NA	Not Sig.	NA	NA
Freq of Instruct. Data Use	Significant (+)	NA	NA	Not Sig.	NA	NA
Teacher Efficacy	NA	NA	NA	Significant (-)	NA	NA

*Odds Ratio of 0.68 (p = 0.01) for Ever Used Acuity

To this point, sections of this dissertation have discussed past research on the conceptual frameworks, teacher attitudes, and teacher practices of DBDM. This ample body of work has furthered understanding of teachers' data use behaviors and of the factors that may systematically encourage or discourage the use of data for instruction, as well as the more specific use of online systems providing access to student data. The next section considers a related and critical question, asking not what teachers do with data, but rather what programs encouraging data use, as well as the actual use of student data, have achieved.

Teacher Outcomes for Data Use Interventions

Along with more access to student data, more training has been implemented in schools on how to use that data. From the federal-level down, administrators have made clear that the use of data in improvement efforts is a priority ("American Recovery and Reinvestment Act of 2009). Interventions to encourage teachers' data use, therefore, have been plentiful and have met with varied degrees of success. Outcomes for these studies aimed at increasing educators' use of data for instruction range from improved teacher attitudes towards data use to an increase in observed usage of data. In a thorough review of interventions supporting teachers' data use, Marsh (2012) reported that most studies of data use interventions focus on implementation outcomes; few consider an intervention's impact on participant attitudes and behaviors, and even fewer address impact on student achievement. Several of the studies reviewed find that teachers perceive data as generally useful and valuable to their practice, though unfortunately, these studies do not examine changes in teacher attitudes before and after intervention, aside from surveying teachers retrospectively about changes in their perception of data use over time.

In several survey-based studies, teachers responded positively to questions about increases in their own knowledge or skills after data-usage interventions (Huffman and Kalnin, 2003; Marsh, 2007). Other studies found more mixed results, with some data teams demonstrating changes in their thinking about DBDM and others not. In some cases, high percentages of teachers indicated the usefulness or value of data, while much lower percentages of the same teachers indicated that the data had changed their thinking about students (Nelson and Slavit, 2007; Quint, Sepanik, and Smith, 2008).

Beyond investigation of data use alone, Gallimore, Ermeling, and Saunders (2009) examined teachers' attributions of student achievement and found that teachers involved in collaborative data teams attributed student gains to the teachers' own efforts, while teachers uninvolved in such teams attributed students' achievement to personal characteristics or conditions of students, such as socioeconomic status.

Overall, Marsh (2012) found that across 20 studies of interventions meant to increase the use of student performance data, only half found an increase in reported use of student data. One of these studies based its finding of increased data use on the principals' report of increased staff use of data (Copland, 2003), while other studies were able to provide more substantial support in the form of correlations between implementation support and frequency of data use (S. Anderson, Leithwood, and Strauss, 2010; Kerr et al., 2006). Several studies also used self-report outcomes related to changes in classroom practice or school decision-making (Denton, Swanson, and Mathes, 2007; Huffman and Kalnin, 2003; Marsh, Sloan McCombs, and Martorell, 2010; Murnane, Sharkey, and Boudett, 2005; Quint et al., 2008; J. A. Supovitz, 2006; Sutherland, 2004). One study of the Getting Results project by McDougall, Saunders and Goldberg (2007) provided stronger evidence of impact through a qualitative analysis of observation, interview, and focus group data comparing intervention and comparison group schools. Impacts on intervention schools included increased attention to academic goals in instructional planning and more systematic collection and use of assessment data, particularly writing assessments, to inform teachers' classroom decisions.

Along with these positive impacts from data interventions, other negative or mixed results have been reported, including instances where interventions failed to impact teacher data use or where teachers leveraged the structure of accountability metrics by focusing instruction on test-taking skills or by directing resources at so-called “bubble kids,” (those closest to proficiency on state tests) instead of at a wider range of students (Kerr et al., 2006; Marsh, Hamilton, and Gill, 2008; Moody and Dede, 2008; Murnane et al., 2005; Porter and Snipes, 2006).

Student Outcomes for Teacher Data Usage

While the multiple process models discussed earlier suggest pathways by which improved access to student data could impact student outcomes, evidence of the impact of teacher data use on student outcomes is extremely limited (Marsh, 2012; Mandinach and Gummer, 2015; Hoogland et al., 2016; Poortman and Schildkamp, 2016). In Marsh’s (2012) survey of data use interventions, only six studies examined impact on student outcomes, and only two of these studies found positive effects. The first, a study of elementary schools implementing the Getting Results model demonstrated an increase of 15.5 Normal Curve Equivalent (NCE) Units as compared to an increase of 11 NCE units for comparison schools, resulting in an adjusted effect size of 0.75 (McDougall et al., 2007). The second, a longitudinal study of the High Reliability Schools (HRS) Project (Stringfield, Reynolds, and Schaffer, 2008) found that improvement in the national exam passing rates was much higher for students enrolled in HRS model schools. While both studies provided promising results, they both investigated schoolwide reform efforts, taking place over multiple years, making it difficult to separate out the effects of data support interventions from the impact of the larger school wide reform.

Two other studies (one related to the High Reliability Schools Project) found an intensity effect, with increased data-use training for teachers correlating with higher student achievement (Marsh, McCombs, and Martorell, 2010; Reynolds, Stringfield, and Schaffer, 2006). Reynolds et al. (2006) found such an effect within the larger evaluation of the schoolwide HRS movement, while Marsh et al. (2010) found a small, but significant relationship ($p < .001$) between time spent by reading coaches on data analysis support and students' reading achievement.

Studies conducted by the Center for Data-Driven Reform in Education on large-scale benchmark testing (Carlson, Borman and Robinson, 2011; Slavin, Cheung, Holmes, Madden and Chamberlain, 2013) demonstrated extremely mixed results across years of implementation, content area, and method of analysis, though generally, the authors suggest that the implementation of benchmark testing and workshops on data use was not sufficient for improvement in student achievement. Larger effects were observed in the later years of the study after schools had adopted and implemented reading or math programs with demonstrated evidence of effectiveness. Again, as with studies related to the Getting Results and High Reliability Schools projects it is difficult to separate the impact of the data use intervention itself from the impact of system wide reforms.

In another large-scale study of the impact of benchmark testing systems, Wayman, Shaw, and Cho (2017) analyze student outcomes specifically as they relate to teachers' use of an online system for benchmark formative assessment. After two years of system use in elementary and junior high reading and math classes, the only significant, but small, relationship was between the consistency (number of weeks) of educator use and elementary reading achievement growth ($p = 0.01$, SD Difference = 0.05). In one of the only other studies relating teacher online data use to student achievement, Tyler (2013) reports that teacher usage of an online interim testing system was unrelated to student growth on state exams or to performance on quarterly benchmark exams for core content area teachers in grades 3 through 8.

Other studies attempted to measure the impact of data systems by evaluating the impact of such systems on a randomly selected group of initial users. May and Robinson (2007), for example, examined the impact of Ohio's Personalized Assessment Reporting System (PARS), an online system for parents, teachers, students, and administrators to aid preparation for the Ohio Graduation Test. In the 51 randomly selected pilot high schools, there was no evidence of increased outcomes for students taking graduation exit exams for the first time, but some impact was found for students who had previously taken and failed a state graduation test (May and Robinson, 2007).

In a similar vein, Konstantopoulos, Miller, van der Ploeg and Li (2016) examined the impact of participating in two interim assessment programs, mClass and Acuity, in Indiana during the 2011-12 school year. Seventy schools were randomly selected from a pool of 157 volunteer schools and then randomly assigned to receive the interim testing programs. With annual Indiana state test scores as outcomes, no significant effects were found for grades 3-8 in mathematics or reading, while significant but negative effects were found for K-2 for the same content areas.

Additional studies have found no impact on student outcomes when comparing comparison and intervention conditions. A district-level intervention which combined several supports for teachers' use of data in an inquiry framework found no differences with similar comparison districts, as well as no differences in intensity of outcomes related to the degree of implementation (Porter and Snipes, 2006). A randomized control trial in Boston tested the impact of coaching teachers to better understand and use reading assessment scores, but also failed to find significant effects (Quint et al., 2008).

In a more promising set of studies for data use in schools, evaluations of the Learning Schools Model (LSM) in New Zealand showed evidence of sustained results in reading achievement (Lai, McNaughton, Amituanai-Toloa, Turner, and Hsiao, 2009; Lai, McNaughton, Timperley, Hsiao, 2009; Lai and McNaughton, 2016). Unlike previously discussed interventions which focused more directly on the implementation of benchmark testing, the LSM had more in common with the Getting Results and HRS projects, implementing a collaborative, problem solving approach where professional learning community teams worked together to identify problems, collect and analyze data related to the problem, and act based on their analysis (Lai and McNaughton, 2016). Along with the implementation of a collaborative, problem solving approach, the LSM model also differed from benchmark testing approaches in several other key areas:

- Developed partnerships between researchers and practitioners, prioritized design and testing of interventions, leveraged networked communities for learning, and examined variation in performance across the system in keeping with design-based research (Anderson and Shattuck, 2012) and improvement science (Bryk et al., 2015),
- Used data from teacher observations and a range of student assessments to connect patterns of teaching to patterns of achievement,
- Used a consistent evaluative framework in collaborative discussions to evaluate hypotheses and proposals, in this case problem-based methodology (Robinson and Lai, 2006),

- Based analysis in local, content-area knowledge, linking data use professional development (PD) with content area PD (Lai and McNaughton, 2016).

Overall, The Learning Schools Model (Lai and McNaughton, 2016), the High Reliability Schools Model (Stringfield, Reynolds and Schaffer, 2008), the Getting Results Model (McDougall et al., 2007), and years three and four of the CDDRE district-reform study (Slavin et al., 2013) provide the most consistent evidence of the impact of DBDM interventions on student outcomes. These data use interventions, however, were only a subset of more comprehensive school reforms, making the separate impact of improved access and use of student performance data difficult to determine. For those interventions limited more strictly to providing access and basic training to benchmark assessment or data systems (Henderson et al., 2007; May and Robinson, 2007; Slavin et al., 2013, Konstantopoulos et al., 2016) few significant results were found. For the very limited studies analyzing teachers' use of online data and assessment systems, as opposed to access alone, the only significant result was found by Wayman et al. (2017), who found a relationship between the consistency of teachers' usage of an online assessment system and growth in elementary reading achievement.

Some researchers argue that this lack of supporting evidence is not surprising given the infant stage of DBDM adoption for most districts and the experimental difficulty of isolating the effects of data use (Hamilton et al., 2009; Mandinach and Jackson, 2012). The lack of evidence supporting the impact of DBDM is perhaps also related to limited understanding of the teacher-level factors enabling and inhibiting teachers' data use. Accounting for the variability among teachers in adopting DBDM may help refine understanding of data use outcomes. A fuller understanding of the individual-level determinants of teacher online data use, alongside a fuller picture of what “use” looks like appear critically necessary for expanding usage of online student performance data and its possible impacts on learning and achievement**.**

Limitations and Next Steps of DBDM Research

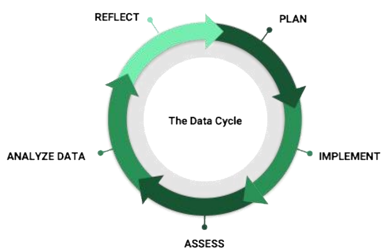
As described in the previous sections, DBDM in schools has not followed a trajectory of research to practice—where an intervention, grounded in evidence and relevant theory, is tested and refined before use. Instead, the current movement of data use in schools has followed a trajectory of expanding DBDM practice with research racing to catch up, investigating a set of practices rapidly diffusing across the k-12 landscape. Under these circumstances, with research trailing practice, studies of DBDM have tended toward the qualitative description of “what is” in the realm of data use in schools. These qualitative descriptions have, in turn, generated process models, which both synthesize descriptions of practice and serve as guide and inspiration for further adoption of DBDM in schools. On the one hand these process models of Data-based decision making are clear and compelling: Plan Do Study Act. Repeat. However, elaborating and qualifying such models is critical to the success of future data use interventions.

Recognizing this need, another line of inquiry in DBDM studies, conducted on a much smaller scale, has worked towards establishing determinant frameworks (Nilsen, 2015) for data use, frameworks that describe the most important factors encouraging and discouraging teachers' use of student data. Figure 5 summarizes this overall movement between process models and determinant frameworks in DBDM research.

Overview of Previous Data-Based Decision Making Frameworks

Type of Framework	Process Models ⟷	General Determinant Frameworks
Description	Models that describe or guide the stages of translating research into action	Models that identify barriers and enablers to implementation outcomes

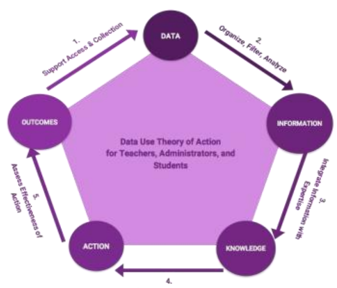
Examples



See Figure 1:
Means et al (2010)

Organizational Level	Facilitators	Inhibitors
District Accountability	Schoolwide focus	Proficiency focus
School Leadership	Organizational learning	Short-term accountability
School Context	Structured collaboration	Limited time and power
Teacher Capacity	Data competence & confidence	Lack of training & confidence
Teacher Beliefs	Assessment helps learning	Assessments are invalid

See Table 4:
Datnow and Hubbard (2015b)



See Figure 3:
Marsh (2012)

Type of Factor	Enablers & Barriers
Organization	Vision & Norms Leadership Support Collaboration
Data	Accessibility of timely data Usability Quality of the data
User	Knowledge and Skills Disposition to Use Data

Adapted from Schildkamp et al (2017)

Figure 5. Overview of previous data-based decision making frameworks

While DBDM research has been successful at generating process models and a few examples of possible determinant frameworks, it has struggled to quantitatively describe or test the strength of relationships between possible factors determining teachers' use of data and teachers' actual use of student data. Even with a strong base in qualitative research, current determinant frameworks have failed to adequately explain variability in teachers' DBDM behavior when applied in large-scale, statistical analyses. Overall, the field of DBDM in schools still lacks strong quantitative evidence linking key enablers and barriers to variation in teachers' data use. See Table 6 and Table 7 for a summary of previously explored enablers and barriers to teacher data use.

While more large-scale studies are needed to generate consensus on these determinants, in the two studies of this dissertation, I apply an exploratory, descriptive, data-intensive approach to generate new lines of investigation. To address the core challenge of testing and improving the determinant frameworks of teacher data use I suggest three expansions of DBDM research.

- An expansion of data and methods
- An expansion of theory
- An expansion of participation

I provide a brief outline for each of these expansions in this section, with the next section on Additional Frameworks, Theories, and Communities providing the more extensive background and survey of the relevant literature.

Expansion of data and methods. A pervasive issue of DBDM research, coloring many of its findings, is an over-dependence on survey outcomes (Snodgrass, Rangel, Monroy, Bell, and Whitaker, 2015) for operationalizing teachers' data use, often due to the difficulty and expense of observing teacher data use in action. Several researchers in DBDM have made the related point that a critical, and less frequent, area for research are educators' actual interactions with data, "zooming in" on the observed daily or weekly practice of data use (Little 2012; Datnow and Hubbard 2015).

As online systems for accessing student data become increasingly ubiquitous, the log files from these systems may provide a valuable window into educator's usage over days, weeks, and months. While a few studies have used log files to analyze teachers' interaction with these systems (Wayman et al., 2009; Shaw and Wayman, 2012; Gold et al., 2012; Tyler, 2013), the metrics generated in these studies have only begun to explore the potentials of log file analysis. Wayman et al. (2017) points out that relevant studies have used limited definitions of usage based on total number or duration of actions to define teachers' engagement with online systems, and that perhaps additional conceptions and metrics of use would further analysis of online systems for accessing student performance data.

To address this over-dependence of DBDM research on survey outcomes and expand the potential of log files analysis, I adopt methods commonly used in the research communities of Educational Data Mining and Learning Analytics (EDM/LA) (Baker and Siemens, 2014) to analyze the log files generated from users' interactions with online systems (Baker et al., 2012, Rodrigo et al., 2012). These methods allow for rich, quantified descriptions of teachers' naturalistic behavior, beyond those used in previous studies of teachers' online data use (Xu and Recker, 2012; Maull, 2013).

Along with an overdependence on survey analysis, the quantitative methods used in previous DBDM studies have focused on variable-centered vs. person-centered approaches to analysis. For example, studies of online data use (Tyler, 2013; Wayman, 2017) have focused on explaining school- and teacher-level factors and had difficulty capturing the variability and the possible subgroups of educators as characterized by their attitudes, behaviors, and outcomes. Drawing from the same EDM/LA communities, I employ data-intensive methods, such as hierarchical cluster analysis as a way of simultaneously identifying patterns across individuals and preserving a view of the overall complexity of variation across users and factors (Bowers, 2010). While expanding the set of data-intensive methods, such as these, may play a key role in future DBDM research, an expansion of the theoretical frameworks guiding this research is just as critical.

Expansion of theory. As Mandinach and Jimerson suggest**,** the underlying logic model of DBDM may be overly simple. To move logically from "(a) train teachers to use data to (b) instructional changes to (c) improved student performance" may be missing a great deal of the "connections to motivational theory, developmental psychology, and other theoretical bases that may influence how data use plays out in classrooms and other organizations" (Mandinach and Jimerson, 2016, p. 4). In keeping with these insights, I propose two areas where theories of psychology and marketing may aid DBDM research in expanding both the determinants and outcomes considered in data use studies. Expanding DBDM studies to include well-supported theories from other behavioral fields may help in identifying the key enablers and barriers to teacher data use.

While the bulk of investigation into data use determinants has focused on the external obstacles or necessary elements for teachers' adoption of data use (Datnow, Park, and Kennedy-Lewis, 2013; Hoogland et al., 2016; Hubbard, Datnow, and Pruyn, 2014), less attention has been focused on teachers' individual attitudes and beliefs and their association with data use. A few researchers have attempted to quantitatively describe relationships between data use behaviors and general data use attitudes, such as those captured by surveys such as the SEDU and TDUS (Wayman et al., 2009; Wayman et al., 2016). However, more such studies are needed, as well as studies which expand the range of attitudes and beliefs considered in relation to DBDM. Social-cognitive theory (Bandura, 1986), for example, is fundamental to much contemporary understanding of motivation, and offers the construct of self-efficacy as a possible determinant of teacher involvement in data use. Only a small number of previous studies have applied social-cognitive theory, in the form of teaching self-efficacy, to the analysis of DBDM attitudes. These studies have begun to establish the reliability and validity of survey scales for measuring self-efficacy in teachers' data use, as well as a prevalent lack of data use self-efficacy among teachers (Dunn, Airola, Lo, and Garrison, 2013a; 2013b).

A related area of common behavioral theory, the Technology Acceptance Model (TAM) (Davis, 1989), also has potential to define key barriers and enablers to teacher data use, particularly when that use is mediated by online information technology systems. TAM employs dual constructs of perceived usefulness (PERUSE) and perceived ease of use (PEOU) to explain variability in the adoption of technological innovation. Moving past the initial adoption of a technology to patterns in its use over the longer term, Use Diffusion theory (Shih and Venkatesh, 2004) proposes that user behavior can be viewed productively through dimensions of frequency and variety. Such a lens provides a possibly valuable framework for studies of teachers' online data use, drawing the focus away from sheer amount of time spent in a system to the quality of teachers' online engagement.

Expansion of participation. While expansions of method and theory may prove useful for DBDM research, an additional expansion of its stakeholders and participants may, in the long run, prove even more valuable, particularly regarding the long-term success of translating research findings to successful practice. Recent movements in Educational Data Science (Piety, Hickey, and Bishop, 2014), Collaborative Data-Intensive Improvement (Krumm et al., 2018), and general frameworks for Learning Analytics (Gasevic et al., 2017) prioritize the integration of research and practice. These movements propose a range of innovations, from the creation of new roles and responsibilities (Agasisti and Bowers, 2017), to the leveraging of knowledge from diverse disciplines (Piety, Hickey, and Bishop, 2014), to designing visual data analytics that leverage complex data and analyses for human judgement (Bienkowski, Feng, and Means, 2012; Krumm et al., 2018).

With the notable exception of the Teacher Data Use Survey (Wayman et al., 2016), few such visualizations or dashboards exist for informing school decision-making around teacher data use. DBDM may be a priority for many districts and schools, yet limited tools exist to help school leaders in monitoring and understanding the status of these efforts. This study generates visualizations designed to efficiently inform school leaders about the complex factors related to teacher data use in schools. Such visualizations may increase schools' abilities to understand the impacts of DBDM initiatives and adapt them for success in local contexts.

In summary, based on review of the literature of DBDM, I have identified three areas for expansion of research in data-based decision making, all with the aim of better explaining the factors impacting teachers' use of student data:

- An **expansion of data and methods** from self-report measures to log files of observed data use and to data-intensive methods from the LA/EDM community.
- An **expansion of theory** to constructs from social-cognitive theory, the technology acceptance model, and use diffusion theory to explain variability and expand outcomes for teacher data use.
- An **expansion of participation** by generating new tools for school leaders to monitor and improve data use initiatives.

In the next section, I will provide additional background and a survey of the literature on relevant methods in EDM/LA, theoretical constructs of social-cognitive theory, technology acceptance and use diffusion, and current movements to integrate research, data-intensive analytics, and education practice.

Expanding Methods, Theory, and Participation of DBDM Research

Expanding data and methods. The emerging disciplines of Educational Data Mining and Learning Analytics (EDM/LA) (Baker and Siemens, 2014) provide this study with a set of methodologies and conceptual frameworks for exploratory investigation of DBDM. The possible contributions of EDM/LA to DBDM research fall into three main areas: elaborated techniques for log file analysis and feature generation, methods for the descriptive visualization of high-dimensionality data, and emerging frameworks for research that bridge data-intensive analysis; collaborative, design-based approaches; and data use in schools. I will briefly introduce these emerging disciplines and then discuss their possible contributions to the investigation of DBDM.

What is Educational Data Mining/Learning Analytics (EDM/LA)? Over the last few decades the methods of data mining and analytics have spread across multiple disciplines. The field of education, though a bit later than some to make use of these methods, has increasingly utilized techniques developed in the data mining community for analyzing and drawing actionable inferences from large datasets. In education research, data mining and data visualization methodologies have taken root largely in the Educational Data Mining (EDM) and Learning Analytics (LA) communities (Baker and Siemens, 2014)**.**

Educational Data Mining as defined by the International Society of Educational Data Mining is "an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students, and the settings which they learn in." ("Home | International Educational Data Mining Society," 2018). While Educational Data Mining has its origins in the analysis of student-software interactions (Baker and Yacef, 2009), its application and categories of analysis have expanded widely from these beginnings.

Applications of data mining and learning analytics methodologies have delved into areas as diverse as providing natural language support for tutoring systems (Nye, Graesser, and Hu, 2014), automated analysis of text features (McNamara, Graesser, McCarthy, and Cai, 2014), at-risk flags for student intervention based on wide range of indicators (Arnold and Pistilli, 2012), and fine-grained explorations of students' affective states and their relationships to learning (Baker, D'Mello, Rodrigo, and Graesser, 2010). To accomplish these tasks, researchers in this field utilize several broad methods of analysis: prediction methods, structure discovery algorithms, relationship mining, distillation of data for human judgement, and discovery with models (Baker and Yacef, 2009). Some of these categories, relevant to the current studies, will be discussed below.

Predictive modeling. Generally, prediction modeling methods attempt to produce highly generalizable mathematical models that predict one dependent variable from various independent variables. The warrant for the generalizability of prediction models is often constructed by following a validation process, where the statistical prediction model is built using one subset of the available data and then tested for its performance on a separate subset of that data. In this cross-validation method, a model is iteratively trained and tested on separate partitions of the data and its performance is averaged across those iterations (Efron and Gong, 1983)**.** As the term suggests, predictive methods have often been used to predict future events—passing state tests (Pardos, Baker, San Pedro, Gowda, and Gowda, 2013), dropping out of high school (Bowers, Sprott, and Taff, 2013), and enrolling in college (San Pedro, Baker, Bowers, and Heffernan, 2013). Along with anticipating future events, however, prediction methods have been used effectively to predict variables that are too cumbersome or disruptive to collect directly (Baker, Corbett, Roll, and Koedinger, 2008) or that can only be known indirectly, such as students' knowledge of a particular topic (Pardos, Gowda, Baker, and Heffernan, 2011).

Structure discovery. Structure discovery methods, another broad category of work in EDM/LA, are used to explore patterns in data without the imposition of a prior schema. These methods contrast with prediction methods in that they lack a pre-defined outcome or label that guides analysis. Structure discovery methods include clustering, factor analysis, social network analysis, and domain structure discovery (Baker and Siemens, 2014). In the EDM/LA context, structure discovery methods have been used to explore the social networks that emerge in Massive Open Online Courses (MOOCs) (Rosé et al., 2014) and to cluster students based on a variety of motivational beliefs and behavior patterns (Beal, Qu, and Lee, 2006). Most relevant to this study, structure discovery, in the form of hierarchical clustering analysis, has been used to explore possible subgroups of students based on their grades over time (Bowers, 2010), their interactions with an online learning management system (Lee, Recker, Bowers, Yuan, 2016; Krumm., et al., 2018), and their academic performance across various instructional methods (Nitkin, 2018).

Assisting human judgement. Another EDM/LA focus is on the distillation of data for human judgement. Analyses, metrics, access to relevant data, and information visualizations generated within EDM/LA have great potential to assist human stakeholders in optimizing their decision-making process (Baker and Siemens, 2014). Research at the Open University UK as well as online additions to traditional live classes has explored teacher use and access to learning analytics (Arnold, 2010; Arnold and Pistilli, 2012; Clow, 2012, 2014). The creation and utility of visualizations and dashboards has been another area of inquiry, with studies of open student models and learning dashboards (Verbert et al., 2014; Verbert, Duval, Klerkx, Govaerts, and

Santos, 2013) providing useful explorations of design and usage. Heatmaps and clustergrams (Bowers, 2010; Lee, Recker, Bowers, and Yuan, 2016), learning curves (Koedinger and Mathan, 2004), and learnograms (Hershkovitz and Nachmias, 2009) are all examples of visualizations of learning or behavior which can inform educators' decision-making about students. Other work in this area has focused on generalized frameworks for producing effective judgements (Wise, 2014), on the integration of data-intensive methods with DBDM frameworks (Agasisti and Bowers, 2017; Bowers, Krumm, Feng, and Podkul, 2016; Bowers, 2017), and on the specific use of enhanced, data-intensive reporting in high-needs and/or secondary school environments (Hawn, 2015; Ocumpaugh et al., 2010).

Contributions to DBDM research. Specifically, in relation to this study, methods used in EDM/LA contribute in three areas: (1) intensive log file analysis and feature generation, (2) person-centered methods for structure discovery, and (3) descriptive visualizations of high-dimensionality data. These specific contributions will be discussed in detail below.

Log file analysis. Previous DBDM studies of teacher interaction with online student data have used a relatively compact set of metrics to analyze log file interactions. These metrics have been overwhelmingly based on rates of access, including session counts (Wayman et al., 2009a), total time spent in the system, and time spent in various subsystems (Tyler, 2013). Studies have also used rate of access as a means of defining user categories (low, mid, and high users) to differentiate analysis (Gold et al., 2012). While these studies have provided valuable initial insight into teacher usage of these systems, practice in EDM/LA suggests a much wider range of techniques for analysis of teacher log file interaction.

From their beginnings, EDM/LA, though particularly EDM, have focused on creating actionable insight from the traces of user behavior captured in log file interactions (Baker and

Yacef, 2009). While most of this effort has been focused on analyzing student log file interaction, EDM/LA studies of teacher interaction with online curriculum and library systems have been conducted as well (Xu and Recker, 2012; Maull, 2013). In addition, some techniques for analysis of student log files may be transferrable to the analysis of teacher log files. At the same time, it is important to recognize that teacher log files of interaction with an online data system do not include the same rich, moment-by-moment, instructional, performance, and learning data as do log files of students' interaction with online instructional applications. While it seems likely that greater understanding can be gained from the application of EDM/LA techniques to the web usage of practitioners, it is also likely that the range of inferences available from students' interactions with tutoring software will remain far wider.

Central to prediction modeling in EDM/LA practice is the creation of an extensive range of features, or inputs, that can be used as descriptions of the dataset in question and as predictors in related data mining analyses. This feature generation approach leads to a much wider range of metrics based on log file interactions. In Baker et al. (2012), for example, 58 features were calculated based on 20-second clips of student interaction with an online tutor. These 58 features were then aggregated as min, max, mean, and sum, for a total of 232 features used to develop automated detectors of student affect. Many of these 58 features relate to students' moment by moment state of learning, and while that level of analysis is not currently inherent in teachers' online, data-use interactions, several EDM approaches to feature generation can still inform future DBDM investigations. For example, the time between actions, usage of help functions, reference to past actions, the normalization of features, the standard deviation, minimum and maximum of interactions, number of unique values entered, and the creation of feature interactions are all commonly used EDM techniques (Rodrigo et al., 2012) that could be applied to the analysis of teacher log files in a descriptive capacity, as well as for their possible predictive utility in extended EDM analyses.

Additionally, a limited number of EDM/LA analyses of teacher log file data do exist as precedent for DBDM analyses. In particular, Maull (2013), Maull, Saldivar, and Sumner (2010b), and Xu and Recker (2012) have analyzed teacher interaction with online curriculum and digital library systems, respectively. In these studies, both authors first generated feature sets summarizing teacher behaviors in the system and then applied clustering techniques, such as K-means, expectation-maximum (EM) likelihood, and Latent Class Analysis (LCA), in order to group teachers by patterns of engagement with the system. In both cases, these online patterns are analyzed in relation to external information about the teachers involved. In the case of Xu and Recker (2012), for example, years teaching and comfort with technology were also analyzed. In the case of Maull (2013) teacher interviews and student outcomes were connected to online usage to create a typology of teacher use of an online science curriculum.

The overall investigation by Maull (2013) includes several key components for consideration. Several surveys, an adoption interview, and a classroom observation cycle were conducted and analyzed in the context of the log file analysis. This qualitative work identified a continuum of teacher practice from less to more classroom differentiation on the part of teachers that could then be tested empirically by assigning each interface widget of the curriculum site a category aligned to this continuum of teaching differentiation. In this case, interface widgets were assigned as either a Publisher, Interactive, or Teacher-created widget, with Interactive and Teacher-created widgets believed to have more potential for differentiation than Publisher-created materials.

Continuing the analysis of log file data, Maull (2013) determined the 20 most frequently used of these interface widgets and used access counts to those 20 widgets as the bulk of 27 total experimental features. The remaining seven features included more general counts of clicks in the system as well as duration of sessions. These features were used in several cluster analyses over both semesters of collected data, utilizing both K-Means and Expectation Maximization algorithms. An additional Principle Components Analysis (PCA) was conducted as a comparison against clustering methods. For the third analysis, combining two semesters of data, Maull refines the feature set to 19, deciding to remove features not explicitly aligned to the differentiation continuum of teacher practice, such as overall duration of use. K-Means and EM algorithms were then re-run, experimentally identifying the number of clusters (K) by testing the within-sum-of-squares values for cluster sizes from 3-15. The work of Maull et al. and Xu and Recker to generate categories of teacher interaction opens the way for future personalization of these systems around the needs and classroom contexts of observed types of teacher behaviors (Maull, Saldivar, and Sumner, 2010a).

A more recent analysis of an online science platform for teachers (Snodgrass et al., 2015) provides a survey of learning analytics methods that might be applied for the purpose of reviewing curricula and applies Hierarchical Linear Modeling (HLM) to compare teacher usage of online curricula to self-report measures to student outcomes. HLM analyses found both significant positive and negative relationships between teachers' access to areas of online curricula and students' science state test scores. When evaluating the usage of online curriculum resources, the authors propose investigating four categories of behavior: adherence to use as intended, patterns of usage, dosage/exposure to the system, and engagement. Patterns of usage, in particular, stand out as an area for possible investigation, since past studies of teacher data systems have tended to focus on rates of usage.

While not strictly addressing educator online behavior, studies of user behavior in Learning Management Systems (LMS) offer other potential methods for analysis and feature generation, some of which may prove fruitful for descriptions of educator usage behavior. Recker and Lee (2016) review the current literature on this topic, discussing the information collected by LMS logs that has been used in predictive or clustering studies. While many tracking variables from LMS systems are not relevant to data use systems, others, such as the regularity of the login interval (Il-Hyun, Kim, and Yoon, 2015) or number of downloads (Yu and Jo, 2014) may be applicable. Other LMS studies have applied clustering algorithms in order to group students according to multiple factors. Such studies have identified participation groupings of very active, active, and non-active students (Romero, Ventura, and García, 2008) and grouped students based on their self-reported use of strategies and their online LMS tool use (Lust, Elen, and Clarebout, 2013). While most LMS clustering studies have used the K-means algorithm, fuzzy c-means was found to be more successful than k-means and subtractive clustering in one study (Yildiz, Bal, & Gulsecen, 2015).

These recommendations for log file analysis more fully embrace the human-computer interaction (HCI) paradigm of investigation into user behavior (Dumais, Jeffries, Russell, Tang, and Teevan, 2014), defined as the extraction and aggregation of user behavior and interaction from systems capable of recording those behaviors and interactions for later analysis (Maull, 2013). The strength of the user behavior approach, and the application of EDM/LA methods, is its ability to generate hypotheses, inferences, and methods for visualization that will ultimately aid the users of such systems. For educational systems in particular, many of which remain proprietary and hidden (Agasisti and Bowers, 2017), such open investigations illuminate educators' complex relationships with evidence, information, and decision-making, feeding back into the design of improved online systems and facilitating their integration into school and classroom process.

In summary, Table 8 lists relevant EDM/LA and data use studies, along with their approaches to metrics, page categorization, and analysis.

Table 8

Summary of Educator Online Data Use Studies

Citation	System Analyzed	Users	Usage Metrics	Report or Page Categories	Analyses
Wayman et al., 2009, 2011 Shaw, 2010 Shaw and Wayman, 2012 Wayman et al., 2017	Interim Assessment	Classroom Teachers: Grades 3-8	Weekly session counts Prevalence: Actions/Year Consistency: Weeks used/Year	Instructional functions: Instructional resources, Management, Reports, Completion status	Hierarchical Linear Modeling, Usage Correlates
Gold et al., 2012	Student Data	Admin. Teachers Specialist Staff	Session counts Duration of sessions Total duration of use	My Students, Reports	Descriptive Statistics
Maull et al., 2010a, 2010b Maull, 2013	Science Curriculum	Science Teachers	Clicks/Session Duration of sessions Clicks on frequently-used website features	Interactive resources, Teacher-contributed, Publisher-oriented	Clustering, Principle Components Analysis, Usage Correlates
Tyler, 2013	Interim Assessment	Core Subject Teachers: Grades 3-8	Total duration of use Duration of use by page category	Class-level reports, Students-in-class, Individual student, Item-level, Resources	Hierarchical Linear Modeling, Usage Correlates
Drake, 2015	Student Data	Principals	Monthly session counts Cumulative session counts	Student achievement, Student demographics, Behavior, Attendance, Teacher, Other	Latent Class Growth Analysis
Snodgrass et al., 2015	Science Curriculum	5th Grade Science Teachers	Category Indices: Engage, Explore, Explain, Elaborate, Evaluate	Engage, Explore, Explain, Elaborate, Evaluate (Website Structure)	Hierarchical Linear Modeling

Descriptive visualizations of high-dimensionality data. EDM/LA practice also contributes to the current study through its development of methods for the visualization of high-dimensionality data (Bienkowski et al., 2012). The distillation of data for human judgement has long been a major focus of EDM/LA (Baker and Yacef, 2009). One approach for maximizing human judgement is to present complex information in the form of visualizations that leverage the large bandwidth and pattern-finding ability of the brain's visual systems (Ware, 2012). Such visualizations are often exploratory in that they do not attempt to confirm specific hypotheses (Krumm et al., 2018). However, an exploratory and descriptive approach is of particular use and potential for schools which struggle to make decisions daily, based on analyses that are insufficiently powered to provide statistical significance. In the absence of sufficient power or in the presence of far more factors than observations, exploratory and descriptive visualizations provide actionable summaries of information (Loeb, Dynarski, McFarland, Morris, Reardon and Reber, 2017). While such visualizations may fail to generalize to other contexts or to account for the role of chance in observations, they still can effectively and iteratively inform local decisions, improvement process, and theories of action.

EDM/LA draws on a long history of exploratory data analysis (EDA) in the data science tradition, from Tukey's (1977) work laying out a justification and disposition for exploration, as well as a set of visualization methods, through Behrens' (1997) attempts at providing a more standardized framework of heuristics for exploration. Of relevance to the current study is work on clustergrams by Bowers (2010). Clustergrams are the visualization of a Hierarchical Cluster Analysis (HCA) combined with a heatmap display of the factors generating the hierarchical clustering. These combinations of dendrograms, heatmaps, and annotations appended to either observations or factors, have offered possibilities for insight and hypothesis generation in multiple contexts. Originally developed in the context of bioinformatics (Gu, Eils and Schlesner, 2016), heatmaps have been implemented in the educational context for studying student grades over time (Bowers, 2010), predictions of student performance (Villagra-Arnedo et al., 2017), patterns of kindergarten students' interaction with mathematics software (Tucker, Lommatsch, Moyer-Packenham, Anderson-Pence and Symanzik, 2017), learners' video viewing behaviors (Kleftodimos and Evangelidi, 2014), students' interactions with an LMS (Lee, Recker, Bowers, and Yuan, 2016; Krumm, Means, Bienkowski, 2018), students' interactions with learning activities (Nitkin, 2018), and even the analysis of learning activities in a museum context (Jorion, Roberts, Bowers, Tissenbaum, Lyons, Kuma, and Berland 2018). In each case, the generation of a clustergram, combining clustering and heatmap visualization allows for the exposure of patterns in large data sets without sacrificing the details of individual observations.

On a more global level, Bowers et al. (2016) discusses the intersection of EDM/LA methods with the systemic improvement models of the DBDM movement, formalizing how previous models for educator data use, such as Marsh's data use theory of action (Bowers et al., 2016) might be integrated with EDM/LA data-intensive methods.

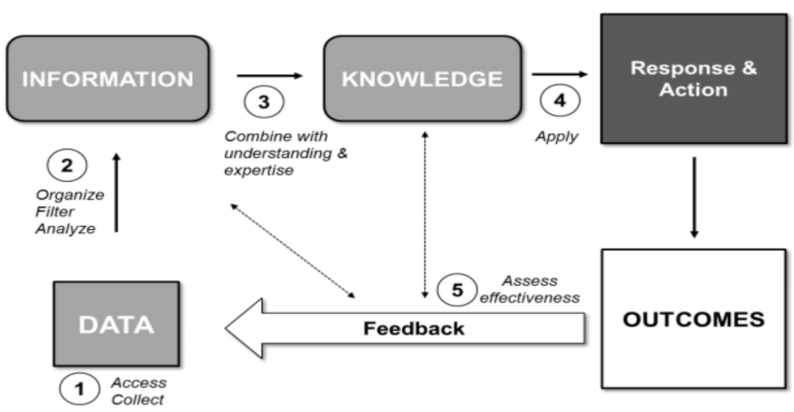


Figure 6. Data use theory of action for teachers, students, school administrators and central office staff. Adapted from Marsh (2012) and reproduced from Bowers et al. (2016) with the permission of the author.

In contrast to the Marsh (2012) theory of action, Bowers et al.'s (2016) Combined Logic Model of Data Analytics places the logic of school data use in the context of a data analytics process, expanding the range of methods for DBDM beyond basic exploratory analyses and providing a roadmap for the incorporation of more complex data analytics, such as those discussed earlier in this section, including predictive modeling, clustering, and visualizations.

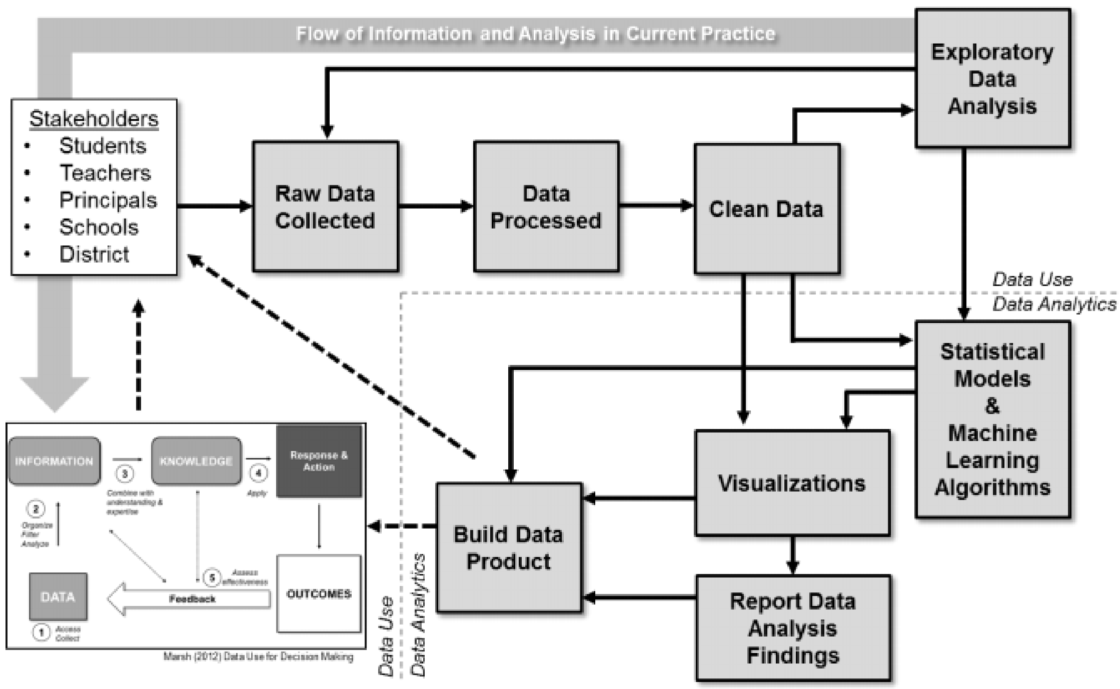


Figure 7. Combined logic model of data analytics for decision making in schools. Adapted from Marsh (2012) and Schutt and O'Neil (2013). Reproduced from Bowers et al. (2016) with the permission of the author.

Expanding theory. Along with an expansion of methods to include innovations from EDM/LA, an expansion of theory may be helpful in considering new determinants for teachers' online data use. Social cognitive theory, specifically self-efficacy is one of the more promising lenses for exploring DBDM behaviors and attitudes. Bandura's (1986) social cognitive theory stands at the base of modern motivational theories and is useful to the analysis of data use behaviors for expanding motivational factors to include the reciprocal interactions between individuals, their behavior, and the environment. Central to social cognitive theory and its descendants is the concept of the "self-organizing, proactive, self-regulating, and self-reflecting" individual (Schunk and Pajares, 2005, p. 86), whose individual beliefs serve as a bridge between the environment and behavior. In the context of the pedagogical use of data, these beliefs can play an especially strong role, as educators struggle with perceptions of fairness to students, institutional accountability, quantitative vs. qualitative understanding, and even their own comfort levels with reading tables, graphs, or interpreting metrics. Within this constellation of beliefs, social cognitive theory defines a central role for the construct of self-efficacy: an individual's beliefs about his or her ability to learn or perform actions at a specified level (Schunk and Pajares, 2009).

Self-efficacy beliefs impact individuals' choices in tasks and behaviors, as well as the effort they expend while working on a task (Bandura, 1977). According to Bandura (1977), individual self-efficacy can be characterized by three dimensions: magnitude, generality, and strength. Respectively, these represent the belief in the level of task that can be accomplished, belief in the specificity of the relevant task domain, and the certainty of these beliefs. These self-efficacy beliefs develop from four sources (a) performance accomplishments (As people encounter success, their efficacy strengthens. As they encounter failure, it weakens.), (b) vicarious experience, (c) verbal persuasion, and (d) emotional arousal (Bandura, 1977).

As a general construct, self-efficacy has been found to relate to task persistence in the face of obstacles, more frequent positive emotions and fewer negative emotions, use of self-regulatory strategies, and increased academic achievement (Bandura, 1997; Pajares, 1996). As a domain-specific construct, research into self-efficacy has addressed the self-efficacy beliefs of teachers, which will be discussed in sections below.

General teaching self-efficacy. Within the professional context, teaching self-efficacy refers to teachers' beliefs about their ability to perform specific aspects of their teaching practice (Tschannen-Moran and Hoy, 2001). Some of these aspects of practice, defined in instruments such as the Teacher Self-Efficacy Scale (TSES) (Tschannen-Moran and Hoy, 2001) and the Norwegian Teacher Self Efficacy Scale (NTSES) (Skaalvik and Skaalvik, 2010) directly engage concepts at the core of DBDM, with subscales such as Efficacy in Instructional Strategies and Efficacy in Student Engagement from the TSES, or, from the NTSES, the Efficacy to Adapt Instruction to Individual Needs. These subscales of Teacher Self-Efficacy relate directly to a teacher's ability to perceive the individual needs of students and respond in an effective manner, either through instructional strategies or methods for student engagement.

In the development of the TSES, Tschannen-Moran, Hoy, and Hoy (1998) bring together two strands of psychological inquiry into an integrated model for teacher self-efficacy: one strand of inquiry into general teaching efficacy (GTE) and another into personal teaching efficacy (PTE). GTE refers, in a nutshell, to teachers' beliefs about the overall efficacy of the teaching profession to create positive change with students, in the face of external factors such as family and community environment and each child's particular cognitive strengths and weaknesses. Work by the Rand Corporation into Teacher Efficacy (Armor et al., 1976) incorporated studies by Rotter (1966) on internal vs. external control, Rose and Medway (1981) on teacher locus of control, and Guskey (1981) on responsibility for student achievement into a construct addressing teachers' beliefs about their impact on students' lives. While causal relationships have yet to be established for these relationships, teacher efficacy, as measured from a combined GTE and PTE perspective, has been correlated to academic achievement (Armor et al., 1976), frequency of disciplinary commands (Rose and Medway, 1981), willingness to implement innovations (Guskey, 1984), and teacher stress (Greenwood, Olejnik, and Parkay, 1990).

Applying Bandura's social cognitive and self-efficacy theory (Bandura, 1977) to this earlier work on teacher efficacy (Armor et al., 1976), Gibson and Dembo (1984), Bandura (1997), and Tschannen-Moran and Hoy (2001) created longer and more reliable surveys, exploring the relationships between the previous concepts of PTE and GTE and Bandura's concepts of self-efficacy and outcome expectations. In particular, Bandura shifted the measurement of teacher self-efficacy considerably by creating an instrument that, more in line with his conception of self-efficacy, attempted to capture teacher efficacy across different types of tasks. Teachers' efficacy, he posited, might be strong in one area, such as student discipline, but weak in other areas, such as instruction. These additional scales, based on a wide variety of more specific teacher tasks, were introduced by Bandura (1997) and further developed by Tschannen-Moran and Hoy (2001).

Tschannen-Moran and Hoy (2001) refine the more general concepts of PTE and GTE into the analysis of specific teaching task and context and the assessment of personal teaching competence within these contexts. Analysis of the teaching task, in keeping with Bandura's focus on the task-specific nature of self-efficacy, considers teachers' perception of a specific category of teaching practice and not the overall capability of the teaching role itself, as would have been examined by the prior construct of GTE. As Tschannen-Moran and Hoy (2001) describe it, the strength of their model is in the combination of *agent-means* perceptions—

perceptions of personal teaching competence and self-efficacy—with *means-end* perceptions—perceptions of the contingent resources and constraints for specific teaching tasks.

This aggregation of competence and contingency beliefs plays out in the phrasing of questions of the TSES that conflate personal competency and situational contingency, as in the question, “How much can you do to get through to the most difficult students?” (Tschannen-Moran and Hoy, 2001, p. 800). This item focuses on personal teacher competence (what can “you” accomplish) while acknowledging the constraints of the context, in this case the reality of “difficult” students. Other items of the TSES continue this general pattern, combining the constraints of context with reflection on personal competence.

Figure 8 below summarizes this relationship between the analysis of the teaching task and personal assessments of competence, as well as connecting teacher efficacy back to the theoretical sources of self-efficacy, mastery experience and physiological arousal.

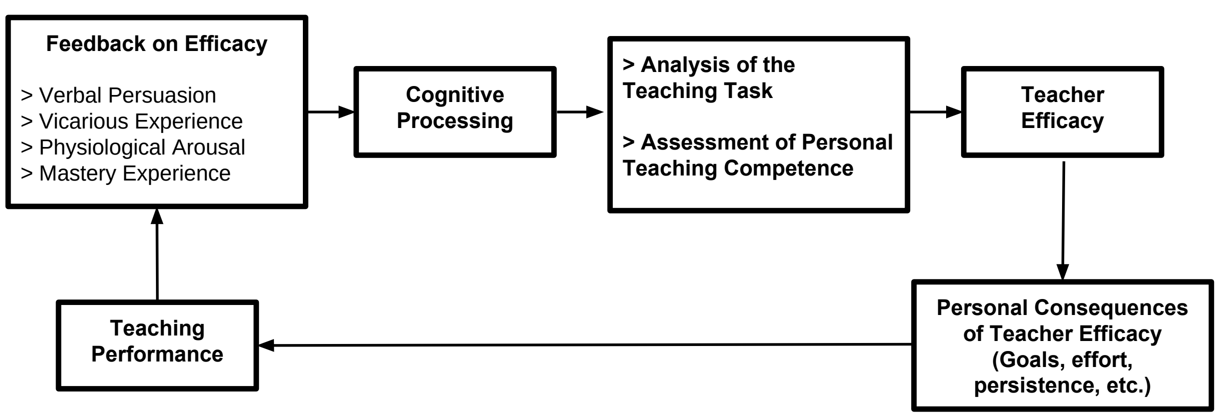


Figure 8. The cyclical nature of teacher efficacy. Adapted from Tschannen-Moran, Hoy, and Hoy (1998)

Over time, several layers of outcome have been connected with teacher efficacy. Researchers have linked teacher efficacy to important teaching behaviors, such as not criticizing students who answer incorrectly, continuing to assist students who have failed at a task, and dividing the class into small groups for instruction. All these behaviors are connected to the effort that teachers will expend in teaching, their persistence against obstacles, and ultimately to teachers' willingness and tendency to use data to inform changes in instruction. Other researchers have found teacher efficacy to be related to experimentation in pedagogy, willingness to try innovative methods, the ability to adapt practice (Ashton and Webb, 1986; Fuchs, Fuchs, and Bishop, 1992; Haney, Wang, Keil, and Zoffel, 2007; Timperley and Phillips, 2003), and a willingness to work with students who experience difficulties (Meijer and Foster, 1988; Podell and Soodak, 1993; Soodak and Podell, 1993), all parallel behaviors to the intentional use of data to inform instruction. Additional studies have provided evidence of relationships with various student outcomes, including students' own self-efficacy beliefs, student engagement, and achievement (R. N. Anderson, Greene, and Loewen, 1988; Midgley, Feldlaufer, and Eccles, 1989; Ross, 1992; Shahid and Thompson, 2001).

Teacher self-efficacy instruments. Two of the instruments recently used for capturing Teacher Self-Efficacy are the Teachers' Sense of Efficacy Scale Short Form (Tschannen-Moran and Hoy, 2001, p. 800) and the Norwegian Teacher Self-Efficacy Scale (Skaalvik and Skaalvik, 2007). As discussed above, Tschannen-Moran and Hoy addressed a need in the assessment of teacher self-efficacy by creating and thoroughly testing a multi-dimensional instrument designed to capture the self-efficacy construct. In the creation of the Norwegian teacher self-efficacy scale, Skaalvik and Skaalvik (2007) expanded the number of scales assessed, matching the dimensions of efficacy under consideration to the role expectations for teachers in Norwegian schools. In addition, they re-introduced a separate measure of the degree to which teachers feel that their performance is constrained by external forces, proposing that the external control or contingency factor of teacher self-efficacy is important in understanding overall teacher efficacy, but is better viewed as a distinct component from the competency factor (Skaalvik and Skaalvik, 2007, 2010).

Teacher self-efficacy for data-driven decision making. Building on work in teacher self-efficacy, Dunn, Airola, Lo, and Garrison (2013b) have delved into teachers' feelings of efficacy specifically related to the realm of DBDM. Their Data-Driven Decision-Making Efficacy (DDDM) and Anxiety Inventory (3D-MEA) takes the domain specificity of the TSES and NTSES one step further, by developing efficacy scales for specific data use tasks. Dunn et al. (2013b) explicitly model their approach on Tschannen-Moran and Hoy's development of the TSES and, therefore, on Bandura's framework for self-efficacy. Teachers' sense of efficacy for DDDM is defined by the authors as "teachers' beliefs in their abilities to organize and execute the necessary courses of action to successfully engage in classroom-level DDDM to enhance student performance" (2013b, p. 88). The goal of the authors in developing the 3D-MEA was to explore the factors inhibiting or encouraging the adoption of data-based decision making practices in schools, taking up the proposal that more needs to be done to understand teachers' psychological states in relation to data usage. Dunn et al. (2013b) developed five subscales (see Table 9) with teachers in the Pacific Northwest. These five scales were later confirmed in work with teachers in a Midwestern state (Walker, Reeves, and Smith, 2016). Discriminant validity between the 3D-ME (an earlier version of the 3D-MEA) and the TSES is also established with correlation coefficients between the two measures ranging from -.02 to .27 (Dunn, Airola, Lo, and Garrison 2013b).

Table 9

Data-Driven Decision Making Efficacy and Anxiety (3D-MEA) Subscales

Subscale	Cohen's Alpha
Efficacy for data identification and access	.84
Efficacy for data technology use	.91
Efficacy for data analysis and interpretation	.81
Efficacy for application of data to instruction	.92
DDDM Anxiety	.88

Note. Scales use a five-point Likert Scale from Strongly Disagree (1) to Strongly Agree (5)

Research utilizing the 3D-MEA has found the subscale "Efficacy for application of data and instruction" to significantly contribute to a multiple regression model predicting teachers' data use practices (Reeves, Summers, and Grove, 2016). Other strong factors in this model included survey items assessing the belief that assessment improves teaching and an item indicating whether teachers had taken an undergraduate course in data use or data-based decision making. Dunn et al. (2013b) incorporate four subscales of the 3D-ME (an earlier version of the 3D-MEA) into a Structural Equation Model supporting links between DDDM Efficacy and advanced stages of teacher adoption of DDDM, as assessed by the Stages of Concern Questionnaire (George, Hall, and Stiegelbauer, 2006) for implementing a pedagogical innovation.

While one previous study has explored relationships between teaching self-efficacy and teachers' adoption of an online data system (Shaw, 2010), as far as I can determine, no studies have examined the relationships between the more specific domain of teachers' DBDM efficacy, general data use attitudes, and behavior in an online data platform. The current study will explore the relationships between teacher self-efficacy, both general and specific to DBDM, as well as other teacher attitudes towards data usage, and teachers' system usage behaviors. Leveraging the construct of self-efficacy may help researchers better understand the motivational determinants of DBDM practice among teachers. On an applied level, research into general teaching and DBDM self-efficacy may help schools create professional development strategies that overcome obstacles to the adoption of DBDM. Knowledge of teachers' self-efficacy as it relates to DBDM may also provide important preliminary feedback to teacher preparation programs as they incorporate and evaluate new training for teachers in data analysis and usage.

Technology acceptance and use diffusion frameworks. Developed within the Management Information Systems (MIS) discipline, the Technology Acceptance Model (TAM) is one of the most widely used frameworks for understanding users' acceptance of information technology (IT) systems (Venkatesh, Thong and Xu, 2016). As such, it is useful for better understanding the factors influencing teachers' adoption of online systems for student data. Broadly speaking, TAM proposes that the intention to use technology, as well as its actual use, are influenced by the perception of how useful that technology will be in helping users perform their related tasks (Davis, 1989; Ali, Asadi, Gasevic, Jovanovic and Hatala, 2003; McFarland and Hamilton, 2006). Along with this construct of "perceived usefulness" (PERUSE), TAM incorporates the influence of users' beliefs that using the technology will be free of effort, referred to as "perceived ease of use" (PEOU). While both constructs are found to impact the intention to use technology (Davis, 1989), perceived usefulness is sometimes posited to act more directly on behavior and behavioral intention, with perceived ease of use acting indirectly through perceptions of usefulness (Venkatesh, Morris, Davis, and Davis, 2003). Questions remain, however, as to the degree to which PERUSE and PEOU relate to the technology system itself or to the task completed in the system. This untangling of task-related perceptions as opposed to computer- or system-related perceptions is a longstanding question in TAM research (McFarland and Hamilton, 2006).

At its inception, TAM adapted the Theory of Reasoned Action (Fishbein and Ajzen, 2010) to the particulars of technology acceptance. While Fishbein and Ajzen (2010) describe and explain extremely generalizable processes connecting attitudes and behavioral intentions to actions, one of the strengths of their approach is the insistence on adequately specifying not just a general action (e.g. data use), but the target of that action, its context, and time (Ajzen and Fishbein, 1977). The related construct of self-efficacy, discussed above, similarly depends on adequately specifying the task or skill in question (Bandura, 1986).

The current study will leverage both the constructs of PERUSE and PEOU as they apply to various elements of DBDM process. For example, teachers' perception of usefulness (PERUSE) will be investigated in relation to their perceived usefulness of the data system itself, the usefulness of types of available data, and the usefulness of classroom activities related to DBDM.

Perceived ease of use (PEOU) will be investigated in relation to teachers' perceptions of their general comfort with technology, their levels of expertise with the system itself, and their perceptions of self-efficacy in regard to multiple data-use tasks. Types of self-efficacy, such as computer self-efficacy, have been identified as one of the most significant contributors to the construct of PEOU (Venkatesh, 2000). Available DBDM-specific measures of self-efficacy (3D-MEA) are therefore used as proxies for PEOU in the current study. Using these DBDM-specific measures allows consideration of users' efficacies across the range of skills necessary for effectively accessing the student data platform, skills of data interpretation, analysis, and use, along with system-related ease of use.

However, the acceptance of a technology innovation is only part of the story, and studies have pointed out the limitations of focusing only on initial adoption instead of ongoing use (Kim and Crowston, 2012; Maull, 2013, Lee and Recker, 2014). Use diffusion theory (Ram and Jung, 1990; Shih and Venkatesh, 2004) supplies a framework for examining the ongoing use of adopters through the dimensions of the rate of use and the variety of use. Rate of use refers to time spent using a product, while variety of use refers to the multiple ways in which the product is used (Shih and Venkatesh, 2004). Combining high and low levels of rate and variety produces a typology of users represented in the grid below.

		Rate of Use	
		High	Low
Variety of Use	High	Intense Use	Nonspecialized Use
	Low	Specialized Use	Limited Use

Figure 9. Use diffusion user typology

Maull (2013) applies use diffusion theory in the context of teachers' use of an online science curriculum. In exploring methods for operationalizing the construct of variety, or the "range of use of a technology" (Maull, 2013, p. 60), the author first applies forms of the Shannon entropy calculation from information theory to generate a course-grained model, dividing users into the four quadrants of the use diffusion typology. While this computational approach was useful in developing methods that might apply across a variety of systems, Maull (2013) goes on to apply a finer-grained, clustering-based approach, still within the use diffusion framework. While segmenting the analysis along the same constructs of frequency and variety, Maull includes a larger number of features for each, expanding the construct of “variety” to include eight features that capture usage across various functions of the curriculum system. This approach yielded a fine-grained typology of user behaviors, more usefully contextualized in the specific functions of the system. Instead of one set of “Specialized” Users, identified in the quadrant-based analysis, the clustering-based approach discovered two types of specialized use, one focusing on the use of interactive resources and one on the community-based functions of the system (Maull, 2013). It may be that the application of a clustering as opposed to quadrant-based approach to the analysis of use diffusion factors could prove useful in other situations for establishing typologies of teacher online use.

Existing typologies of teacher data and technology use are either based solely on frequency of use, dividing teachers into categories of light to heavy use (Gold et al., 2012), or are based on qualitative analysis (Bill & Melinda Gates Foundation, 2015), which may be difficult to apply to new teacher observations. The application of a use diffusion framework to the analysis of teacher data use and attitudes may provide a valuable means of exploring relevant subgroups of teacher behavior.

Expanding participation. In recent years, researchers within the learning analytics community have enthusiastically generated frameworks for integrating research related to DBDM into the everyday experience and practice of educators. Integrating trends in education driven by technology, online networks, and information systems (Piety, 2013), these frameworks have expanded the role for practitioners in education research. These syntheses have produced multiple conceptual frameworks, cataloging and framing clusters of research and practice across EDM/LA practice, DBDM research, and other disciplines.

Educational data science. In one of the earliest of these frameworks integrating questions of research and practice, Buckingham Shum, Hawksey, Baker, Jeffery, Behrens, and Pea (2013) reflected on the scarcity of the Educational Data Scientist in the wild. Piety, Behrens, and Pea (2013) and Behrens, Mislevy, Piety, and DiCerbo (2013) systematically sketch out the lines of a broad sociotechnical movement, identifying shifts in the type and quantity of evidence available for analysis in education, as well as qualitative shifts in educational practice. They describe how a forthcoming "digital ocean" of artifacts will combine with three educational shifts—a movement away from institutional control, along with a shift towards a wider range of competencies and blended and personalized learning. To meet the challenges of investigation in this new era, Piety et al. (2014) suggest that the Educational Data Sciences draw from the established fields of statistical data analysis, classroom/learning technology, learning sciences, information sciences, organization and management science, and decision science, all supported by foundational methods of computer science (Figure 10).

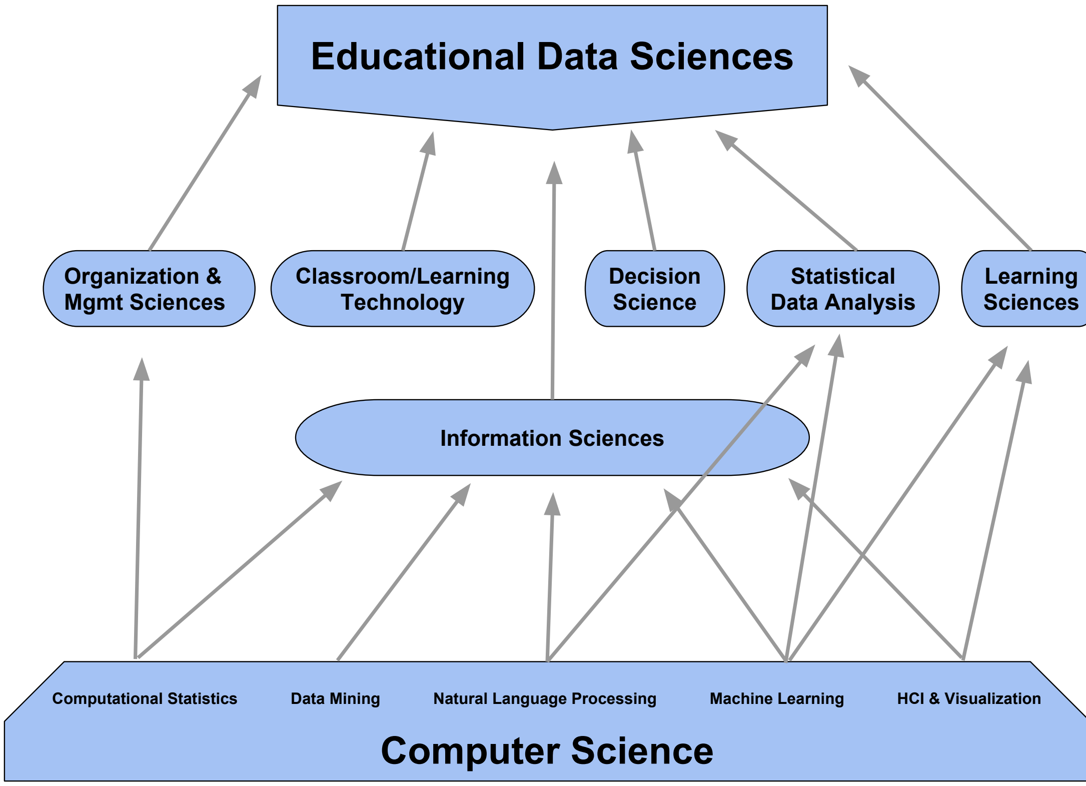


Figure 10. Education data sciences interdisciplinary connections. Adapted from Piety, Hickey, and Bishop (2014)

Having identified a broad picture for the future of the Educational Data Sciences, Piety et al. (2014) go on to identify four emerging communities arising from these same sociotechnical shifts: Academic/Institutional Analytics, EDM/LA, Learner Analytics, and Systemic/Instructional Improvement. Significant areas of overlap exist between the four communities and Piety et al. (2014) suggest that the boundaries of these communities will continue to blur, recognizing that they are differentiated by the scale at which they examine educational context (individual to system) and the stage of learning (early childhood to career) but united in their embrace of larger sociotechnical trends in education.

Education leadership data analytics. In a similar vein, work by Agasisti and Bowers (2017) and Bowers (2017) focuses on the intersection of data-intensive approaches with educational leadership, advocating and explaining the emerging professional role of the Educational Data Scientist. Some of the important themes from earlier work carry over in the goals for this position, including the need to facilitate communication between technical experts in data analytics, the education decision-makers at the school and district level, and teachers themselves. As the domain of the Educational Data Sciences spans from the individual student to the school district (Piety, Hickey, and Bishop, 2014), so the Educational Data Scientist must be prepared to facilitate across all these levels (Agasisti and Bowers, 2017).

Collaborative data-intensive improvement. Continuing to elaborate and refine specific areas of Educational Data Science, Krumm, Means, and Bienkowski (2018) provide an extensive treatment in the aptly named *Learning Analytics Goes to School*, drawing together multiple research examples into an applied framework for learning analytics. Krumm et al. (2018) continue to emphasize the need for combining data-intensive topics and methods (EDM/LA) with research on data use in schools (DBDM), while adding an important focus on collaborative research approaches. By incorporating work in design-based research and improvement science, Krumm et al. provide a highly useable framework for co-developing data-intensive solutions for schools.

Excitingly, these new research approaches are providing DBDM with an expanding set of frameworks for investigation. Educational Data Science, specifically, along with the larger EDM/LA communities offer pathways for the integration of complex data analytics methods and products into previously defined DBDM logic models, often in the form of visualizations that use data-intensive methods to present information in ways intended to leverage human judgment.

Bienkowski et al., (2012) groups similar approaches, from visual searches for patterns to data dashboards, as visual data analytics aimed at distilling trends from complex data sets. Such approaches offer techniques for the expansion of participation, as the audience for data-intensive methods becomes educational decision makers themselves, both at leadership and instructional levels.

Work in visual data analytics and EDM/LA visualizations, in turn, relies on broader frameworks for understanding and generating effective visualizations, frameworks that apply in a much wider range of cases, from the humble bar chart to complex high-dimensional representations. Topics in data visualization are many, from the cognitive foundation of humans' perception (Ware, 2008; 2013), to human-computer interaction techniques for matching suitable visual idioms to tasks (Preece, Rogers, and Sharp, 2015). Visualizations do not operate in a contextual vacuum, though, and work by Munzner (2015) provides a useful structure for distinguishing the different levels at which design decisions can impact the effectiveness of visualizations, including the domain situation, the task abstraction, the visual encoding or idiom, and the algorithm. The first two of these categories notably deal with the task of determining the visualization needs of the particular context and the effective means of abstracting data for those contexts. The second two levels deal with more specific questions of how these abstractions become a visual reality (Munzner, 2015), levels which often receive more focus in the visualization field.

These dissertations studies, conducted in one local educational context, yet with data-intensive methods, provides a possible example of Educational Data Science at work, attempting to leverage an expanded set of behavioral theory with data-intensive methods to generate transferable understandings of data use, as well as concrete tools for practitioners ready to interrogate their own local context of DBDM.

Conceptual Framework Summary

In the previous sections, I identified a core need of DBDM research: to improve understanding of the enablers and barriers to teachers' use of student data. I then described three expansions of research to address that need: expansions of methods, theory, and participation. The two studies of this dissertation implement these expansions in order to explore a key question of current DBDM practice: *How are teacher roles, attitudes, and efficacies related to their online use of student data?*

While research into DBDM in schools has produced fascinating and extensive qualitative work and a rich literature of guidance for educators, results from large-scale quantitative and experimental studies have been mixed as to the key dynamics and impacts of data use. Large-scale survey studies have left the majority of the variation in teacher data use unexplained (Shaw and Wayman, 2012; Tyler, 2013). Additionally, the results of studies evaluating data use interventions themselves have been mixed as to the effectiveness of these interventions for increasing teachers' use of data or for increasing student achievement outcomes (Wayman et al., 2017; Tyler, 2013). Given this lack of certainty in the literature, these two studies expand the range of possibility for DBDM factors and outcomes, exploring data use through the lens of methods and theory commonly applied in the EDM/LA communities. Instead of attempting to confirm existing hypotheses of data use, Studies 1 and 2 work towards generating and exploring high-dimensionality, quantitative descriptions of data use in a local school context. Rich, quantitative descriptions of teachers in one school—their roles, their attitudes towards data use, and their use of an online data system—aim to identify additional measures and determinants of teacher data use, as well as generate practical tools for future use in schools. In this section, I will summarize the conceptual framework for these two studies, as well as discuss some of the strengths and limitations of the selected approach to research, one that is descriptive, local, and small-scale.

Conceptual framework. This dissertation draws heavily on methods applied in the EDM/LA communities and on theories from cognitive psychology and information science to supplement the traditional methods and questions of DBDM research. These methodological and theoretical additions are discussed extensively in the previous section. In summary, Study 1 applies use diffusion theory and learning analytics methods for log file analysis to describe and visualize teachers' online interactions over time and to generate and explore a range of outcomes for online behavior. Study 2 uses data-intensive visualizations to explore the possibilities of teachers' roles, self-efficacy and technology acceptance for explaining variation in DBDM practice. Figure 11 provides a visual summary of the overarching approach of this dissertation.

While the context of this study—the teachers of one school using assessment data over the course of a semester—is a core setting for DBDM practice, as a setting for research it raises several questions. The most critical of which, may be, "Why use rich quantitative description as a framework for investigation?" and "Why conduct an investigation in only one school with a small population of teachers?"

I will attempt to address both of these questions next.

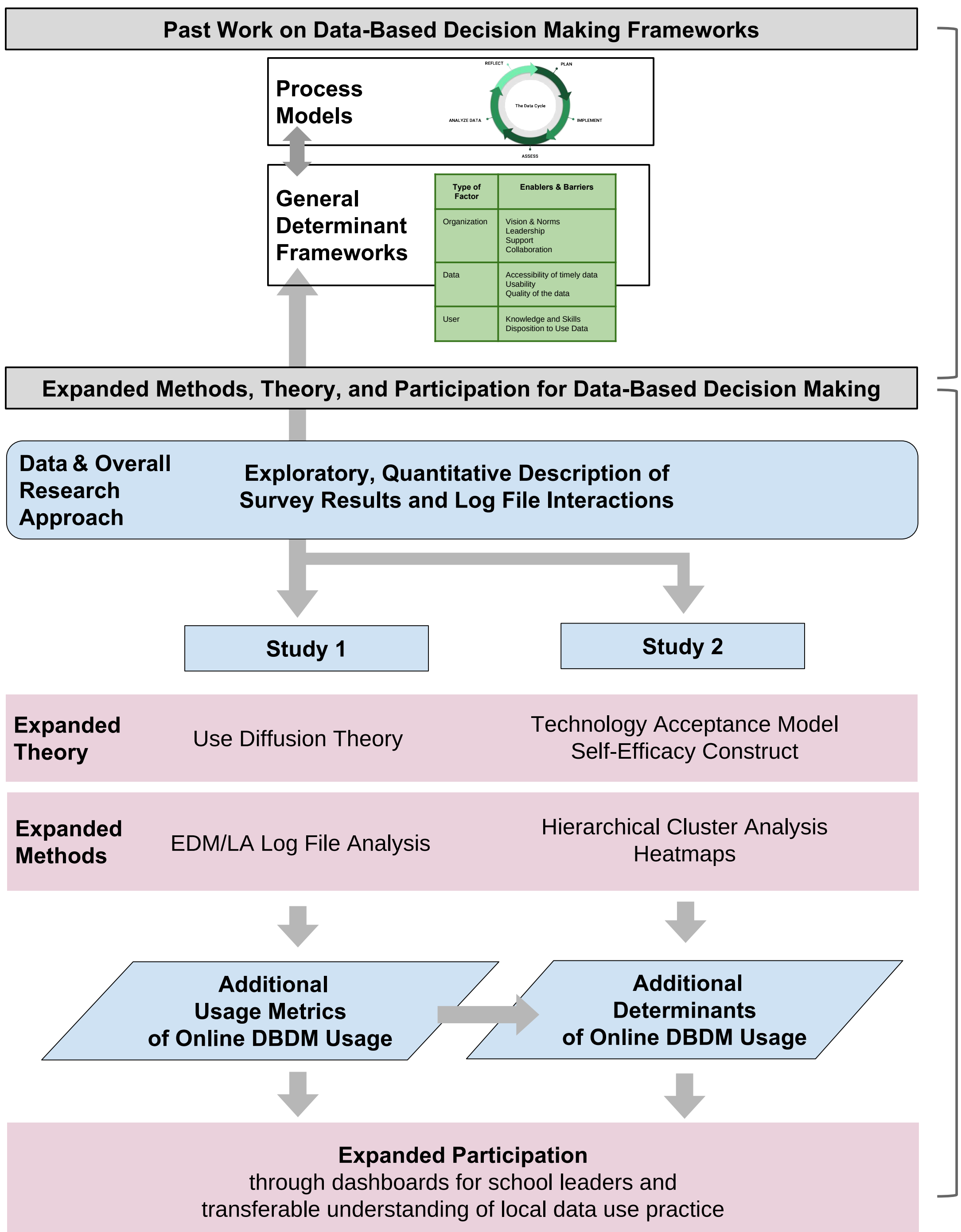


Figure 11: Expanding methods, theory, and participation for data-based decision making

Why use rich quantitative description as a framework for investigation? Loeb et al. (2017) points out that while causal research may seem more prominent, most research is, in fact, descriptive. For any research problem, sufficient description is essential: establishing the quantitative landscape of who, what, where, when, and to what extent informs the possible impacts of interventions, and even which interventions are needed in the first place (Loeb et al., 2017). While such descriptive work on its own cannot establish causation, when properly applied it can, "prioritize causal mechanisms, generate hypotheses and intervention strategies, interpret the findings of causal research, diagnose problems for practitioners and policymakers to address, and identify new issues to study" (Loeb et al., 2017. p. 1). Particularly where large datasets are available, descriptive work can distill and reveal actionable patterns (Loeb et al., 2017).

Other researchers have also emphasized the importance of quantitative description for both scientific and practical ends. Work by Bowers, Blitz, Modeste, Salisbury, and Halverson (2017) has suggested the term "quantitative phenomenology," to describe research that provides a "rich contextual analysis of relationships" (Bowers et al., p. 50). While this approach has in the past been applied to national survey data (Bowers et al., 2017; Graves and Bowers, 2018), the current study follows a similar overall pattern: exploring a specific phenomenon through disaggregated visualizations and person-centered methods for patterning across a range of attitudes, behaviors, and characteristics. As a form of descriptive analysis, such work shares the larger goals of descriptive studies to inform the generation of new theories and hypotheses, as well as applied decision-making.

A second related approach has been discussed in organizational science by March, Sproull, and Tamuz (1991) in their provocatively titled, "Learning from Samples of One or Fewer." The authors describe how organizations, in order to learn, must interrogate unique, local events (samples of one) for actionable meaning. One of the ways they attempt to decrease the variability and error around the interpretation of these events is by examining an individual case with greater intensity, creating a richer description by "aggregating over multiple observers" (March et al., 1991). While methods of analysis in March et al. are only generally described, the overall themes of increased intensity and variety of observations appear to apply to the current study as well. March et al.'s (1991) proposal for generating more reliable understandings at a local, organizational level recognizes that organizations, public schools included, already use, and will continue to use, their local knowledge to generate understanding and make decisions, whether or not that knowledge was generated with methods that met scientific criteria for reliability, validity, or generalizability.

In relation to the current study, the landscape of teachers' interaction with online systems is still sufficiently unknown so as to make an exploratory description of these interactions informative. When combined with survey data mapping the landscape of teachers' data use attitudes and school roles, the current studies offer the possibility of capturing meaningful dimensions and patterns of teachers' data use in schools.

Why investigate in only one school with a small number of teachers? The two studies of this dissertation investigate one of the fundamental contexts of data-based decision making: teachers in one school using assessment data over the course of a semester. The small scale of this study, both in terms of context (one school) and participants (approximately 40 core content teachers) is both a strength and a limitation. Since the goal of this research is to generate new hypotheses as to the determinants of data use in schools, a solidly-contextualized, longitudinal, multi-dimensional description of data use and attitudes in one school may be a better fit than a larger-scale approach, likely to overlook the complicated nature of within-school roles and structures. To borrow an anthropological term, I attempt to provide, a "thick" (Geertz, 2017) yet quantitative description of data use phenomenon in one school, with the hope that this style of description may provide more useful insights than a less-contextualized, “thin” description of data use on a larger scale.

The sacrifice, of course, is that an in-depth understanding of a small-scale phenomenon can fail to generalize to a larger population or achieve statistical significance in its results. This failure to statistically generalize, however, does not mean that the patterns and themes of observation gained through small-scale, contextual studies do not retain “transferability” (Yanchar, 2006) for other researchers or practitioners searching for patterns in what they deem similar circumstances. Particularly, with the advent of instructional software and online data systems, the sources of data for even small numbers of participants can be exceedingly rich, leading to investigation of transferable patterns based on what Geertz might have deemed “exceedingly extended acquaintance with extremely small matters” (Geertz, 2017, p.23).

Along with these more theoretical justifications of the value of small-scale research, considerations of practicality and feasibility are important to note as well. Perhaps most importantly, by holding the school context constant, these studies are able to investigate how the variation in teacher-level factors is displayed against a unified backdrop of school-based scheduling, teams, training, and leadership. Keeping the school context constant allows for increased exploration and description of the heterogeneity within that school. Another result of limiting these studies to one school is that the data represent 83% - 89% of the core content teachers in the school and are generally complete across variables. As opposed to a sample of users from a wider population, these studies come close to representing the entire population under investigation. This focus and completeness of data allow for a more thorough investigation of how teachers’ data use attitudes and actions interact within the complex internal structures of a school.

Additionally, the study of Progress Secondary School provides a unique opportunity to study teacher data use for several reasons, not the least of which is that my experience with the school allows for a more contextual presentation of school structures and processes. Along with this opportunity for a more richly described context, the school's implementation of nine-week interim testing appears similar to the interim testing process commonly described in data use literature, making investigation of data use in this one school more transferable to interim testing contexts on a wider scale.

Also, as a combined middle and high school, Progress Secondary provides a context where some overarching school factors are constant, while other school systems differ between middle and high schools. Overarching characteristics, such as data systems, school location, student demographics, training systems, team structures, and schoolwide leadership are held constant across schools, while at the same time, middle and high school have some different characteristics due to participation in different state-level testing regimes and accountability requirements. While one school for most purposes, Progress Secondary provides a valuable lens on how the attitudes and data use of middle and high school teachers may differ as they react to differences in statewide accountability and assessment structures. With school-level one of the only school factors consistently related to teacher data use (see Table 6), and almost no examination of observed online data use at the high school level, this ability to investigate both middle and high school data use side by side is particularly valuable.

While limiting investigation to one urban secondary school hinders the generalizability of these studies, that same limitation comes with the opportunity to describe and explore teachers' online usage and attitudes towards data within a rich, school-level context, often unavailable in larger-scale studies. By richly describing "what is" through the lens of relevant theory these studies aim to generate new hypotheses, measures, and tools to inform future research and implementation of DBDM in schools.

As the overarching conceptual framework (Figure 11) suggests, the work of Studies 1 and 2 can be thought of a series of expansions to the DBDM literature, of its methods, theory, and participation. While the general goal of these expansions is to extend current determinant frameworks of teacher data use in schools, three specific research questions guide this work across the two studies.

Study 1 primarily addresses research question one (R1): **“**To what extent and in what ways do teachers use online data and assessment tools?” Past studies of teachers’ online use, log file analysis, use diffusion theory, and methods for visualization guide response to this question.

Study 2 addresses the second research question (R2) “What attitudes do teachers have towards data use?” Survey responses are viewed in relation to established subscales of data use attitudes, as well as in relation to the perceived usefulness and perceived ease of use constructs of the Technology Acceptance Model. After examining teachers’ data use attitudes, Study 2 turns to the third research question (R3) “To what extent are teachers’ data use attitudes, technology acceptance, self-efficacy, and roles related to their online use of student testing data?” Exploratory correlations and various methods for data-intensive visualization are used across Study 2 to address research questions two and three.

To expand the participation of school leadership in ongoing efforts to better understand and implement initiatives in DBDM, Study 2 also presents dashboards of teacher data use and attitudes, drawing from analyses in both Studies 1 and 2 and geared towards data use at a schoolwide level.

Study 1 has three main goals:

1. Expand the quantitative descriptions of teacher interaction with online data systems, particularly in regard to visualizations and metrics for online behavior,
2. Explore subgroups and patterns of teacher online behavior along dimensions of frequency, consistency, and variety of use, and
3. Inform school practice and software design to facilitate teachers' access and use of student data.

Study 2 builds on the results of Study 1, adding several related goals:

1. Describe relationships between data use attitudes, self-efficacies, and teachers' roles,
2. Identify subgroups of attitudes, self-efficacies, and technology acceptance in relation to teacher roles and online use of student data,
3. Explore possibilities for organizing teacher roles, self-efficacies, and technology acceptance into a determinant framework for online use of student data, and
4. Suggest possibilities for guiding school practice in ways that improve teachers' use of student data for instructional decisions.

Overall, results suggest that to facilitate teachers' engagement in DBDM, schools should shift from accountability -driven, "one-size fits all" approaches to differentiated approaches to data use, varied according to teachers' perceptions of usefulness, content-area needs, and the internal, professional structures of schools.

By expanding the methods, theories, and participation of DBDM research, this dissertation aims to assist both researchers and practitioners as they work to facilitate the adoption and improve the effectiveness of data use systems in schools.

Study 1: Exploring Teachers' Online Usage of Student Testing Data

Summary and Purpose of Study

During the 2015-16 school year, Progress Secondary implemented a new data and assessment platform: Benchmark Data (a pseudonym). As part of that implementation, log files from January 2016 until the end of the school year in June 2016 were analyzed and visualized with methods similar to those of the few comparable studies (Gold et al., 2012; Tyler, 2013; Wayman et al., 2009a, Wayman et al., 2017), and with additional analyses drawn from the EDM/LA community. These log file analyses attempt to add to the limited available public research which describes teachers' use of online data systems. This initial and exploratory study attempts to address the following research question:

(R1) To what extent and in what ways do teachers use online data and assessment tools?

Research Context

The school. As described earlier, Progress Secondary School was located in an urban area of New York State and served approximately 500 middle and high school students at the time of the study. The student body included large percentages of minority, low-income, and special education students and was divided into a middle school, serving grades 6-8 and a high school, serving grades 9-12. While sharing a location and overarching leadership, each of these two schools followed different assessment, professional development, and student schedules.

Assessment at Progress Secondary. At Progress Secondary middle school, formative interim assessments were administered in both ELA and Math three times over the course of the 2015-16 school year prior to end-of-year state testing, which took place in May of 2016. All interim assessments were administered on paper, with multiple choice answer sheets printed from and scanned directly into the Benchmark Data system using document cameras.

ELA interim tests consisted of approximately five reading passages in a variety of genres, each followed by four to six multiple choice items testing reading comprehension. ELA assessments also included a writing task, scored by teachers according to the same rubrics used to assess student writing during the state test. In addition, shorter ELA progress monitoring quizzes were administered weekly and also scanned directly into the Benchmark Data system. Items for both the ELA interim assessments and progress monitoring quizzes were selected from sets of previously released items from past state tests. The specific items for ELA interim testing were selected by administrators in collaboration with middle school ELA teachers.

In Math, middle school interim tests were administered on a similar schedule to ELA, three times over the course of the school year. Prior to each interim testing session, Math teachers across the middle school grades identified the state standards that they had covered to that point in the school year. In collaboration with Math teachers, administrators then selected items representing these standards from previously-released state testing items, as well as from the standards-aligned item bank provided by Benchmark Data. Each Math interim test consisted of approximately twenty-five multiple choice items and five to seven constructed response items. Constructed response items were scored by Math teachers according to state testing rubrics.

In Progress Secondary high school, interim testing was organized around the New York State Regents testing system, a system of high school course exit exams required for graduation (Part 100 Regulations, 2018). For each high school course which prepared students for a state Regents exam, interim tests were administered four times throughout the school year. These Regents-preparatory courses included English, Living Environment, Earth Science, Global

History and Geography, U.S. History, Algebra 1, Geometry, and Algebra 2/Trigonometry. In English, interim tests were administered in 9th, 10th, and 11th grade in preparation for the English Regents administered at the end of the 11th grade year.

For each of these high school interim assessments, teachers identified the standards they had taught to that point in the year. School administrators then selected items for the assessment in collaboration with the course teachers. All items, both multiple choice and constructed response, were selected from past New York State Regents exams.

Benchmark data and assessment system (Benchmark Data). During the 2015-16 school year, the Benchmark Data system allowed for the collection, distribution, and analysis of student testing data, particularly interim testing and progress monitoring quiz results, but also historical state test and computer adaptive test scores. Through this online system, teachers and other school staff had the ability to view student data individually or in class or demographic groupings, create reports that combined multiple data sources, view assessment results in a variety of formats and analyses, and create and administer their own assessments using either items provided within Benchmark Data or their own items.

The process of sharing interim testing results was conducted both through email, by sharing links to Benchmark Data assessment reports, and in person, by conducting collaborative professional development sessions, where teachers accessed interim assessment results on laptops with guidance from school administrators. Staff were provided with training on multiple occasions over the course of the school year in the use of the data system.

Across both schools, several teacher leaders were involved, not just in the analysis of testing data after each administration, but also in the preparation of exams and the management of scoring within the Benchmark Data system. These preparations may have included activities such as creating tests with appropriate answer keys and items weights, assigning state standards to items, printing student answer documents, scanning student answer documents, and generating score reports in the Benchmark Data system.

Because the Benchmark Data system provided access mainly to testing results, the assessment schedule for both the middle and high school provided the overarching context for teacher access to the system throughout the school year. The list below details school year dates relevant to the use of the Benchmark Data system for the both the middle and high schools.

***2015-16 Relevant School Dates*.**

1/11 - 1/15	HS Interim Testing Administration
	ELA and Social Studies Interim Testing Administration, HS Data Team Meeting
1/26 - 1/29	HS State Regents Testing Administration (January)
	MS Interim Testing Administration
	Benchmark Data Training – for administrators
2/1	MS/HS Benchmark Data Training
2/8	HS Interim Testing Results emailed to teachers
2/15 - 2/19	Midwinter Vacation
3/3 - 3/4	MS Interim Testing Administration
	HS Benchmark Data Training
3/7	MS Interim Test Results Emailed to teachers
3/18 – 3/21	HS Interim Testing Administration
3/19	HS Drop Out Predictions Emailed to teachers
3/26 – 3/28	HS Interim Test Results emailed to teachers
	MS State Test Predictions emailed to teachers

4/5 - 4/7	MS ELA State Test Administration
4/13 - 4/15	MS Math State Test Administration
4/25 - 4/29	Spring Vacation
5/10 – 5/12	HS Interim Testing (Science, English) Administration
5/23 – 5/25	HS Interim Test (Math, Social Studies) Administration
	HS ELA Interim Testing Results emailed to teachers
6/1 – 6/4	HS ELA, Math, Social Studies and Science Interim Testing results emailed to teachers
6/14 – 6/22	HS State Regents Testing Administration (June)
6/28	Last day of school

Implementation of Benchmark Data began just prior to the period of this study, in the fall of 2015. Unfortunately, log file records of staff usage of Benchmark Data were only available beginning in January of 2016. Ideally, analysis would have included the initial months of implementation, but technical issues with the logging of the data made this impossible.

Data Available in Benchmark Data. In general, teachers were able to access student testing data for students enrolled in their classes. The following general types of data, outlined in Table 10, were available to teachers and other staff through Benchmark Data.

Table 10

Benchmark Data Metrics and Data Categories

General Category	Description	Metrics and Fields
Student Demographics and Classes	Basic student data	Student ID number, Birth date, Grade level, Courses, English language learner and Special Education status
Middle School State Test Scores	ELA and Math State Test Scores starting from 5th grade	Various metrics, including Performance level, Scale score
Regents Test Scores	Regents test scores from all courses and all attempts	Testing course, Date of test, Scale score
Regents Testing, Item-level Results	Item-level responses from past Regents testing	Test name, Date of test, Item-level metrics and linked standards
Computer Adaptive Test Scores	Computer adaptive score history in reading and math	Test subject, Date of test, Scale score, Grade Level Equivalent score, Test gain
Interim Testing and Progress Monitoring Results	Interim testing and progress monitoring reporting	Test name, Date, Standards alignment, Percent correct, Performance by item, Performance by standard
Early Warning Indicator Flag	High School early warning Indicator for dropout risk (Yes or No for Drop-out Risk)	Implementation of the Chicago Consortium 9th grade early warning indicator, (Allensworth, 2013)
State Test Predicted Performance	Individual student predicted performance on ELA and Math Tests 2015-16	Linear regression prediction under a cross-validation framework

Methods

Data set and data filtering.

Benchmark data log file description. Benchmark Data log files for second semester—January 2016 through June 2016—contained approximately 8,000 rows of user interaction by 39 core content teachers, which were analyzed in the current study. Each user login to Benchmark Data began a user session, which was identified with a unique session ID. All personally-identifiable information had been removed from the data set prior to analysis. The following relevant fields were contained in the log file (Table 11).

Table 11

Log File Description

Column Name	Description	Example or List of Possible Response
timestamp	Date and time of each user action	*Example:* 2016-03-07T15:13:51.000Z
extra_user_id	De-identified numeric ID for user	*Example* 779
extra_php_ session_id	Unique ID for user session	*Example:* 35s2sa3smhiu8p8f40pd01****
message	General action category	*All Possible Responses:* Answer Sheet printed, Answer Sheet scanned, Page Viewed, User Login, User Logout, Widget Added, Widget Removed, Widget Settings Changed
extra_page_class	Specific page classification	*Examples (83 total possibilities):* MainWelcome, AssessmentListController, StudentProfile, Assessment_ReportController, Gradecam_SheetServiceController

To facilitate analysis, a small set of data cleaning activities were applied to the log files to minimize the inclusion of sessions or actions where the user was inactive or had effectively stopped using the system but had not logged off. Kovanovic et al. (2015) provides a helpful overview of the complexities of estimating time-on-task from log files, including approaches for estimating the length of the last action of a session. As a domain-specific example, when dealing with similar log files, Gold et al. (2012) imputed last action length based on the median action length for a particular task, performed by the appropriate user type, during the same week. Given the limited number of users and actions in the current data set, a simpler approach was taken. Expressed in terms of Kovanovic et al. (2015), the current study followed a protocol for limiting outliers (or abnormally long actions), while dealing with last-action estimation, by counting the time of last action as zero.

Any sessions lasting beyond four hours were assumed to be due to a failure to log off the system and truncated to the end of the last user action before the end of the session. Since the longest session length was two and half hours long, this rule was not applied. In approaching this same issue of inactivity, Gold et al. (2012) excluded all sessions lasting more than one hour. Maull (2013), in contrast, excluded sessions lasting less than 30 seconds and those lasting more than eight hours.

To avoid losing session and action level data, while minimizing the impact of user inactivity on analyses, I applied an action-level rule, where any actions lasting an hour or more, were truncated to half of their original length. The intention of this rule was to mitigate the impact on analysis of any extra-long actions, where the user may have stepped away from the system for some period. Since these actions were only truncated, any relevant information contained in those actions would not be entirely lost. Such pauses in online activity may be common in schools, where teachers' computer use is frequently interrupted by interactions with colleagues or students.

In multiple cases, the system initiated an automatic time out of user sessions after two hours of inactivity. These cases were treated according to the prior rule, where the user action immediately prior to the automatic time out was truncated to half its recorded length.

Usage metrics. A set of measures used in previous studies were generated from log files to capture teachers' online usage of the system (Gold et al., 2012; Maull, 2013; Tyler, 2013; Wayman et al., 2009a, 2011). Broadly speaking, these metrics included measures of prevalence, consistency, and variety.

Measures of prevalence or frequency. Previous measures of prevalence (Wayman et al., 2009a) and/or frequency (Shih and Venkatesh, 2004) appear to describe a similar set of metrics intended to capture the count and duration of users' activities. Such frequencies can be calculated at different scales of time (action, session, week, semester) and for different sets of activities. While frequency and prevalence metrics might refer to a broad range of time scales, they have tended to describe users' total use of a system over a set period. For this study, I calculated the following measures of frequency for each user, along with their summary statistics.

Counts (Min, Max, Mean, SD)

- Counts of total actions/clicks by session, week, and semester
- Counts of actions by type of user action
- Count of total sessions by week and semester

Durations (Min, Max, Mean, SD)

- Duration of total actions by session, week, and semester

- Duration of actions by type of user action

In this study, measures of duration were more commonly used than counts of actions or access. Previous studies have used a variety of measures to account for prevalence or frequency of use: the number of times user access or login to the system for a session (Wayman et al., 2009a, 2011), the duration of time spent accessing the system (Tyler 2013), or the count of actions taken while logged into the system. While action counts, session counts, and duration of use are all related measures of frequency, the use of action and session counts may be prone to discrepancies when comparing across systems and studies. Sessions, for example, may be counted differently in situations where a user opens multiple browser windows to run multiple instances of the data platform (Kaushik, 2010; Kovanovic et al., 2015). While in the current study, sessions were counted solely by unique session ID, it is difficult to know how previous studies have approached this issue. More explicit descriptions of the methods used to count unique sessions may benefit future studies, as well as investigation into user behaviors involving overlapping “unique” sessions across multiple tabs or browsers.

The use of action counts as measures of frequency also includes potential pitfalls given the complexity of logging user actions. As described more thoroughly in Dumais et al. (2014) and Maull (2013), the types and grain-level of user interactions recorded in server-side log files are subject to the recording infrastructure and log-recording specifications of each online application. It seems likely that different log file recording parameters for Tyler (2013), Wayman et al. (2017), Gold et al. (2012), and this study may make comparisons across action counts problematic. This is not to say that threats to internal validity exist within each study, but only that comparisons of action counts between studies should be taken with a grain of salt.

Measures of consistency. As a complement to prevalence or frequency of usage, measures of consistency provide a useful means of differentiating between users who may have used the system for several hours on one occasion and users who engaged with the system for shorter periods of time across multiple weeks of the school year. In these situations, depending on measures of frequency only would leave this critical kind of variation unexplored. The number of weeks a system is accessed (Wayman et al., 2009a), provides a useful measure of how consistently a user accessed the system over time. Consistency metrics calculated for this study include:

- Number of weeks accessed
- Maximum weekly duration/Total duration

Maximum Weekly Duration as a percentage of Total Duration describes the opposite of consistency, attempting to simply quantify the degree to which a users' behavior is limited in time. The higher the value of this outcome, the more concentrated a users' access was to a particular week of usage. Since usage timelines indicate that time in the system tended to concentrate in weeks with professional development or training sessions, a high value on Max. Weekly Usage/Total Usage may stand as a proxy for the degree to which a user focused their system access to scheduled, school-mandated sessions. A low value on this metric may indicate a more distributed and independent pattern of access.

Measures of variety. As conceptualized in Use Diffusion Theory (Shih and Venkatesh, 2004) and implemented in Maull (2013), usage variety describes the range of use across the available functions in the system. Of all the actions available in Benchmark Data, for example, how many and what kind of actions did the user take advantage of? Metrics generated include:

- Duration of each type of user action as a percentage of that users' total duration of use.

This metric is calculated for all types of user actions, representing the proportion of time spent performing each type of action, without regard to absolute time.

Creating categories of user action. Aside from methods for generating metrics and features, another area of preparation should be mentioned briefly: methods for grouping user actions for analysis. Since each row of a log file can specify one of a large variety of clicks on specific webpages, menus, or functions, studies of educators' online data use have generally grouped these specific clicks into more general categories, based on the content or function associated with each click. Wayman et al. (2017) for example classified teacher online actions into instructional and non-instructional functions and then further categorized instructional actions as (a) assigning content to students, (b) managing student information, (c) accessing reports, and (d) tracking students' completion of assignments. Tyler (2013) grouped webpages viewed into categories such as (a) class-level aggregate pages, (b) students-in-class pages, (c) individual student-level pages, (d) item pages, and (e) resources pages. While these functional categories are clearly useful for describing user behavior, other methods for grouping user actions by frequency of teacher usage (Maull, 2013) or clustering techniques may provide additional insights.

Benchmark Data, like other assessment and data use platforms, provided a wide range of pages, functions, and reports in order to help users prepare and administer tests, as well as generate, search, and filter test results. Table 12, below, outlines broad categories of user actions and the associated log file records for the current study. These categories of user action were developed through consultation with an experienced system administrator. Each recorded log file action was associated with a user function based on firsthand experience across all system functionality. In reviewing the categories presented further below, there are several important distinctions in how testing data is viewed in Benchmark Data, some of which are not immediately obvious based on the titles of functions alone:

- *Assessment Preparation vs. Assessment Administration vs. Assessment Navigation vs. Assessment Search.* Broadly speaking, *Assessment Preparation* activities are associated with the creation of an assessment, its answer key, and associated standards. *Administration* activities refer specifically to actions used to create, print, and scan answer documents. *Navigation* functions include clicks required to move between system pages or functions, while *Search* functions refer specifically to the filtering and search bar tools used to call up a particular assessment for reference.
- *Assessment Overview vs. Assessment Report*. The *Assessment Overview* page appears for users when they first access an assessment and provides summary charts of the average score, percentages of students meeting various pre-defined performance thresholds, and aggregate performance on standards and items. *Assessment Reports*, in contrast, offer more in depth and specific views of testing data, often related to item and standard analysis.
- *Reports vs. Assessment Reports.* General *Reports* combine data from multiple assessments or other sources of student data into a combined table view, while *Assessment reports*, as mentioned above, are related to in-depth viewing of the results for a particular test.

Table 12

Benchmark Data Functions and Associated Page/Click Types

General User Function	Examples of Associated Log File Pages and Actions	
Assessment Preparation	General Preparation Assessment_Field Standard Controller Assessment_Group Controller Assessment_Index Controller Assessment_ItembankFieldController Asssessment_Material Controller	Preparing Assessment Standards Assessment_StandardController Preparing Interpretation of Results Assessment_PerformanceBandSetController Using Itembank ItembankController
Assessment Administration	Preparing/Scanning Answer Sheets Assessment_SheetDesignerController Gradecam_SheetServiceController	Entering Student Responses Assessment_StudentResponseImportMapController Assessment_StudentResponseRestoreController
Assessment Report Generation	Specifying Assessment Report Fields Assessment_AggregatorController Assessment_ReportController	Assessment_ReportStatisticsController Assessment_ReportStudentCountsController Assessment_StandardProgressController
Assessment Navigation	Basic System Menus Assessment_FieldController	
Assessment Search	Search for Assessments AssessmentListController	
Assessment Staff Sharing	Sharing Assessment Reports with Staff Assessment_PermissionController	Assessment_ReportResponseDistributionController

Assessment Student Filtering	Filtering for Student Groups Assessment_StudentController	Assessment_StudentFilterController
Assessment Student Sharing	Printing Results for Students Assessment_ReprtStudentSlipController	
Assessment Overview	Overview of Assessment Results AssessmentDetailsController	Assessment_OverviewController
Assessment Report View	Specific Views of Assessment Results PrebuiltReport_CodeBasedController	AssessmentExportStudentResultsController
Report Search	Search Filters and Search Bar for Reports Reports_IndexController	
Report Creation	Creating Reports from Student Data ReportCreator_ChartController ReportCreator_ColumnAdd	ReportsCreatorManageEdit ReportsCreatorNlp ReportsCreatorSortEdit
Reports Data View	Viewing Report Data ReportsCreatorView	ReportsStudentProfile
Student Data View	Viewing Individual Student Data StudentDemographics	StudentProfile
Student Grouping	Creating Student Groupings Student_GroupStudentsAddController StudentGroupDetail	StudentGroupProgramsEdit StudentGroupReportsEdit

Student Search	Searching for Individual Students	StudentAdvancedSearch
	SisChooseStudent	
	SisQuickSearch	
System Function	Basic System Functions	SisError
	Auth_SessionTimeoutController	SisLogout
	DashboardPage	UserPassword
	MainWelcome	

Analyses. The following analyses were conducted in response to research question one: **(R1) To what extent and in what ways do teachers use online data and assessment tools?**

Unfortunately, the ratio of observations (39 users) to a much larger number of dimensions did not allow for the use of adequately powered inferential statistics. As a result, methods for analysis were descriptive and visual. Though descriptive methods pose a clear challenge to the generalizability of this study, they more closely align with the practical limitations of school-based analysis and can contribute to the framing of future, larger studies (Loeb et al., 2017).

Aggregate user metrics. Descriptive statistics were calculated for multiple metrics, aggregating the usage of all active core content teachers over the course of the entire semester. Additionally, these aggregate measures were broken down by school level (middle or high) and by core content area (ELA, Math, Science, and Social Studies). Where appropriate, user totals, averages, and standard deviations were calculated for session count, action count**,** duration of use, seconds/session, and seconds/action. Along with total usage metrics, usage by type of action was also calculated across school and content area in order to examine variation in how the online data system was used according to school roles. Histograms were generated for overall usage to explore the distribution of use across users.

Usage timelines. As a means of capturing and describing teacher usage over time, weekly and daily timelines of teacher usage were generated alongside relevant testing and teacher training dates from the 2015-16 school year. Line graphs and stacked bar charts were used to represent aggregate usage of the data system over time both in terms of total usage and broken out by specific types of user actions.

Individual usage pathways over time. To capture variation in usage across users, as well as over time, individual usage pathways were represented in the form of bubble charts which capture both the type of user action as well as its duration. Multiple forms of sorting and grouping within these charts allows for visual exploration of patterns by user, school, and content area.

Clustering users by online behaviors. Loeb et al. (2017) indicate that clustering techniques are a valuable tool for descriptive studies, offering a form of analysis for grouping units that have similar traits, in this case teachers' online use. One such method, based on methods developed in bioinformatics and applied to education by Bowers (2007; 2010), combines the clustering and dendrogram produced by agglomerative hierarchical clustering analysis (HCA) (Hastie, Tibshirani and Friedman, 2017) with a heatmap of the clustering factors for all individual observations. In this study, such HCA heatmaps, clustering users in rows and usage factors in columns, allow for visual inspection of the HCA clusters within the full context of individual variation in usage, represented by the colored cells of the heatmap. Annotations displayed to the right of the heatmaps allow for additional inspection of key factors and outcomes, such as school level, content area, and overall usage, in relation to both the dendrogram hierarchy and heatmap representation.

HCA employs both a dissimilarity measure to establish the distance between each single pair of observations, as well as a method of linkage used to define the dissimilarity between groups of observations (James, Witten, Hastie and Tibshirani, 2013). In agglomerative clustering, the method of linkage is employed iteratively to group the most similar observations and groups of observations. HCA and dendrograms in this study were generated using R version 3.5.1 in RStudio 1.1.43 with code adapted from Bowers and Zhao (2018) and based upon work supported by the National Science Foundation under grant no. 1546653.

HCA heatmaps and dendrograms in Study 1 employ the common method of centering all clustering variables at a mean of zero, with their standard deviations scaled to one. Because of large differences in overall usage between middle and high school users, all usage factors in Study 1 are group-mean centered by each user's school, either middle or high. Euclidean distance is used as a measure of dissimilarity (James et al., 2013) and the linkage between groups of observations is established with Ward's Method (Kovanovic, Gasevic, Joksimovic, Hatala, and Adesope, 2015), which minimizes the within-group sum of squares at each stage of linkage in the hierarchical clustering (Murtagh and Legendre, 2014). Ward's Method was implemented in R with Ward.D2 of the hclust function (Murtagh and Legendre, 2014). Ward's method has been found to perform better than other clustering techniques in reproducing an original structure of clusters (Blashfield, 1976; Hands and Everitt, 1987).

A separate method, average linkage (Bowers, 2010; Lee et al., 2016), which computes the average of all pairwise dissimilarities between two clusters, was also attempted, as its robustness to missing data makes it an extremely practical option. In this case, since rates for missing data were low, Ward's was implemented for its tendency to produce well-defined clusters.

While one of the strengths of the HCA heatmap methodology is to make all clusters transparent within the dendrogram structure, some of these clusters will be more distinct and others more closely related as measured by the vertical distance of their union on the dendrogram. The interpretation of clusters or even the choice of which clusters to interpret is often highly dependent on manual analysis of the dendrogram, as well as on the research context and domain (James et al., 2013). Throughout these analyses, the interpretation of HCA heatmap clusters will be determined by visual inspection of the dendrogram for a sensible number of clusters in relation to factors and by a voting recommendation for the best numbers of clusters, as implemented in R by the NbClust package (Charrad, Ghazzali, Boiteau, and Niknafs, 2014). NbClust applies 30 indices for determining the number of clusters and returns the recommendations of these methods. Because of the small sample size, the minimum number of clusters was set to two in NbClust and the maximum to seven. For each HCA heatmap, results of the NbClust analysis are provided along with a dendrogram of the HCA results indicating the height at which the dendrogram was cut in order to identify a number of clusters for interpretation (see Appendix C).

Results

Aggregate user metrics.

Distribution of usage and descriptive statistics. Overall usage of the system by all core content teachers was 154 hours over the course of one school semester, which consisted of 112 instructional days. This overall usage represents 3.85 hours of use per core content teacher, or 1.38 hours per instructional day, or about ten minutes of access a week for every core content teacher.

As in previous studies of teacher data systems, the distribution of this access was highly positively skewed (Wayman et al., 2009, Tyler, 2013). The largest numbers of teachers tended to access data systems either not at all or for only short periods, with a minority of teachers accessing the system for much longer amounts of time and producing a characteristic long tail of users to the right of the distribution. This distribution is easily seen in the current study (

Figure *12*), with almost half of teachers accessing the Benchmark Data system for less than an hour over the entire period of the second semester, rather than the roughly four hours of average use suggested by total system usage.

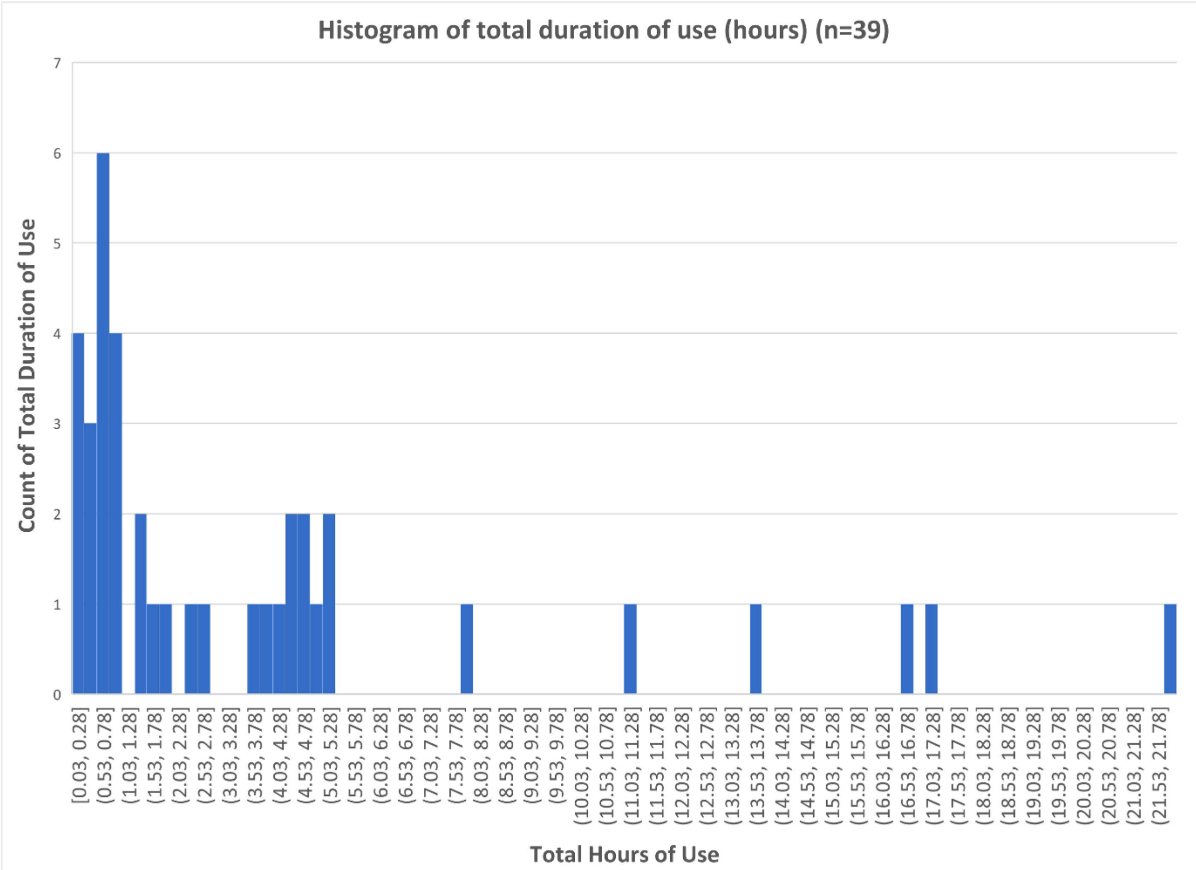


Figure 12. Histogram of total duration of use (hours) (n=39)

A breakdown of this skewed distribution begins with an overview of teachers' total and average online usage across school levels (middle and high school) in Table 13, and across core content areas (English Language Arts, Math, Science, and Social Studies) in Table 14 and Table 15. In comparing usage between middle school teachers (n = 20) and high school teachers (n = 19), high school teachers accounted for 69% of time spent and of actions completed in the system. Proportion of sessions followed a somewhat different pattern, with high school teachers accounting for 59% of total sessions.

Table 13

Total Online Actions, Sessions, and Use: Middle and High School

	School Level				
	Middle School n=20		High School n=19		All Active Users n=39
Usage Metrics	Total	% of Total	Total	% of Total	Total
Count of Online Actions	2,442	31.0	5,447	69.1	7,889
Count of Sessions	155	41.1	222	58.9	377
Duration of Use (Hours)	47.1	30.6	106.7	69.4	153.9

Within the middle and high school, core content areas followed distinct patterns of use (see Table 14 and Table 15). For example, Math teachers dominated system usage in the middle school, contributing 79% of the total duration of middle school use, while in the high school, Math teachers accounted for only 43% of total high school duration of use. English Language Arts (ELA) teachers used the system with the second highest frequency in middle school, for 15% of total usage, while Science and Social Studies each used the system for less than 5% of the total middle school access to the system.

Given that Benchmark Data, as implemented for middle school teachers, presented mostly interim testing results in preparation for ELA and Math state testing, it is completely understandable that middle school Science and Social Studies teachers spent less time accessing the system. On the other hand, the curricular standards addressed in ELA interim testing, particularly those standards addressing reading comprehension, have possible affordances for instruction in Science and Social Studies classrooms, especially regarding the appropriate level of texts and the mastery of comprehension strategies.

Overall, the relationships between three frequency metrics (action counts, session counts, and duration) appear to vary widely by school and content area. At the school level, the proportion of middle school actions to high school actions closely parallels their proportion of time spent in the system. The percentages of sessions, however, accounted for by each school, move in opposite directions, with middle school accounting for a higher than anticipated number of sessions (Table 13).

Within content areas, other differences between frequency metrics apply. The percentage of actions, sessions, and time accounted for by high school Math teachers are all roughly the same, while ELA teachers in Table 15 account for a lower percentage of overall actions (23%) than they do for overall duration of use (31%), suggesting possible differences in use between content areas.

Table 14

Total Online Actions, Sessions, and Duration of Use: Middle School Content Areas

	Middle School Content Areas								
	ELA n=4		Math n=10		Science n=3		Soc. Stud. n=3		Middle School n=20
Usage Metrics	Total	% of Total	Total	% of Total	Total	% of Total	Total	% of Total	Total
Online Actions	217	8.9	2,023	82.8	79	3.2	123	5.0	2,442
Sessions	21	13.6	121	78.1	5	3.2	8	5.0	155
Hours Used	6.9	14.6	37.0	78.5	1.2	2.5	2.1	4.5	47.1

Table 15

Total Online Actions, Sessions, and Duration of Use: High School Content Areas

	High School Content Areas								
	ELA n=8		Math n=4		Science n=2		Soc. Stud. n=5		High School n=19
Usage Metrics	Total	% of Total	Total	% of Total	Total	% of Total	Total	% of Total	Total
Online Actions	1,272	23.4	2,357	43.3	1,602	29.4	216	4.0	5,447
Sessions	56	25.2	96	43.2	46	20.7	24	10.8	222
Hours Used	33.4	31.3	45.6	42.7	20.4	19.1	7.4	6.9	106.7

In the high school, content areas tended toward a more equal distribution of use than in middle school. Math teachers still used the system for the largest percentage of time (43%), but with ELA (31%), and Science (19%) following more closely behind. Social Studies, with only 7% of overall use in the high school, demonstrated the lowest use across both schools. This more equal distribution of access by content areas in the high school matches the accountability structure of high school state testing, where Regents tests (and therefore interim testing) were required in all four content areas, as opposed to middle school state testing, which primarily tested students in ELA and Math only.

In terms of overall mean usage, the average user across schools logged in for ten sessions, lasting for about 25 minutes each. The average user completed about 16 actions per session, with each action lasting a little less than two minutes. However, the differences in overall usage between the middle and high school hold when considering middle and high school average usage (Table 16). High school teachers averaged 70% longer actions, 35% longer sessions, and four more total sessions than middle school teachers, though the average number of actions per session across schools was roughly the same. The standard deviation for high school users was also larger across all metrics, reflecting a wider range of teacher behavior.

Table 16

Average Online Actions, Sessions, and Duration of Use: Middle and High School

	School Level					
	Middle School n=20		High School n=19		All Active Users n=39	
User Metrics	*M*	*SD*	*M*	*SD*	*M*	*SD*
Action Count	122.1	173.6	286.7	472.9	202.3	357.7
Session Count	7.8	10.8	11.7	12.5	9.7	11.7
Hours Used	2.36	3.19	5.62	6.39	3.95	5.21
Avg. Sec./Actions	84.6	46.0	141.9	90.7	112.5	76.1
Avg. Act./Session	15.7	6.2	16.0	12.8	15.9	9.8
Avg. Sec./Session	1287.6	772.2	1737.2	1014.8	1506.6	915.5

Average usage in high school Math and ELA was much higher than in the middle school. In both schools, Math teachers averaged about three times as many actions as ELA teachers and two to three times the overall duration of use (Table 17 and Table 18). At 11.39 and 10.21 hours, respectively, high school Math and Science teachers had the highest average use over the course of the semester, followed by high school ELA teachers (4.17) and middle school Math teachers (3.70). Average use by Social Studies teachers was the lowest in both middle (0.71) and high school (1.47). Such low access is easily explainable in middle school, where Social Studies teachers did not have access to interim testing directly related to Social Studies content (as opposed to general reading comprehension). Low access in high school, however, is less easily explained, since most high school Social Studies classes received interim testing results through Benchmark Data in preparation for one of two end-of-course Regents exams.

Table 17

Average Online Actions, Sessions, and Duration of Use: Middle School Content Areas

	Middle School Content Area									
	ELA n=4		Math n=10		Science n=3		Soc. Stud. n=3		Middle School n=20	
User Metrics	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Action Count	54.3	62.4	202.3	218.2	26.3	25.1	41.0	23.4	122.1	173.6
Session Count	5.3	6.5	12.1	13.7	1.7	1.2	2.7	2.1	7.8	10.8
Hours Used	1.71	2.36	3.70	3.90	0.39	0.45	0.71	0.16	2.36	3.19
Seconds/Actions	103.4	29.3	94.3	55.4	41.7	17.6	70.1	21.4	84.6	46.0
Actions/Session	11.4	7.7	17.1	4.2	14.3	8.1	18.4	8.3	15.7	6.2
Seconds/Session	1039.3	398.5	1565.4	892.6	673.0	492.9	1307.2	727.7	1287.6	772.7

Table 18

Average Online Actions, Sessions, and Duration of Use: High School Content Areas

	High School Content Area									
	ELA n=8		Math n=4		Science n=2		Soc. Stud. (n=5)		High School n=19	
User Metrics	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Action Count	159.0	195.3	589.3	679.0	801.0	1021.1	43.2	50.3	286.7	472.9
Session Count	7.0	4.5	24.0	17.9	23.0	18.4	4.8	4.7	11.7	12.5
Hours Used	4.17	3.75	11.39	9.31	10.21	9.25	1.47	1.68	5.62	6.39
Seconds/Actions	159.0	118.3	105.9	44.3	103.5	90.4	158.5	77.8	141.9	90.7
Actions/Session	17.4	12.9	19.3	13.7	25.1	24.3	7.5	2.1	16.0	12.8
Seconds/Session	2212.7	1341.8	1625.8	288.7	1498.7	249.6	1160.9	704.8	1737.2	1014.8

Summarizing metrics from Table 17, Table 18, and Table 19 (below) highlights the main usage behaviors resulting in longer durations of access in high school over middle school. While access for all content areas was longer in the high school, the table describes how access increased due to different user behaviors in different content areas. Math teachers in the high school, for example, experienced sessions with similar lengths and number of actions to middle school Math teachers, but on average they completed twice as many of those sessions. ELA teachers in high school, on the other hand completed only slightly more sessions than middle school teachers, but they spent more time in those sessions, taking a greater number of longer actions.

Table 19

Usage Behaviors Accounting for Average High School Duration of Use

Content Area	Difference in Average Duration of Use from Middle School (Hours)	Reasons for difference in Average Duration of Use
ELA	+ 2.46 in High School	More seconds per action and more actions per session lead to longer sessions
Math	+ 7.69 in High School	Twice as many sessions of about the same length
Science	+ 9.82 in High School	12 times as many sessions, where the sessions are twice as long
Social Studies	+ 0.76 in High School	Twice as many sessions, where the sessions are slightly shorter, with fewer actions

Correlation of usage outcomes. In addition to these aggregate statistics, a correlation matrix was generated to compare relationships between the usage outcomes (Table 20). These various outcomes had pros and cons as measurements of usage. Total Action Count, for example, was sensitive to inflation by rapidly completed user actions, such as the scanning of student answer documents. While Duration of Use was not susceptible to the same inconsistencies as

Total Actions, it had its own shortcomings in presenting as equal two types of users: those with fewer longer sessions and those with more frequent shorter sessions. Still, total actions and total sessions were highly correlated with total duration, both at $r = 0.92$. Total sessions and total actions were only slightly less-highly correlated at $r = 0.86$.

Average session length, in contrast, was found to be unrelated to other usage outcomes, while teachers' maximum weekly usage as a percentage of their total weekly usage had a strong negative relationship with all other usage outcomes, except for average session length. These low and negatively correlated outcomes may be explained by a subset of teacher behaviors related to required school events, such as training sessions or team meetings devoted to viewing interim testing results. For example, teachers who only accessed the system during such required sessions would produce high values for maximum weekly usage/total usage, as well as a higher average duration of sessions (as they worked though the forty to fifty minutes of a training session) but low overall duration of use due to their lack of independent access at other times in the school year.

Table 20

Intercorrelations for Usage Metrics

Measure	1	2	3	4	5	6
1. Total Actions	—					
2. Avg. Session Duration	.05	—				
3. Weeks Used	.76	-.05	—			
4. Max Weekly/Total Duration	-.62	-.01	-.68	—		
5. Total Session Count	.86	-.07	.91	-.76	—	
6. Total Duration of Use	.92	.18	.78	-.60	.92	—

Timelines of teacher usage. Along with descriptive statistics of teacher usage, several types of timelines were generated to capture the variation in teacher usage over time. Values graphed over weekly timelines include total weekly count and duration of online actions (Figure 13), weekly unique users (Figure 14), duration of use by type of user action for all teachers combined (Figure 15, Figure 16, Figure 17, and Figure 18), and by individual teacher (Figure 19) Multiple graphs and timelines are presented in relation to key testing and professional development dates of the 2015-16 school year.

In Figure 13 (following) teachers' weekly count of online actions is contrasted with their weekly duration of actions. Given the common pattern of nine-week interim tests culminating in end-of-year state testing, this decreasing pattern of usage over the course of the year is somewhat surprising. A reasonable alternative might be to expect an increase in system access closer to the state test, with the largest usage of interim testing results during the session immediately prior to the state test. Instead usage declines over the course of the year, with successive rounds of interim testing associated with less and less access.

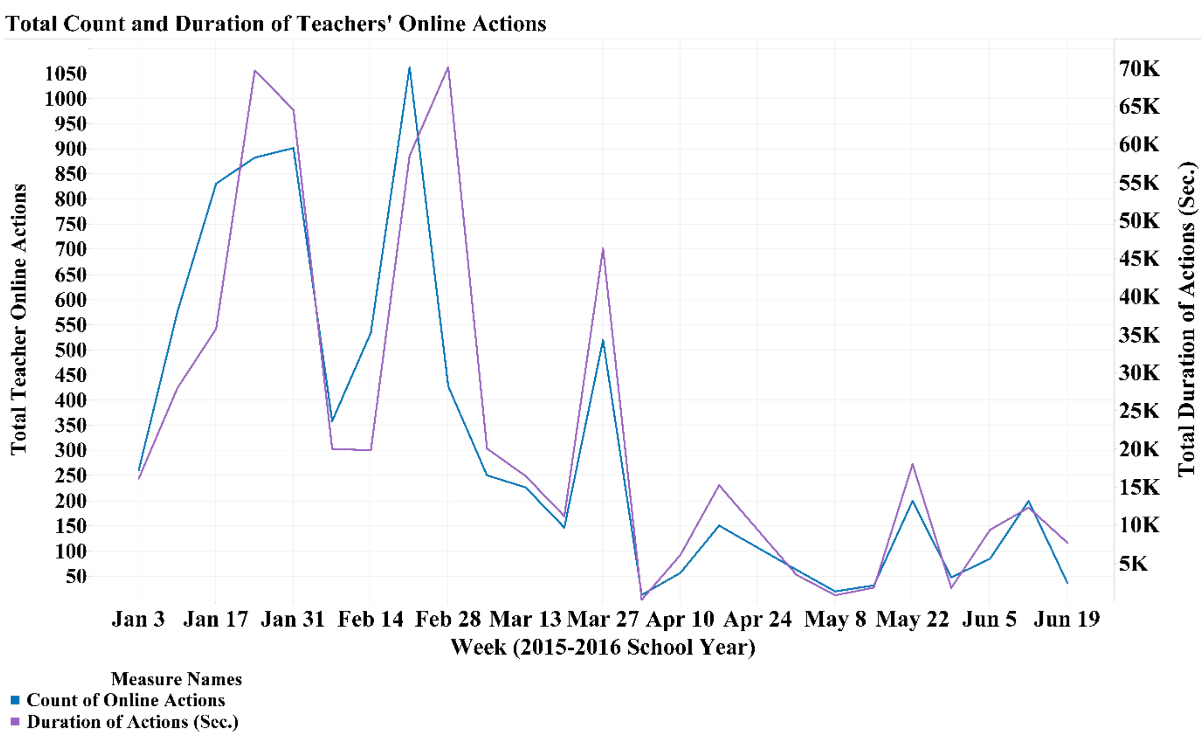


Figure 13. Total action count and duration of teachers' online usage

While for the most part the overall count of actions follows the same trends as the overall duration of use, there are some exceptions. During a few weeks in January and February, for example, a peak and subsequent dip in the number of actions arrives a week out of phase with a subsequent peak and decline in the overall duration of use. The largest number of actions (1,063), for example, takes place during the week of February 21st, while the greatest duration of use (70,174 seconds, 19.5 hours) takes place a week later.

In order to better present user access in relation to individual teachers, Figure 14 presents duration of use in relation to the weekly count of unique users. The pattern of unique users across weeks is much more stable over time, with an average of only 8 different teachers accessing the system each week. The same difference in timing between actions and duration of use noted in Figure 13 can also be seen in Figure 14, as the number of unique users during weeks 1/24 and 2/21 remains low while the overall duration of use rises dramatically. Prior to spikes in unique teachers using the system, the number of unique teachers remains close to average, while the duration of their use reaches some of the largest amounts of the semester. This pattern of use highlights the important role that a small number of teachers can take in the preparation of interim assessments and their out-sized impact on overall usage.

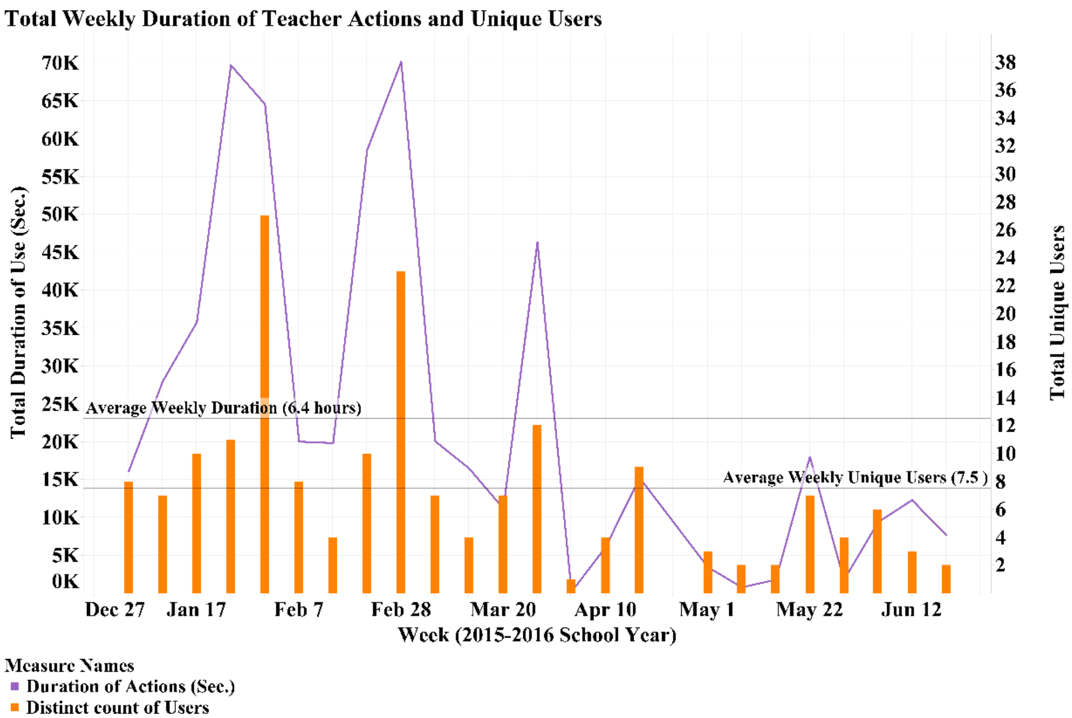


Figure 14. Total weekly duration of teacher actions and unique users

Building on the broad picture of usage presented in Figure 13 and Figure 14, the bar charts presented below add key elements, such as assessment-related dates from the school calendar and multiple views of relevant categories of online action. With these additions, Figure 16 and Figure 17 allow for easier analysis of school events driving usage across the semester. Figure 15, immediately following, displays all types of teacher actions categorized from the web page classification system employed within the Benchmark Data system. Figure 16 limits the actions displayed to the top ten most frequent categories of use, and Figure 17 displays only those actions directly related to the viewing of test results, eliminating any usage related to the preparation and administration or tests, or the search for assessment results. This highly filtered view presents a pattern of peaks and troughs, distinct from more inclusive charts of teachers' overall system interactions.

While somewhat dense, the stacked bar charts make visible some patterns of overall usage, such as the large investment of teacher time in test preparation activities in the weeks preceding interim assessment administrations, as well as what appears to be shorter periods of test preparation activities happening on a regular basis.

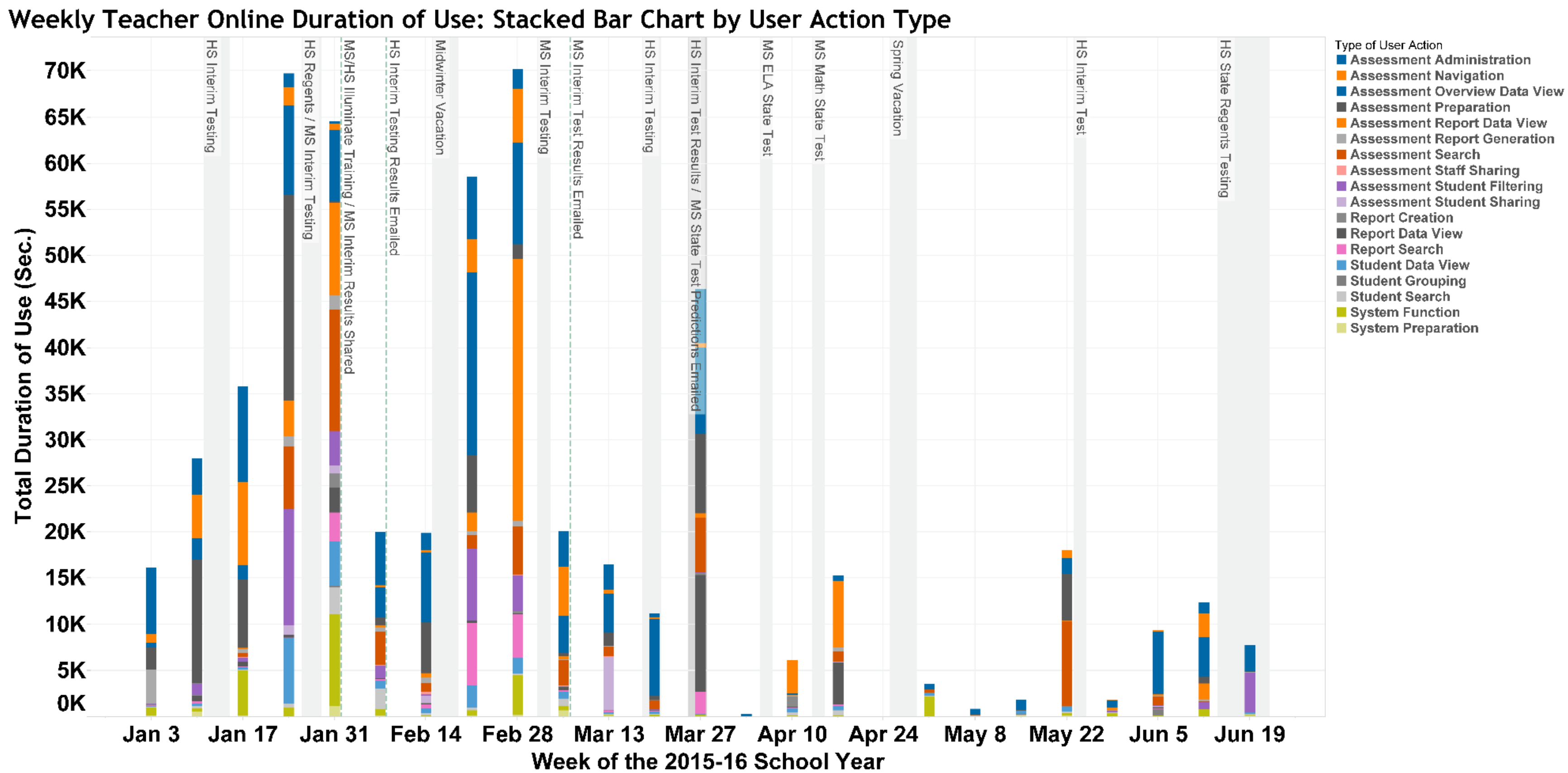


Figure 15. Weekly teacher online duration of use: stacked bar chart by user action type

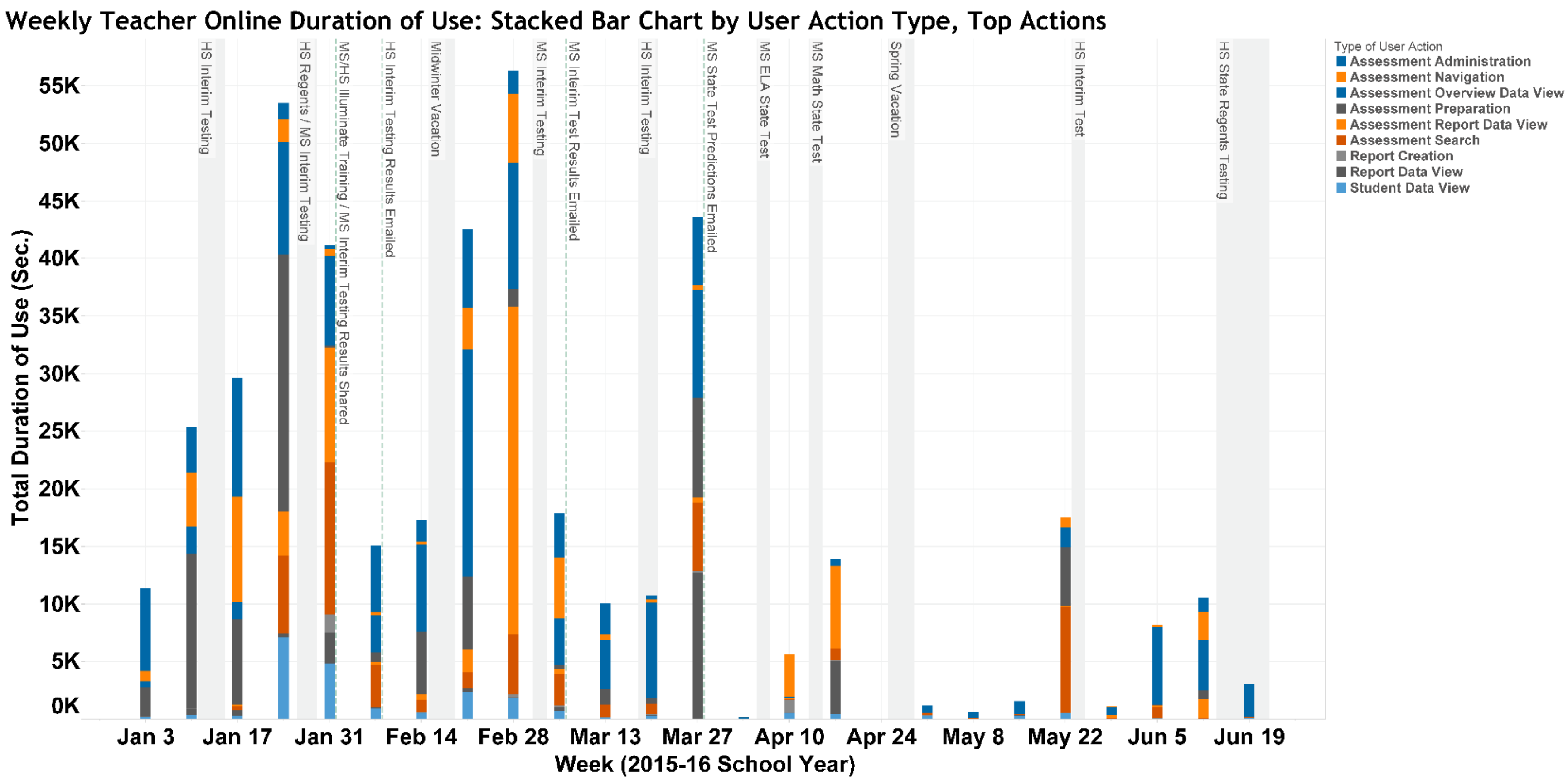


Figure 16. Weekly teacher online duration of use: stacked bar chart by user action type (top actions)

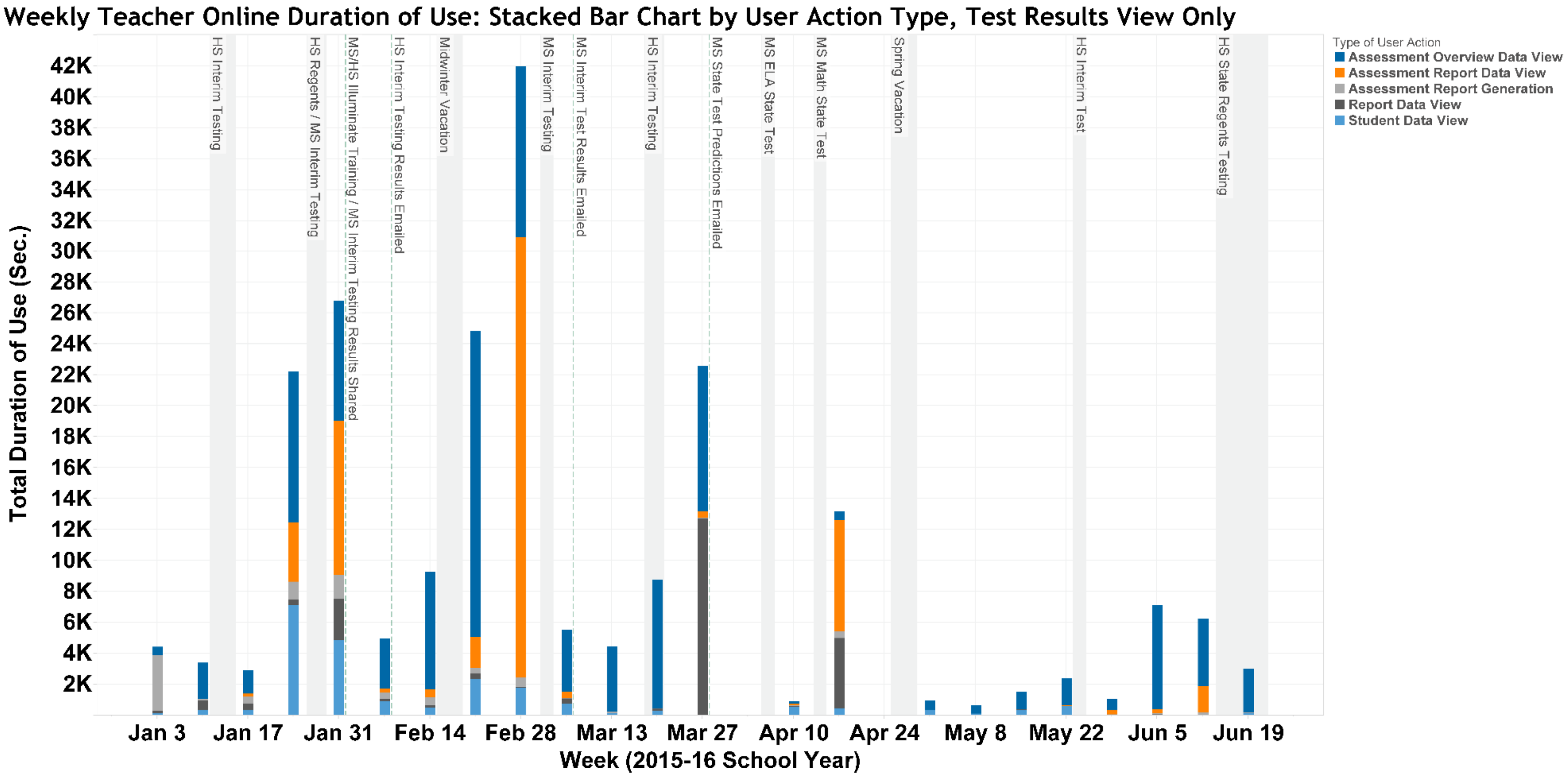


Figure 17. Weekly teacher online duration of use: stacked bar chart by user action type, test results view only

Including relevant dates in the three figures above allows for contextualization of spikes and dips in usage over the semester. The largest spike in usage on February 1st, for example, aligns with a combined middle and high school training and viewing of interim testing results. High school interim testing and Regents testing administration align with multiple spikes in usage in late January, March, and May. Middle school interim testing aligns with a spike in usage on March 3rd. Other moments of high usage are less easily explained by identified events, such as the spike in usage on February 18-24th. Also identifiable in these figures is how large spikes in aggregate usage often consist of different combinations of online actions. Surges in usage can occur for different reasons and consist of multiple types of teacher actions, which can vary widely depending on the day or week in question.

Additionally, several types of teacher actions seem to follow regular, but much lower levels of activity, such as viewing individual-level student data or accessing basic system functions. The relationship over time between three different ways of viewing test results is also visible: (a) viewing the initial Assessment Overview Data View (b) viewing more in-depth and specific Assessment Report Data Views, and (c) viewing Report Data Views which combine results from multiple assessments and/or contextualize results data by school performance bands. While the most common way of viewing test results appears to be through the overall test view page, substantial time is also spent in more specific assessment reports. Report views which combine multiple types of student data or contextualize test results are the least viewed form of student data, though there is a spike in usage in the week of March 27th, coinciding with the release of a report view which included predicted values of Middle School state test scores.

Finally, Figure 18, in a similar vein, displays the top ten categories of use on a daily, instead of weekly, basis. While cluttered, this view has some possible use in describing patterns of use that may occur at a more fine-grained level. For example, what appears as a spike in usage during the week of February 28th in Figure 17, is shown in Figure 18 to consist of more intense usage over multiple days, when viewed as daily usage.

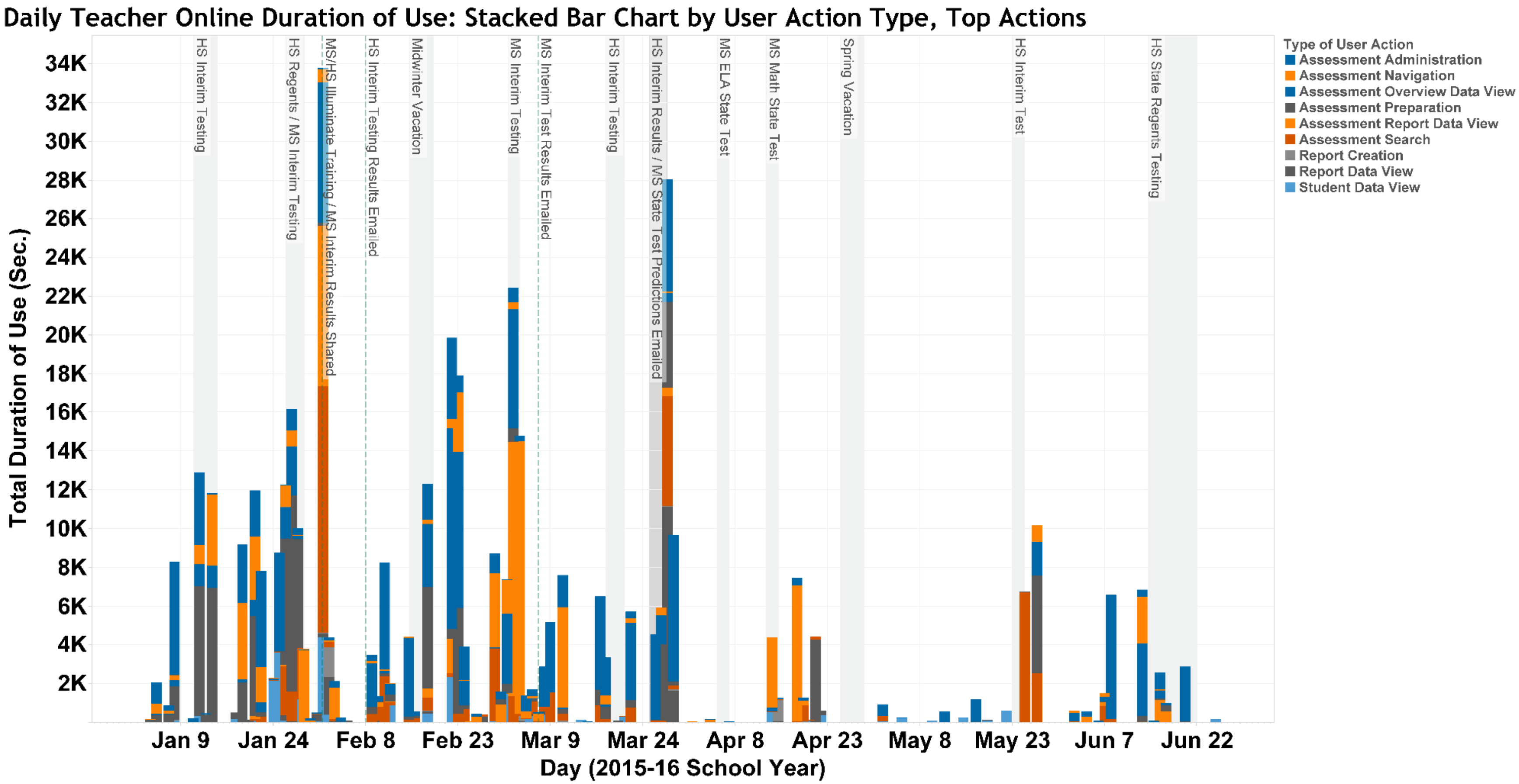


Figure 18. Daily teacher online duration of Use: stacked bar chart by user action (top actions)

Individual User Pathways. While the figures above allow for inspection of overall trends in access, Figure 19 (see below) provides a means of exploring patterns of individual use over time, while still in the context of school and content area. Each user's interaction with the data system is presented by colored squares representing moments of interaction, arranged on the *x*-axis in rows over the days of the second semester. The size of each square represents the duration of the user action, while the color indicates the type of action. Though the aggregate charts above allow for some similar observations, Figure 19 brings added specificity to visualizing individual use along with group trends. With these individual pathways as a guide, several interesting patterns emerge:

- Figure 19 highlights the impact of training and professional development on usage, with many teachers focusing their use of the system almost entirely within scheduled school training and planning sessions.
- Email appears to be an ineffective means for administrators to share testing results with teachers. Teacher usage of Benchmark Data after instances where email was used to notify staff of the availability of testing results demonstrates only minimal increases, especially when compared to sessions of in-person training.
- Several users in the high school show patterns of heavy usage related to preparation and management of assessment.
- Neither the middle nor high school appears to make extensive use of the more intensive assessment reports available in Benchmark Data, choosing instead to use the assessment overview as their major point of access for test results.
- While most users move directly to viewing the test results themselves, some users appear to spend substantially more time than others in searching and navigating to assessments.

- While most usage is at the class and assessment level, a subset of users appear to focus on accessing individual-level student data.

Sorting by overall usage makes visible the extent to which a minority of users account for the bulk of extended use. In middle school, for example, only 37% of teachers demonstrate a pattern of more consistent and intensive use over the course of the year. High school users appear to follow a similar overall pattern, with 37% of users demonstrating more consistent semester-long use. In the high school, however, consistent users are drawn from a wider variety of content areas: ELA, Math, and Science. In both middle and high school, Social Studies teachers demonstrate minimal usage.

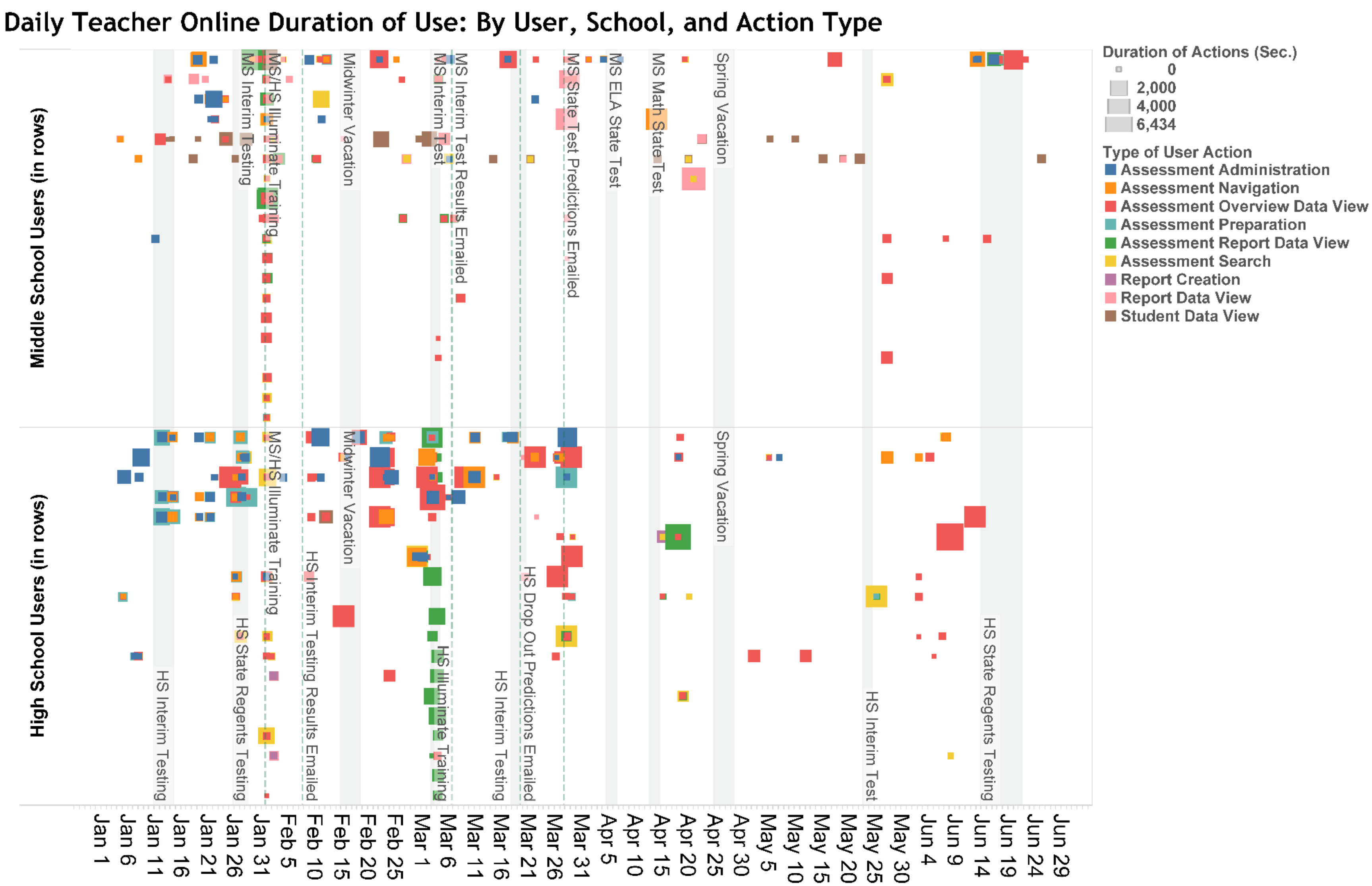


Figure 19. Daily teacher online duration of use: by user, school, and action type

HCA heatmaps of usage factors. As a means of further exploring subcategories of teacher usage, I applied hierarchical cluster analysis and heatmapping to multiple factors of teacher usage. Three HCA heatmaps are included in this analysis. The first, Figure 21, clusters users according to factors of frequency and consistency, following Wayman et al.'s (2009a; 2017) analyses. The following heatmap, Figure 22, clusters users according to their variety of usage, as described in use diffusion theory (Shih and Venkatesh, 2004) and represented in this case by the proportion of time that teachers access the available online functions. The third heatmap, Figure 23, follows Maull (2013) in applying clustering to factors defined in terms of frequency and variety, the two dimensions of use diffusion theory. In all three heatmaps (Figure 21, Figure 22, and Figure 23) usage features related to dimensions of frequency, consistency, and variety are combined in order to explore possible subgroups of user behaviors.

Interpreting HCA heatmaps. Each HCA heatmap presented in Study 1 and Study 2 has common characteristics. Figure 20, following, acts as a brief guide to interpreting these common elements across studies, identifying and describing the essential parts of an HCA heatmap, such as the factors used for clustering, the colored heatmap cells, the dendrograms clustering rows and columns, and the annotations that help identify relationships between HCA clusters and other factors.

A brief note is also included with each heatmap, describing the decision to identify a cut off for the number of clusters to present and discuss. In each case the decision of how many clusters to consider is guided by both visual inspection and the voting recommendations from the NbClust package in R (Charrad et al., 2014). For reference, a version of all dendrograms that indicates the height at which clusters were separated is included in Appendix C

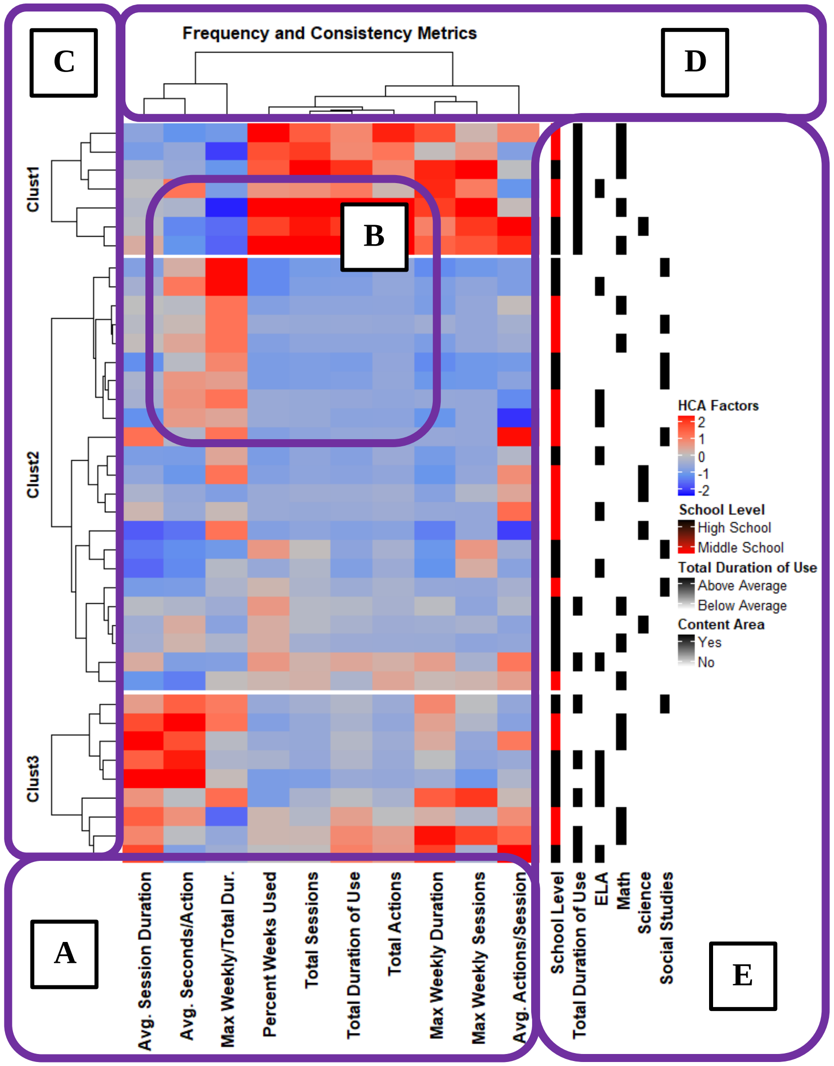


Figure 20. Guide to reading hierarchical cluster analysis (HCA) heatmaps

A These variables are used by the Hierarchical Clustering Analysis (HCA) algorithm to create clusters of users (*rows*) and factors (*columns*).

B Colored rectangles (*cells*) indicate users' standardized scores for the metric indicated by the column. Standardized scores are presented along a spectrum of blue to red values, with the darkest blue rectangles representing values furthest below the average and the darkest red rectangles indicating values furthest above the average.

C This branching dendrogram indicates how each *row* was clustered by the HCA algorithm. The height of lines in the dendrogram measures the differences between rows and clusters. Height is measured by moving up the dendrogram hierarchy from the origin of the line to the level at which two rows or clusters are connected.

D This dendrogram at the top of the figure indicates how *columns*, or factors, were clustered by the HCA algorithm. One larger cluster made up of seven factors and one smaller cluster of three factors are visible.

E These black and red rectangles (*annotations*) help the viewer relate the clusters identified by the dendrogram to important factors or outcomes. All the heatmaps in this study include the same annotations: (1) whether the user taught middle or high school, (2) whether the user demonstrated above average duration of use, and (3) what content area the user taught. Here, for example, annotations allow the viewer to recognize that users in Clust1 tend to have higher online use and teach Math.

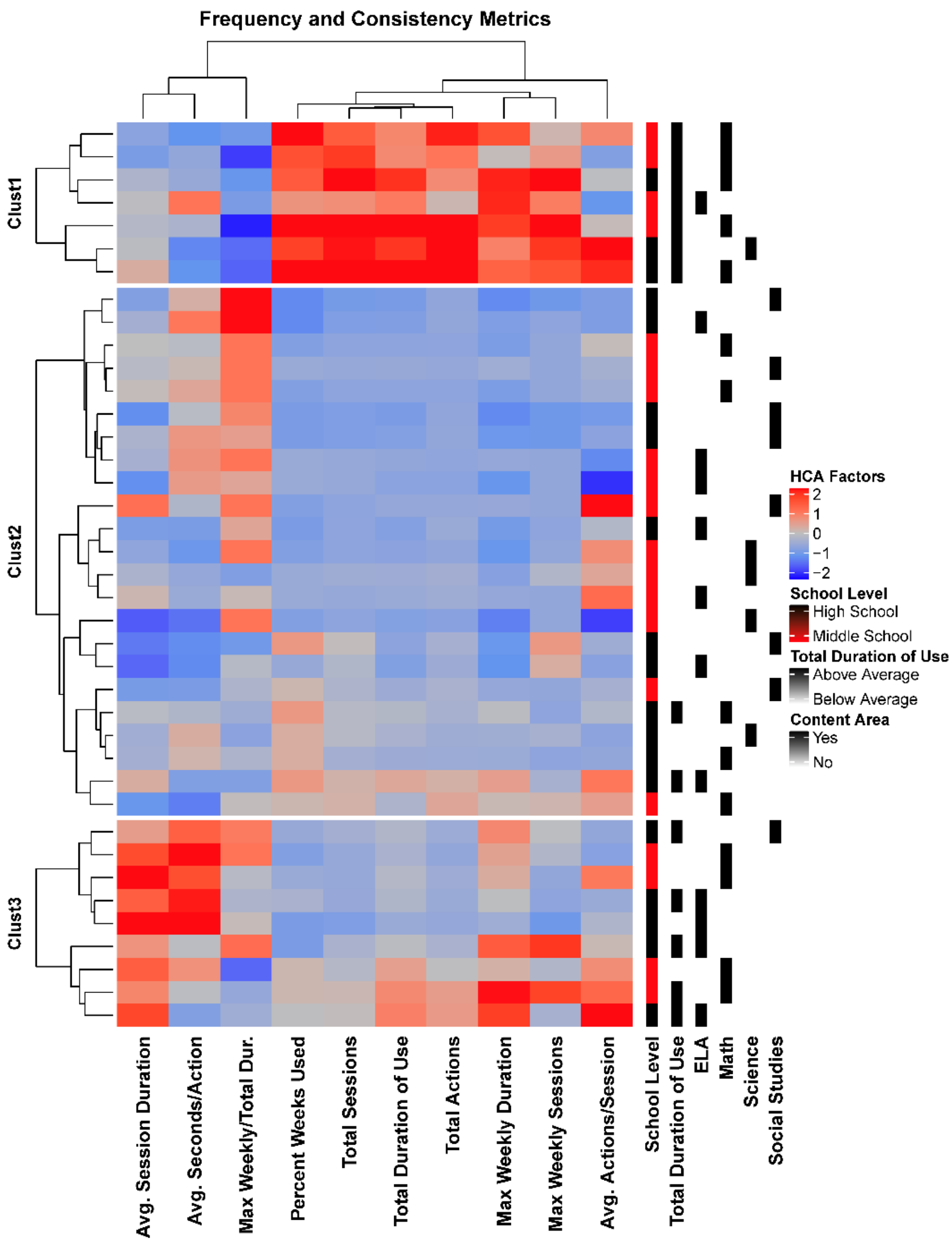


Figure 21. HCA heatmap: frequency and consistency of online use. NbClust found that ten methods proposed two clusters, while eight methods proposed three clusters. Visual inspection found three clearly-defined groups.

Clustering by online frequency and consistency. The HCA Heatmap in Figure 21 clusters users by a variety of features related to prevalence and consistency of use (Wayman et al., 2009). Where past studies have tended to use one major indicator of prevalence or frequency—total actions (Wayman et al., 2009) or total duration (Tyler, 2013; Gold et al., 2012)—Figure 21 clusters several indicators of prevalence and frequency at the level of the individual action to session-level behavior to total usage. Clustering across these multiple features indicates how multiple usage outcomes may contribute to a fuller picture of teachers' online use of student data.

Clustering along features of frequency and consistency, Figure 21 divides users into three major groups. Two of these groups, Clust1 and Clust3, contain more users with above average use duration, while Clust2 consists almost entirely of users with below average duration of use. Interestingly, Clust1 and Clust3 users appear to reach their higher levels of usage through two different paths. Clust1 users increase their prevalence of use through more frequent and consistent sessions of average length, while Clust3 users achieve above average use through fewer, but longer sessions. In contrast to the other two clusters, most Clust2 users are characterized by high values of Max. Weekly Duration/Total Duration, indicating that their interaction with Benchmark Data was highly focused during one week of the year, most likely during a training session, a usage pattern also seen in Figure 19. However, a subcluster of Clust2 does appear to use the system with slightly higher consistency over time, with two of these users demonstrating above average usage.

Group-mean centering for middle and high school users appears to leave them fairly evenly-distributed across the clusters, as opposed to what Table 16 indicates about greater usage in the high school in absolute terms. Content area, on the other hand, appears to be related to the identified clusters of usage, with higher usage Clust1 and Clust2 consisting almost entirely of Math and ELA teachers and lower-usage Clust2 containing almost all of the identified Science and Social Studies teachers.

The usage factors themselves are also clustered in interesting ways. In the middle of the columns, factors for Percentage of Weeks Used, Total Session, Total Duration of Use, and Total Actions cluster tightly together with only a slight separation for Percentage of Weeks Used from the other three factors. In contrast, usage factors related to session-level behaviors, such as Average Actions/Session, Average Seconds/Action, and Average Session Duration, appear critical in describing differences between the three identified subgroups.

Clustering by online variety of usage. Having explored possible subgroups based on features related to frequency of use, the next heatmap (Figure 22) clusters features related to users' variety of online use, a dimension of usage drawn from the use diffusion framework (Shih and Venkatesh; Maull, 2013).

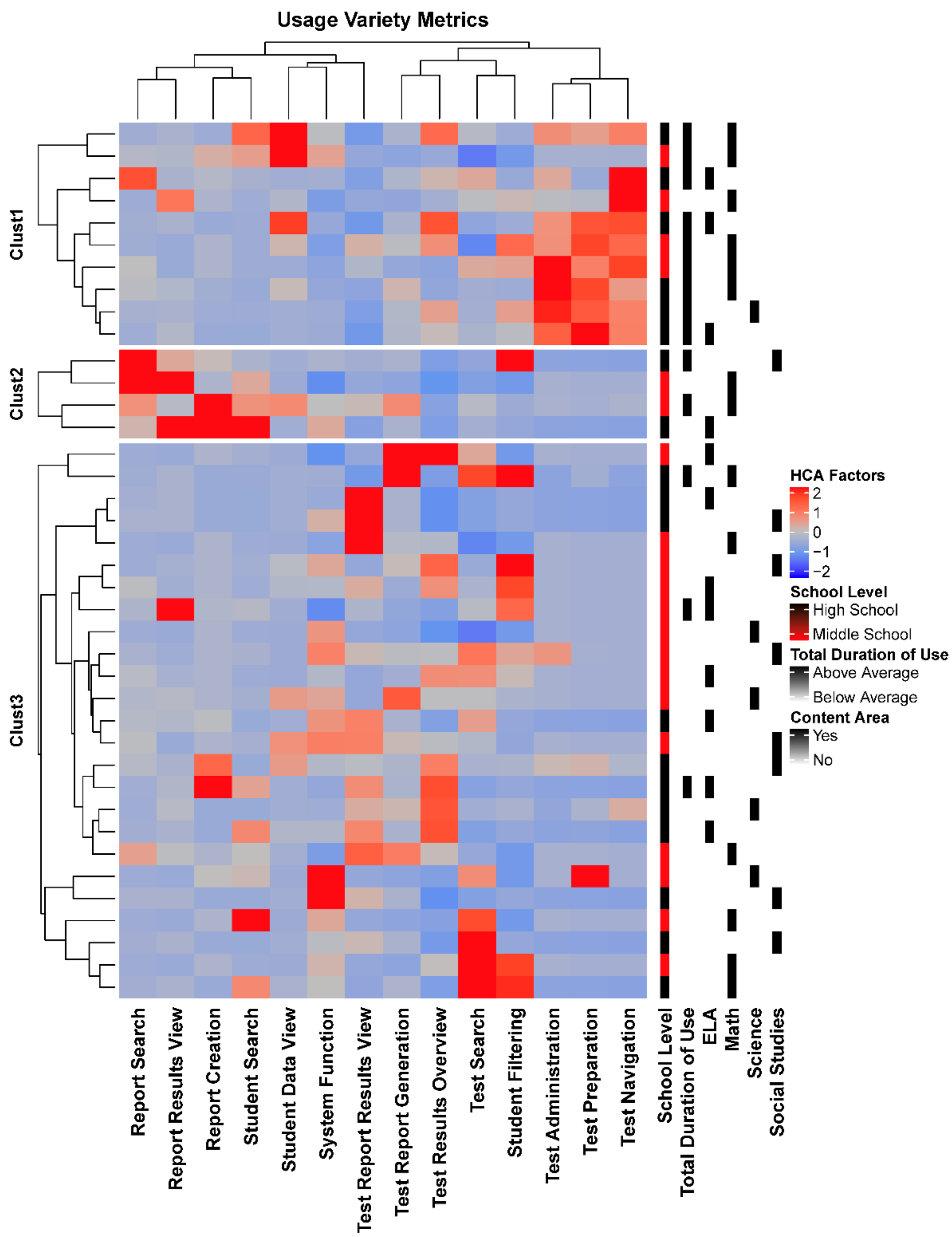


Figure 22. HCA heatmap: variety of online use (duration of page type use as a percentage of total use). NbClust found that eight methods proposed three clusters as optimal, while ten methods proposed seven clusters as the best option. Visual inspection suggests three main clusters.

Figure 22 identifies three major clusters of usage behavior based on the proportion of total use that users allocated to various system functions. Users in Clust1 spent a much higher proportion of their usage time on test administration, preparation, and navigation activities, with a small subgroup of Clust1 focusing on searching and viewing individual level student data. Users in the smaller Clust2 focused their usage on viewing and creating reports that included metrics from multiple student assessments. Users in the largest cluster, Clust3, spent the largest proportion of their time viewing testing results and reports and filtering student groups. This largest group of Clust3 users appear to focus on behaviors directly related to the intended use of the system for classroom teachers, functions such as viewing overall interim testing results, or more specific reports related to item- or standard-level performance on interim testing. Interestingly, then, teachers in Clust1 and Clust2 spent a larger proportion of their online time outside of this core set of assessment features, either working with the creation and management of assessments or with reports that combine information from multiple assessments. Since Figure 22 operationalizes variety of use as a proportion of total use, it is impossible to say from this figure whether teachers' specialized uses in Clust1 and Clust2 represent *additional* uses of the system, above and beyond usage of core assessment reporting or *alternative* uses of the system, substituting for the core function of analyzing interim testing results. In either case, though, combining a disaggregated yet patterned approach to visual data analytics (Means et al., 2010) with metrics for usage variety (Shih and Venkatesh, 2004) has helped distinguish useful subgroups of teacher data use.

It is also important to note that while some groupings are similar between these first two heatmaps, particularly Clust1, overall the HCA heatmap by variety of use (Figure 22) finds different clusters of users than the HCA by frequency of use (Figure 21). These different outcomes suggest the importance of multiple usage outcomes when characterizing teacher online data use.

Clustering by online frequency and variety of use. The third and final HCA heatmap of Study 1 combines features from the previous two heatmaps, clustering along both dimensions of use diffusion theory: frequency and variety.

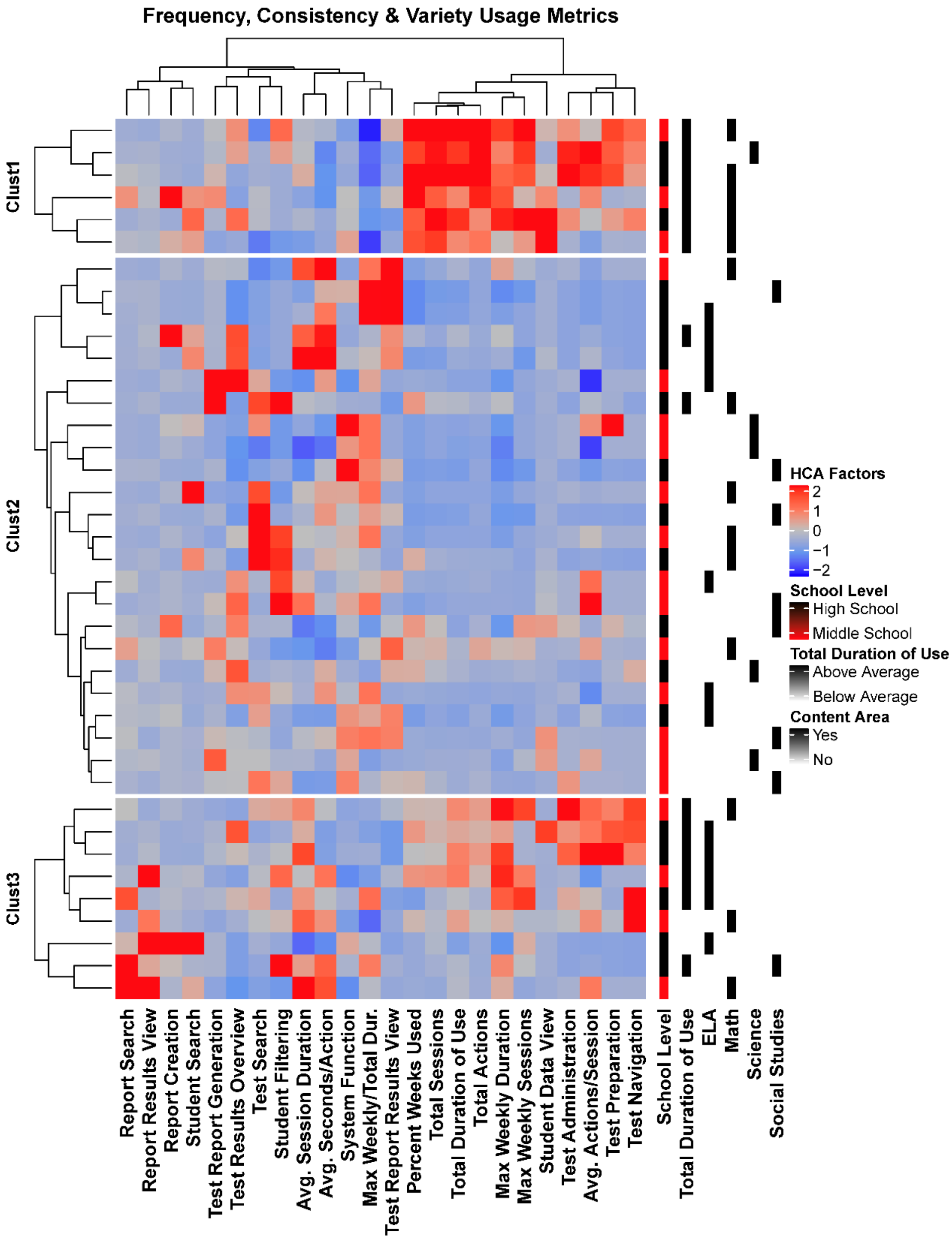


Figure 23. HCA heatmap: frequency, consistency, and variety of use. Eleven methods of NbClust suggested three as the best number of clusters. Visual inspection suggested three clusters as well.

As with the previous two heatmaps (Figure 21 and Figure 22), Figure 23 finds three main clusters for interpretation. Consistent with Figure 21 and Figure 22, users in Clust1 demonstrated some of the highest-levels of overall usage combined with a focus on test creation and administration. Clust2 users demonstrated lower-levels of overall use, while focusing on core functionalities for viewing and analyzing test results. Clust2 users also appeared to concentrate their usage during scheduled training sessions, as opposed to independently accessing the system over time. Clust3 users are divided into two subgroups, both demonstrating longer than average length of sessions. The upper group of Clust3 demonstrates similar patterns of use to Clust1, accessing functions for test creation and management, but with lower overall duration of use. The lower group of Clust3 overlaps with Clust2 from Figure 22, with users focusing on viewing and generating more complex reports that span multiple predictors of student performance. While their membership is generally mixed across middle and high school, the three clusters of Figure 23 demonstrate strong correlation to content area, with Clust1 relating to Math, Clust3 relating to ELA, and Clust2 containing almost the entire set of Science and Social Studies teachers.

Along with clusters of users, interesting clusters of factors are also generated in Figure 23. For example, after dividing into two main clusters, the cluster of columns on the left consists mainly of features related to viewing assessment results and reports, along with two features related to session length and one related to the degree that users limit their activity to one week of the semester (Table 21). In contrast, the second cluster of factors consists of several features related to total usage frequency alongside a few features related to test preparation and student data view. This mixing of dimensions of frequency and variety across clusters highlights the connections between what users choose to access in online data systems and how they access it, with higher levels of overall use related to testing preparation and features related to session length and focus in time more tightly clustered with assessment analysis functions (Table 21). The discussion section following will provide more interpretation of subgroups defined in Figure 23.

Table 21

HCA Clusters for Frequency and Variety Factors

	Cluster 1: Assessment and Report Results		Cluster 2: Test Administration and Frequency Metrics	
Sub-Cluster A	Report Search	[Variety]	Avg. Actions/Sessions*	
	Report Results View	[Variety]	Test Administration	[Variety]
	Report Creation	[Variety]	Test Preparation	[Variety]
	Student Search	[Variety]	Test Navigation	[Variety]
Sub-Cluster B	Avg. Session Duration		Percentage of Weeks Used	
	Avg. Seconds/Action		Total Sessions	
	Max Weekly/Total Dur.		Total Duration of use	
	Test Report Generation	[Variety]	Total Actions	
	Test Results Overview	[Variety]	Max Weekly Duration	
	Test Search	[Variety]	Max Weekly Sessions	
	Student Filtering	[Variety]	Student Data View	[Variety]
	System Function	[Variety]		
	Test Report Results View	[Variety]		

Note. *Shading indicates frequency or prevalence metrics

Discussion

The analyses of Study 1 yield interesting results in several areas: from the differences in aggregate usage metrics across school levels and content areas, to the relationships between usage over time and school-based training and assessment schedules, to correlations between usage factors, and finally to exploratory clusters of users and factors segmented by fine-grained indicators of usage frequency and variety. This variety and depth of description applied to online data use highlights the factors and outcomes of teacher data use most important to consider in future studies and provides insights to inform school data use interventions. This discussion will focus on the related three contributions below:

1. Expanding the quantitative descriptions of teacher interaction with online data systems, particularly in regard to visualizations and metrics for online behavior;
2. Exploring subgroups and patterns of teacher online behavior along dimensions of frequency and variety of use; and,
3. Informing school practice and software design to facilitate teachers' access and use of student data.

In considering these results, it cannot be stated strongly enough that the intention of these studies is to use teacher interaction data to find ways to make teachers' jobs more efficient and more effective, and not to evaluate teachers in any way. As Study 1 suggests, teachers' decisions whether and how to use an online system are informed by complex determinants, many of which are beyond teachers' control, and as previous studies suggest (Tyler, 2013; Wayman et al., 2017), these data use decisions may have a limited relationship, if any relationship at all, to teachers' core effectiveness in engaging students and helping them learn.

Expanding quantitative description of teacher interaction with online testing data.

Patterns in overall usage. While previous studies differ greatly in the usage metrics presented, some comparisons are possible. Tyler (2013), for example, in his district-wide analysis of an interim assessment system, calculates an average of seven hours of use by teachers over a 10-month period, or 0.7 hours per month. This study found an average of 3.95 hours of use over 6 months, or 0.65 hours per month. While the different timeframes under investigation present strong obstacles to comparison, these similar durations of average use are striking. On a different usage metric, Wayman et al., (2017) reports an average of 127.5 actions per user over 10 months, or 12.75 per month, while the current study found an average 202.3 actions over six months, or 33.7 actions per month. While at first glance, users in the current study appear to have used the online system for twice as many actions, it is also entirely possible, but difficult to establish, that differences in the level of actions logged by the two systems, and not actual differences in user behavior, account for this overall difference.

In terms of distribution of use over time, Study 1 finds a strong impact of required professional development and training on the timing of online usage. Figure 18 and Figure 19 clearly show stronger usage by multiple users in both the middle and high school for those weeks with professional development and training session. Even the variety of usage, as indicated by the color of squares in Figure 19, seems consistent across users as they participate in professional development and access similar reports. While these observations follow only from visual inspection of usage timelines, Tyler (2013) finds a related pattern where a third of teachers focused their use of the online system in one of four testing intervals. Tyler suggests that this pattern may be due to prompting by supervisors or the timing of school evaluation systems. In both studies, online usage for a large subset of users appears largely determined by institutional schedules, as opposed to personal or classroom-specific timing. While his findings are for principals and not teachers, Drake (2015) describes patterns where principals' usage of a data warehouse maps onto the timing of leadership decisions for teacher hiring and evaluation. These patterns in principal data use are clearest when contrasting the behavior of principals with high levels of usage with those demonstrating the lowest level of use.

Usage metrics, their relationships and importance. Along with patterns in overall usage, patterns and relationships between usage metrics also merit interpretation. In addition to metrics for online frequency and consistency of use analyzed in previous studies (Wayman, 2017; Tyler, 2013), Study 1 analyzes usage in relation to: time per action, actions per session, session length, maximum weekly duration, maximum weekly sessions, percentage of weeks used, total sessions, total actions, total duration, and maximum weekly usage/total usage. Many of these additional metrics were found to capture less correlated aspects of behavior or to differentiate between possible subgroups of online usage. These findings speak to the limitations of overall frequency or prevalence of use for capturing the most relevant factors of teachers' usage, particularly in those cases where multiple patterns of online behavior, such as more frequent/shorter sessions and less frequent/longer sessions, both resulted in larger than average duration of use.

Overall, little difference was found between metrics related to overall use, such as total actions, total sessions, total duration, and to a lesser extent, percent of weeks used. Though percentage of weeks used, or what Wayman et al. (2009) refer to as consistency of use, was slightly less correlated to total duration (r = .78) than was total actions (r = .92) or total sessions (r = .92) the correlation was still strong and all four of these factors clustered tightly together in Figure 21's HCA heatmap of frequency and consistency metrics. Since weeks used presents a less skewed distribution than total actions, durations, or sessions, lends itself to easier interpretation, and is the only usage metric to demonstrate a significant relationship with student outcomes (Wayman et al., 2017), this metric may be the most useful of the four measures of total usage.

Alongside total usage, metrics for session-level usage and concentration of usage provided description of additional aspects of user behavior. HCA Clusters in Figure 21 depend on values for these usage metrics and later in Figure 23 and Table 21 these same factors of usage cluster alongside different varieties of user behavior, indicating their possible relevance beyond empirical description of user behavior. Maximum weekly duration and maximum weekly sessions, in contrast, appear to add little in clustering analysis (Figure 21 and Figure 23) to clustering of user behaviors and may not prove as useful for future analyses. Table 22 summarizes some of the most promising metrics for capturing user differences in frequency and consistency of user behavior.

Table 22

Summary of Useful Metrics for Online Frequency and Consistency of Use

Type of Description	Usage Metric
Total Usage	Percentage of Weeks Used
Session-Level Usage	Time per Action Time per Session Actions per Session
Concentration of Usage	Maximum Weekly Usage/Total Usage

Patterns in middle/high school and content area usage. One of the more critical takeaways from Study 1 appears to be the degree to which school level, in this case middle and high school, as well as content area are deeply related to teacher use of online testing data. Some of these differences are most likely due to how state-level accountability impacts assessment. Particularly in middle school, lack of online access by Science and Social Studies teachers is easily explained by a lack of end-of-year state testing and, as a result, a lack of benchmark testing in these areas. While state testing results, computer adaptive testing results, or the literacy focus of ELA interim testing might be useful, at least hypothetically, to middle school Social

Studies or Science teachers, these teachers do not appear to access the available reports much beyond the minimum for required training. However, in middle school Math and ELA, both content areas with interim assessments and end-of-year state tests, teachers demonstrate large differences in online usage, with Math teachers completing more than twice as many sessions as ELA teachers on average and completing more actions within those sessions.

Though neither of these previous studies includes high school usage, Tyler (2013) and Wayman et al. (2017) describe similar usage differences between Math and ELA, with higher levels of use for math versus reading teachers. On the other hand, where the current study found higher online use for high school than middle school, Gold et al. (2012) briefly describes higher usage for middle school over both elementary and high school, though the data system they investigated was a platform for summary student data and not, specifically, for analyzing interim testing results. Given their data and functionality, direct comparisons between the two systems may not adequately consider important differences.

Extremely limited research on high school data systems (Gold et al., 2012), combined with the accountability structure for New York State, make Study 1 findings on high school data use particularly interesting. Even under a New York state-wide assessment system where all four core content areas are required to administer multiple end-of-year competency tests, and where interim testing was conducted at Progress Secondary in all four core content areas in preparation for these tests, large differences in usage are still found across content areas, with average online use in Math and Science exceeding that of ELA and Social Studies by a large amount.

Along with these differences in total frequency of use, Study 1 also finds that content area teachers may demonstrate different patterns of use at multiple levels, during sessions and even individual actions. For example, while average use is similar in high school Math and

Science across multiple indicators, ELA and Social Studies demonstrate several differences, with Social Studies teachers spending a similar amount of time per action but completing fewer actions per sessions for an overall lower duration of use.

Another interesting distinction is the patterns of usage by which Math and ELA teachers reach high overall durations of use. As indicated in Figure 21, while Math teachers appear to reach high levels of usage by either (a) more frequent/shorter sessions or (b) less frequent/longer sessions, ELA teachers appear to almost exclusively follow pathway (b), with less frequent/longer sessions, to achieve higher levels of use. While many have discussed the need for increased time dedicated to data use (Schildkamp and Kuiper, 2010; Datnow et al., 2012) it is interesting to consider, based on these results, that Math and ELA teachers might benefit from not just more time for data use, but different distributions of time. Paralleling their observed online usage, Math teachers may be able to effectively utilize shorter more frequent sessions to analyze multiple discrete items on a test, while ELA teachers may benefit from more extended sessions to better understand extended reading passages and analyze student open-response writing.

In order to interpret these findings more fully—both in terms of overall usage and session-level behaviors—more qualitative analysis analyzing teachers' perspectives on their own usage alongside more intensive study of individual clickstream patterns would be helpful. More extensive research on classroom data use establishes the general importance of pedagogical content knowledge (Blanc et al., 2010; Datnow et al., 2012; Fuchs et al., 1999; McNaughton et al., 2012). However, even in the area of classroom data-use, few studies appear to directly address differences by content area (Hoogland et al., 2016). So, while the overall importance of content area goals and knowledge seems clear in both classroom and online data use, additional studies are needed in both contexts to unravel how exactly the dynamics of DBDM play out in specific disciplinary contexts.

One speculative hypothesis, at least regarding online data use, builds on the fact that different content areas depend on different types of items and methods of analysis within online assessment systems. These content-specific items and analyses in turn may encourage different patterns of online use. Even when analyzing similar standardized, multiple-choice testing results, Math teachers' online interactions with stand-alone multiple-choice items may appear noticeably different from ELA teachers' online interaction with small groups of multiple-choice items, which directly refer to preceding reading comprehension passages of varying length and complexity.

Differing hierarchies of state standards may also make a difference for online interaction. While Math teachers can generally place each Math test item on a ladder of difficulty tied to grade-level standards, ELA teachers must place items in reference to much broader reading comprehension standards, as well as take into account the difficulty of a reading passage in multiple genres. Or, in the case of Science and Social Studies, online interaction may look different where multiple-choice items assess students' knowledge of domain-specific content, as opposed to students' ability to perform disciplinary skills, such as interpreting texts or performing calculations. In a similar vein, Hoogland et al. (2016, p. 382) report that focus group discussions on classroom data use emphasized the importance of "clear insight into curriculum" and "clear, continuous and identifiable learning trajectories." That state-level math standards come closer to providing a clear learning trajectory may help explain the tendency of Math teachers to make greater use of online assessment data aligned to these standards.

Exploring subgroups of teacher behavior related to frequency and variety of online usage. Analysis of HCA heatmaps in Figure 21, Figure 22, and Figure 23 reveals additional subgroups of teacher online behaviors distinct from differences based on school level or content area. Features clustered along dimensions of frequency and variety, and therefore aligned to use diffusion theory (Shih and Venkatesh, 2014), prove valuable in Study 1 for exploring subgroups of teacher online behaviors. Additionally, the disaggregated visual analytics of the heatmap view allow for more extensive interpretation of the distribution of behaviors across the population than do measures of central tendency and dispersion alone. For example, a metric indicating low average access to an online report may hide the underlying pattern that a subgroup of users accesses this report intensively, while other users simply do not. Visual analysis of HCA heatmaps, in contrast, allows for the identification of these possible subgroups of users and behavior, as well as tentative interpretation of subgroups based on the clustering of factors. Additional features of the HCA heatmap, such as the standardization of factors and inclusion of annotations also enhance their utility for effective visual analysis.

Following this visual analytics approach, Study 1 explores the possibilities of use diffusion theory to generate clusters of behavior along the dimensions of frequency and variety of use. Based on Maull's (2013) methods and findings, analysis in Study 1 was not aimed at grouping usage into one of the four quadrants defined by the intersection of frequency and variety. Instead, multiple features related to frequency and variety of use were used to cluster users and behaviors in a search for potentially meaningful subgroups of online behavior. Figure 24 summarizes some of these exploratory subgroups of usage identified in Study 1 and places them in the context of use diffusion factors. Overall, several clusters can be interpreted as instances of limited, specialized, or intense use. Interpreting clusters as nonspecialized, however, proved difficult. Maull (2013) notes a similar challenge where the positive relationship between frequency and variety makes it difficult to find users who have both low frequency and high variety of use. As usage increases, the variety of pages accessed tends to increase as well.

		Rate of Use	
		High	Low
Variety of Use	High	Intense Use -Assessment Management & Analysis (Intensive)	Nonspecialized Use
	Low	Specialized Use -Multiple Measures Analysis -Assessment Management -Student-Centered Analysis	Limited Use -Assessment Analysis (Training-based)

Figure 24. Exploratory data use clusters related to use diffusion typology

Each of the exploratory subgroups in Figure 24 is briefly discussed in the next section. Though these clusters are often quite small, within an already small population, they find parallels in previous typologies related to teachers' use and attitudes towards data and technology. These relationships to previous work will be discussed in the sections below, as well as summarized in Table 23.

Interpreting limited use. The largest cluster of teacher behaviors identified in this study follows a pattern of limited use. Visible as "Assessment Analysis (Training-Based)" in Figure 25, this subgroup is defined by a low rate of overall use, combined with low variety of use, as its members focus on the core assessment features of the system: Test Results Overview, Test Search, and Student Filtering. In terms of frequency of use, this same cluster tends to have higher values for Maximum Weekly Use/Total Use, indicating a probable focus of use during required training sessions, where teachers viewed various reports on recent interim testing in a guided setting.

This subgroup of limited use finds parallels across previous typologies, from Maull's (2013) "Uninterested Non-Adopter", to Wayman et al.'s (2009) teachers "Opposed to Data", to Graves and Bowers (2018) "Evaders". While based in slightly different domains—online curriculum use, general data use attitudes, and technology use, respectively—these categories all indicate limited use of data and/or data use technologies. While this group of users may be proportionally smaller when considering teacher technology use more generally (Graves and Bowers, 2018) or when considering online data or curriculum use, the category of limited use appears to apply to a larger percentage of users (Maull, 2013; Tyler, 2013, Wayman et al., 2012, Gold et al., 2012). Study 1 found 62% of users to fall into the limited use category: Assessment Analysis (Training-Based). While this proportion of limited use is larger than in other typologies, it may be that required organizational trainings effectively combined limited and moderate users into one category, where in Maull, 2013, for example, they cluster in two separate groups.

At the same time, a separate typology, generated from a large-scale survey of teachers' attitudes towards data use and data use technologies (Bill & Melinda Gates Foundation, 2015), found multiple types of users within, what might be considered, the larger limited use category. Within the Gates Foundation typology, Aspirational Users, Perceptives, and Traditionalists all consider themselves limited users of data, but for different reasons. Aspirational Users believe generally in data use but find applied use difficult; Perceptives prefer to use their own observations when making decisions about their teaching; and Traditionalists prefer to use student performance captured in the form of grades when making instructional decisions. While Study 1 only analyzes teachers' observed data use, it is interesting to consider, as the Gates

Foundation Typology suggests, that different attitudes about and obstacles to data use may manifest behaviorally in the same large category of Limited Use.

Understanding specialized use. Several smaller clusters of specialized online behavior are identified in Figure 25, specifically clusters of online use related to “Student-Centered Analysis,” “Multiple Measures Analysis,” and “Assessment Management,” The subgroup for Multiple Measures Analysis tends toward a longer session length, with usage focused on viewing student data gathered from multiple assessments into a single tabular view. One example of a multiple-measures report might be one that collects scores from several assessments related to reading: ELA state testing data, computer adaptive test scores in reading, and results from a decoding assessment.

A second specialized usage category, “Assessment Management,” demonstrates more intensive use of functions related to test preparation, administration, and scoring. More specifically, these functions might include activities such as selecting items from an item bank, formatting answer documents, linking items to standards, or using a document camera to capture test results. The use of such assessment management functions would be highly dependent both on their availability within the online system, as well as on school-level policies for system use. Nevertheless, with many online testing systems offering teachers’ functionality to create their own tests, actions devoted to assessment management are important to include as part of the time devoted to DBDM.

A final subgroup of specialized use, Student-Centered Analysis, privileges searching and viewing student data at the level of the individual student. As opposed to reporting multiple metrics for multiple students, student data views provide more comprehensive reporting on student metrics for only one student. This tendency to view individualized student data views may be comparable to the desire of the Data Maven subgroup (Bill & Melinda Gates Foundation, 2015) to view holistic student data on multiple measures. While Tyler (2013) finds that teachers tend to access student-level data at very low rates, Study 1 raises the question of whether this low level of access to student-level data represents generally low access or, instead, a small subgroup of users intentionally accessing student-level data, while the vast majority fail to use such views.

While the domains under investigation are slightly different, categories of specialized use, such as those identified in Study 1, have been identified in several previous typologies. Wayman et al. (2009), for example, identifies a category of teachers as using "Data as a Supplement," suggesting that data use may have been incorporated into teaching practice in a specific, specialized context. Past typologies frequently identify substantial categories of users or behaviors focusing on specialized uses of data, whether they are Interactive Resource or Community Seeking Specialists engaged with an online science curriculum (Maull, 2013), Scorekeeper teachers focusing their use of data on assessment (Bill & Melinda Gates Foundation, 2015), or the specialized Assessors and Presenters found in a nationally-representative survey of teacher technology use (Graves and Bowers, 2018).

The impact of roles and responsibilities. An important factor for understanding specialized usage patterns is the possible impact of teachers' additional roles and responsibilities. While not explicitly identified in data collection for the current study, teachers frequently fulfilled additional roles and responsibilities within the school. These roles may have been linked to a job title, such as Learning Specialist, assigned as part of committee work, or contracted for outside of regular school hours. Related data-use tasks may have included monitoring student performance by grade level or by demographic groups, reporting information on students' Individualized Education Program (IEP) for Special Education services, advising on class groupings, or helping create, administer, or score schoolwide tests. Study 1 finds several clusters of specialized use, with some teachers devoting a larger proportion of their time to managing assessments, viewing individual level student data, or analyzing groups of students by multiple student measures. Without additional teacher data it is impossible to determine the degree to which specialized use may be due to additional assigned responsibilities, as opposed to individually-driven inquiry.

Similarly, Gold et al. (2012), found large differences in usage between classroom-based teachers and teachers with schoolwide roles and responsibilities. Teachers with schoolwide roles were found to use the ARIS online data system for more than twice the duration of classroom-based teachers. Such specialized teacher roles and responsibilities may play a large role in determining online usage among teachers and may be at the root of usage differences in the current study as well. Eliciting more information on teachers' specialized responsibilities may be critical in explaining variation in online data use, especially in interpreting the behaviors of high frequency users. Are such users highly invested Data Mavens and Growth Seekers (Bill & Melinda Gates Foundation, 2015), independently-focused on classroom-level DBDM, or do they have additional professional responsibilities that require access to online student data, or both?

Schildkamp et al. (2017) highlights a related issue for general data use, where the same actors can use data for multiple purposes, in this case for accountability, school improvement, or instructional purposes. The authors report that schools tend to make greater use of data for school improvement and accountability purposes, as opposed to instructional ones. If data use for school improvement and accountability tends to be linked more directly to specific teacher roles and responsibilities, this may be one possible explanation for schools' greater use of data in these areas than for guiding instruction.

Differentiating intensive use. While high frequency of use appears related to several specialized patterns of teacher access, one subgroup, “Assessment Management and Analysis (Intensive)” (Figure 25) appears to use the system with higher frequency across assessment management and assessment analysis functions. Even these users, however, do not appear to equally distribute their usage across categories, but instead spend the greatest proportion of their time in assessment management functions. While the most intensive users, then, may also demonstrate greater variety of use, it is possible that the category of “Intensive Use” might be better conceptualized as a more intensive form of specialized use, or as a category that combines functions of specialized use. The Assessment Management and Analysis (Intensive) category, for instance, appears to combine higher usage for assessment management and assessment analysis, but not for student-centered views or reports of multiple measures. The variety of use for these teachers, then, does not stretch to all categories, but only encompasses a subset of functions.

The Gates Foundation typology (Bill & Melinda Gates Foundation, 2015) appears to support this diversified view of the intensive user by finding two categories of the most intensive data users, the Data Mavens and Growth Seekers. While both types are intensive users of a wide range of data, they differ in that Growth Seekers include a larger inclination to apply data as a means of self-reflection for improving practice.

Overall, a visual analytics process employing HCA heatmap views in a local school context has provided interesting findings that both support and qualify earlier typologies and categories of teachers’ engagement with data use and technology. Table 23 provides a summary of the typologies referenced in the discussion above alongside the exploratory clusters found in Study 1.

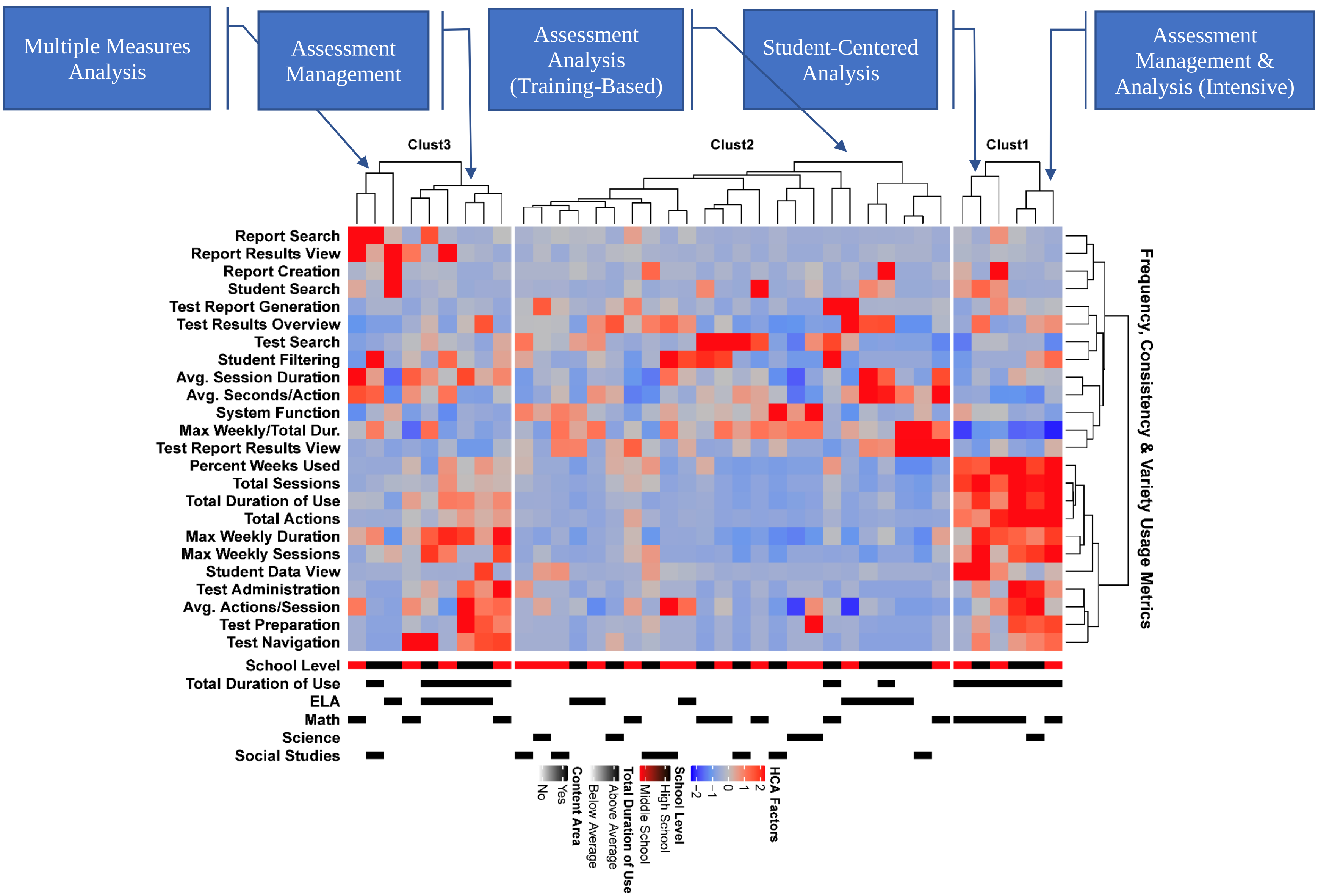


Figure 25. HCA heatmap: possible subgroups of frequency, consistency, and variety of use

Table 23

Comparison of Typologies for Teacher Data Use and Technology

Teacher Typology	**Types of Teachers, Attitudes, or Usage**			
Use Diffusion Typology (Shih & Venkatesh, 2004)	Limited Use: Low frequency Low variety	Nonspecialized Use: Low frequency High variety	Specialized Use: High frequency Low variety	Intensive Use: High frequency High variety
Data Use Attitudes (Wayman et al., 2009)	-Opposed to Data		-Data as a Supplement	-Data is Essential
Usage of Online Science Curriculum (Maull, 2013)	-Uninterested Non-Adopter (32%)*	-Moderate Generalist (36%)	-Interactive Resource Spec. (10%) -Community Seeker Spec. (14%)	-Ardent Power User (8%)
Attitudes to Data and Data Use Technology (Bill & Melinda Gates, 2015)	-Perceptives (14%) -Traditionalists (10%) -Aspirational Users (17%)		-Scorekeepers (11%)	-Data Mavens (28%) -Growth Seekers (20%)
Technology-Using Teachers (Graves & Bowers, 2018)	-Evaders (22%)		-Assessors (28%) -Presenters (25%)	-Dexterous (24%)
Teacher Online Data Use (Current Study)	-Assessment Analysis, Training-based (62%)		-Assessment Management (15%) -Student-Centered Analysis (8%) -Multiple Measures Analysis (8%)	-Assessment Mgmt. & Analysis, Intensive (8%)

* Where available the percentage represented by each subcategory is presented in parentheses.

Informing school and software design decisions to facilitate teachers' access and use of student data. While Study 1 describes only a small sample of teacher behaviors, that sample provides rich information for school leaders and software designers as they support instructional decision-making. Analyses from Study 1 might inform decision making at several levels, from overall school process, to grade-level or content-team planning, to specific teacher roles and responsibilities.

Most fundamentally, with 62% of teachers clustered in a limited use category, findings from Study 1 support past findings on the low rates at which most teachers access online student data (Wayman et al., 2012; Tyler, 2013, Gold et al., 2012) or are satisfied with data use technologies (Bill & Melinda Gates Foundation, 2015). At the same time, district investment in a variety of online data and assessment system continues to climb steeply (Molnar, 2017), making the need to increase meaningful teacher interaction with student data ever more pressing. As more scarce resources are diverted to DBDM solutions to educational problems—such as closing the achievement gap or meeting state goals for student proficiency—these solutions must see stronger outcomes than they have thus far (Tyler, 2013; Wayman et al., 2017).

The exploratory analyses of teacher usage in Study 1 suggest several possible ways to increase meaningful interaction with online student testing data, primarily by adapting to usage patterns expressed by content areas and specialized subgroups of teacher online behaviors.

Supporting content area needs. Study 1 suggests that teachers' decisions to access testing data may be related to a web of organizational factors, such as school level, content area, and specialized professional responsibilities. Differences in access by content area in particular, as demonstrated in aggregate statistics (Table 17 and Table 18), and in HCA results (Figure 21 and Figure 23), bear further investigation. While Math teachers have the highest rates of use, lower rates of use for ELA, Science, and especially Social Studies teachers indicate areas for possible collaboration and partnership between content teams/content specialists and educational data scientists, mapping the discipline-specific needs of data use and testing ways in which those differences can be accommodated and leveraged in school processes and online software. Content-specific frustrations and success with classroom-based formative assessment, especially in Science and Social Studies, would seem a fruitful area for research-practitioner partnerships.

Supporting access for subgroups of specialized teacher access. Since the decisions of teachers with specialized roles and responsibilities may have impact across classrooms or student groups, informing these decisions with data may have a highly leveraged impact. Knowledge of subgroups of specialized use, such as those identified in HCA heatmap Figure 22 and Figure 23, may have multiple uses for both schools and software designers, from identifying content for more specialized, school-based training, to the eventual creation of recommender systems (He, Parra, and Verbert, 2015) focused on the recommendation of student information and analyses to teachers. Such systems might provide automated prompts and suggestions based on recommender algorithms that use explicit labels (of users and content) alongside implicit user behaviors to make recommendations based on the preferences of similar users and the similarities of content items (He, Parra, and Verbert, 2015).

In the Benchmark Data system, currently, a system user with sharing privileges intentionally assigns reports to users, many with possibly unknown, temporary, or ad hoc roles and responsibilities. Perhaps eventually, Benchmark and similar systems focused on data and assessment reporting for teachers, might combine analysis of teacher online use with system-delivered surveys to help define the kind of reports and access that each educator finds most useful. A brief survey, for instance, on the perceived usefulness of content area information, types of assessments, student-centered reports, multiple measures reports, and other forms of student data might feed system functionality in recommending either reports, available metrics, or training opportunities to teachers, while also providing guiding information to system designers and instructional leaders about the analyses and evidence that teachers search out and find useful. Teachers' rankings of assessment reports and views, combined with basic information about their roles and responsibilities might prove a powerful feedback tool for system designers and educational leadership, as well as useful input for information and reporting-based recommender systems for educators.

The HCA findings and identification of teacher subgroups in Study 1 and Study 2 may help address some of the existing challenges for recommender systems around transparency, cold start issues, and identification of important contextual information (He, Parra, Verbert, 2015).

Supporting through training and school structures. Also at the school level, Figure 16, Figure 17, and Figure 19 suggest that the importance of in-person training and sharing of results is difficult to overestimate. The most extensive teacher usage occurs during structured sessions of training and sharing of results. No such increase in activity occurs after links to testing results are emailed to teachers. While increasing the amount and consistency of required trainings might be the initial response to such findings, it is important to carefully manage any increased demands on teachers' time in terms of the quality and the focus of training, teachers' ongoing evaluation of such professional development, and, most critically, whether any required increases in time spent accessing student data result in meaningful analysis, actions, and improvement of student engagement and learning.

Supporting through meaningful distribution of test results. At the same time, in-person meetings require substantial effort to organize, implement, and attend, so finding more effective asynchronous means of distributing test results may be an important line of inquiry as well. Focus groups and structured interviews conducted around varying ways of distributing results might be an interesting place to start for identifying factors impacting teachers' decision to take a closer look at test results.

Managing assessment management. Several of the most intensive users in Study 1 dedicate a large proportion of their usage towards creating and managing tests (see Figure 23), as opposed to viewing and analyzing testing results. While teachers' participation in the preparation of schoolwide assessments may build valuable skills, increase transparency around testing, and foster teacher buy-in, the demands of such involvement may also detract from teachers' use of online features dedicated to planning for their own classroom instruction. Tracking this overall balance of time spent on administration versus analysis over multiple years may help schools better prioritize their time and effort in relation to online testing systems.

What is "successful" online use? While Study 1 makes advances in identifying possible subgroups of teacher online usage, it is unable to support inferences on the success of that usage, from either the user or an organizational perspective. The inability of this and other studies of online data use to identify or effectively discuss "successful" teacher interactions raises the need for another related area of inquiry, the evaluation of successful use of the system, a topic related to web analytics' focus on conversion (Kaushik, 2010; Dumais, et al., 2014). While in e-commerce terms, "conversion" often refers to an online visitor clicking on an ad and going on to make a purchase, the applications of the concept can be considered much more broadly in terms of the proportion of visits or visitors that complete their desired task or a task that the organization feels important. Commercial web analytics is clear in defining the metrics for "successful" access, as distinct from metrics for basic frequency or even variety of use (though these metrics may play a role in measures of success) (Kaushik, 2012). The usage of a customer who goes online and quickly purchases the exact book they want is an online success, even if both frequency and variety of use were lacking in the interaction. In a similar way, teachers may make highly successful and even impactful use of an online data system, by quickly and efficiently retrieving important student data, even if such an interaction would demonstrate a low duration and variety of use.

Thinking critically about and defining “successful” user interactions for online data use systems could open log file analysis to more actionable interpretations by schools. Clearly, reproducible and actionable definitions of “success” would be a work in progress and heavily impacted by what teachers themselves deemed as a successful interaction with the system. Yet, the process of identifying and monitoring even tentative markers of prescriptive success, such as accessing a report within a particular time frame, might provide a foothold for logic models that connect teachers’ use of student data to more critical teacher and student outcomes.

Another possible method for evaluation of data and assessment systems is to apply existing frameworks for evaluation of learning analytics dashboards or online curriculum resources (Bodily and Verbert, 2017; Snodgrass et al., 2015; Yoo, Lee, Jo, and Park, 2015)

Limitations

Generally, Study 1 is limited by its use of exploratory and descriptive analyses, analyses that do not test any hypotheses and cannot support any causal inference. Study 1 is also limited by its small sample size, which makes generalizing any findings impossible. The use of log file analysis also introduces several limitations to Study 1. While the use of log files attempts to address an over-dependence on survey methods in studies of DBDM, log files of teachers’ data use have their own severe limitations. Teacher access to an online system is, in the end, only that: time spent on a website. It is not observation of actual data use that impacts teaching or even evidence that a teacher has seriously considered the evidence presented on their screen.

This analysis of log files was also limited by the use of an expert-derived categorization of web functions and pages and the level of specificity of the analysis. Empirically-derived categories for web pages or more fine-grained analyses of teachers' online process may have arrived at useful distinctions, such as distinguishing between teacher use related to specialized roles and responsibilities and use related to core classroom instructional responsibilities.

Along similar lines, log file analysis itself has limitations, particularly in capturing the time of the last action in a session or in identifying and discounting over-long actions where a user is no longer engaged with the website. The choice of Study 1, to not estimate a time for the last action in each session, will have underestimated usage for all users in ways that are difficult to predict. Metrics for variety of use may also have limitations due their calculation as a proportion of online access and not an absolute measurement.

Another limitation of Study 1 is the amount of subjective interpretation involved in HCA, from the somewhat subjective determination of the number of clusters for interpretation to the interpretation of the similarities and differences between subgroups.

Conclusion

As noted above, this study has severe limitations. Even with these limitations, though, the results and analyses provide previously unseen glimpses of teachers' online use of student data. Though similar studies exist, no other study of online teacher data use that I have been able to find has compared teachers' data use between middle and high schools and across content areas, critical organizational features of districts and schools. No other study has explored teachers' data use across the combined metrics of frequency and variety of use or has visualized these factors beyond the representation of central tendency, a representation that has tended to obscure variations in use across important subgroups of teachers. Given the size of the online assessment market, projected at 1.73 billion by 2020, it is imperative that more such public, exploratory descriptions be generated, to inform the design of online data systems to the benefit of teachers and to inform school and district leaders of realistic expectations for adoption and use as they investigate, adopt, and implement future systems for DBDM.

Study 2: Connecting Teacher Roles and Data Use Attitudes to Online Behaviors

Building on Study 1, Study 2 expands the methods, theory, and participation of research into Data-Based Decision Making (DBDM). Specifically, Study 2 employs methods for visual data analysis employed in learning analytics alongside theories of self-efficacy and technology acceptance to explore determinants of teachers' online use of testing data. Where previous large-scale studies have been unable to find consistently significant determinants of teachers' online data use (Table 6 and Table 7), this small-scale, descriptive study, embedded in a local school context generates exploratory metrics and determinant frameworks for teacher data use that may inform future studies.

Toward the end of the 2015-16 school year, core content teachers at Progress Secondary participated in surveys exploring their frequency of data use, perceived usefulness of school software programs and data types, general data use attitudes, and self-efficacy for data-driven decision making. While any staff member interested in providing feedback was encouraged to do so, only the responses of core content teachers were included in the current study.

Purpose and Research Questions

Only a limited number of studies have explored the relationships between teachers' data use attitudes and teachers' online data use behaviors (Wayman et al., 2009, 2011; Shaw, 2010). Wayman et al. (2009, 2011) include findings from the Survey of Educator Data Use (SEDU), one of the surveys employed in Study 2. Shaw (2010) administered a survey on Teaching Self-Efficacy (Tschannen-Moran et al., 1998), subscales of which are also used in Study 2. These two studies found contradictory results regarding the significance of general data use attitudes (See Table 7), along with surprising relationships where both a subscale related to data's effectiveness for pedagogy (Wayman et al., 2009) and a teaching self-efficacy scale (Shaw, 2010) were found negatively related to online use (p = .01 and p = .05, respectively). Study 2 revisits possible relationships between these survey constructs tested previously and attempts to expand on their results by exploring relationships between teachers' online usage and self-efficacy in data-driven decision making. Study 2 also applies the lens of the technology acceptance model to teachers' online data use and explores more nuanced visual analyses of the relationships between these factors. Study 2 addresses the following research questions:

(R2) What attitudes do teachers have towards data use?

(R3) To what extent are teachers' data use attitudes, technology acceptance, self-efficacy, and roles related to their online use of student testing data?

In addition to these questions, the Discussion section of Study 2 explores ways in which schools might leverage log file and survey data to evaluate school data use, plan professional development, and inform decision-making.

Methods

Participants. Of 42 total core content teachers, 37 (88%) participated in the survey, and 35 (83%) participated in both the survey and logged into the online Benchmark Data system at least once.

Survey Methodology.

Recruitment and consent procedures. Participants were recruited through a principal letter, and with the principal's permission, two voluntary sessions were held for the online administration of the survey in the spring of the 2015-16 school year. The voluntary nature of the survey was emphasized.

Survey procedures. After participants signed the consent form, they were given a unique participant ID, a choice between online and paper administration, and reminded to use their participant ID when responding to the survey. No personally-identifiable information was requested as part of the survey, which took approximately twenty-five minutes to complete.

Teacher background factors. For general background, the survey asked participants about the number of years they had been employed in education, which position they held currently at Progress Secondary, and which grades they worked with.

Technology acceptance factors. According to the technology acceptance model (TAM), the factors of Perceived Usefulness (PERUSE) of a system and the Perceived Ease of Use (PEOU) are two of the most important influences on technology acceptance (Davis, 1989). In this study PERUSE was operationalized as participants' perceived usefulness of the Benchmark Data system, as well as the perceived usefulness of relevant data types. Perceived Ease of Use was operationalized as items indicating participants' general comfort level with technology and their level of use for the Benchmark Data System (possible responses for both items were: "Non-user," "Novice User," Average User," or "Expert User"). Additional items for the PEOU factor were drawn from those assessing self-efficacy for DBDM-related analyses. While some studies separate PEOU and self-efficacy as highly-related, but distinct factors (McFarland and Hamilton, 2004, Venkatesh, 2000), computer self-efficacy has consistently been found to be a strong significant determinant of PEOU (Venkatesh and Davis, 1996, Venkatesh, 2000). In its earliest conception, in fact, the Technology Acceptance Model (TAM) explicitly drew upon Social-Cognitive Theory and Bandura's construct of self-efficacy as an expression of internal control, related to the PEOU construct (Davis, 1989). Given the strong, previously identified relationships between these constructs and the exploratory nature of Study 2, self-efficacy items related to DBDM are included when considering teachers' perceived ease of use for the online DBDM system.

Survey instruments and subscales. Surveys contained subscales from several different instruments: the Survey of Educator Data Usage (SEDU) (Wayman et al., 2009b, Wayman, Cho, Jimerson, and Spikes, 2012), The Data-Driven Decision Making Efficacy and Anxiety Survey (3D-MEA) (Dunn, Airola, Lo, et al., 2013b), the Teacher Self-Efficacy Scale - Short Form (TSES) (Tschannen-Moran and Woolfolk Hoy, 2001), and the Norwegian Teacher Efficacy Scale (NTES) (Skaalvik and Skaalvik, 2010). Table 25 summarizes subscales and reliability metrics for these instruments. Please see Appendix A for complete survey scales, where publicly available.

The Survey of Educator Data Usage (SEDU) was used to capture teachers' attitudes towards several aspects of data use. Responses for all scales in the SEDU were set on a 4-point Likert scale ranging from 1 to 4, with responses appropriate to each question. Participants' scale scores were created by averaging their responses to all items in a scale, with item scores ranging from one to four. Descriptions of each scale are presented below:

The **Data's Effectiveness for Pedagogy** Scale included five items which asked educators how well data could improve specific aspects of instruction, including planning, identifying learned concepts, and selecting learning goals.

The **Data Use Practice** Scale used five items to assess the degree to which teachers apply data use to instruction. Items asked educators to agree/disagree with statements about their use of data to plan lessons, identify learning needs, and alter instruction.

The **Data Attitudes** Scale included four items about teachers' high-level approach to data, whether they find data useful and whether it helps their instruction.

The **Computer Data Systems** Scale included four items relating to computer systems in use at the school, specifically their ease of use, their efficiency, and to what degree they provide a diversity of data.

With six items, the **Supports for Data Use** scale questioned teachers about how supported they felt in data use through professional development and dedicated staffing.

Framed differently than other SEDU scales, the **Instructional Resources Scale** captured how frequently educators reported using data for eight instructional purposes, such as forming small groups, tailoring instruction, or identifying learning goals.

Other selected items from the SEDU did not represent general constructs, but instead gathered teachers' perspectives on school-specific data systems and types. These school-specific items were grouped into categories of overall comfort with technology, reported frequency of data use, perceived usefulness of data types, and perceived usefulness of software programs. Specific items for each category are summarized in Table 24.

Table 24.

School-Specific Items from the Survey of Educator Data Use (SEDU)

Response Options	Items for Response
Comfort with technology and level of use	
(1) Non-user	1. Student Information System
(2) Novice user	2. Benchmark Data
(3) Average user	3. Google Drive
(4) Expert user	4. Google Classroom
Perceived usefulness of data types for understanding students and targeting instruction	
(Null) I'm not sure what this data is	1. Your own assessments
(1) Not at all useful	2. Your own observations
(2) Slightly useful	3. Your own gradebook/records
(3) Somewhat useful	4. Data from students' use of software
(4) Very useful	5. Students' performance on State Standards

(5) Extremely useful	6. IEP Information
	7. Students' Past grades
	8. Interim Testing Data
	9. Progress Monitoring Quiz Data
	10. Computer Adaptive Test Data
	11. Data on Student's Cognitive Skills
	12. ELA and Math State Test Scores
	13. Regents Testing Data
	14. Statistical predictions about how students will perform on state tests
	15. Attendance and Tardiness Data
	16. Discipline and Behavior Point Data
	17. Statistical predictions about whether students are at risk for dropping out
	18. Lexile Levels
Perceived usefulness of software programs for accessing data	
(Null) Not applicable	1. Student Information System
(1) Not at all useful	2. Google Docs and Sheets
(2) Slightly useful	3. Benchmark Data
(3) Somewhat useful	4. Google Classroom
(4) Very Useful	
(5) Extremely Useful	
Reported frequency of data use for specific types of data	
(Null) Never	1. Your own assessments: tests/quizzes/assignments
(1) Less than once a month	2. Discipline data: discipline incidents/points
(2) Once or twice a month	3. Student Software data
(3) Weekly or almost weekly	4. Formal assessments: Interim and computer adaptive
(4) A few times a week	5. State Achievement Tests: ELA/Math/Regents
	6. IEP Goals/Information

Instead of general attitudes towards data use, subscales for the DDDM Efficacy and Anxiety Survey (3D-MEA) (Dunn, Airola, Lo, et al., 2013b) assessed teachers' self-efficacy regarding specific data use tasks by asking their level of agreement on a five-point Likert scale, ranging from Strongly Disagree to Strongly Agree. Subscales administered from the 3D-MEA included the following:

Efficacy for Data Analysis and Interpretation assessed teachers' confidence in understanding assessment reports and interpreting scores representing student performance.

Efficacy for Application of Data to Instruction asked teachers about their confidence using data to guide instruction, specifically their confidence in identifying students with special needs, gaps in curriculum, and gaps in student understanding.

The **Efficacy for Data Identification and Access** subscale assessed teachers' self-efficacy for accessing state testing results and for knowing which reports to access for understanding student performance.

The **Efficacy for Data Technology Use** subscale asked teachers about their confidence in using the data systems provided by the school to retrieve information, filter students, and search for standards.

Along with the SEDU and 3D-MEA, subscales of the Teacher Self-Efficacy Scale (TSES), the Norwegian Teacher Self-Efficacy Scale (NTSES), and the Norwegian External Control Scale (NECS) provided insight into more general areas of teacher self-efficacy deeply related to data use. The TSES asked teachers about their ability to perform a range of teaching functions, prompting, "How much can you do to. . ." and having teachers respond on a nine-point Likert scale ranging from None at All to A Great Deal. The NTSES asked teachers to respond to the prompt, "How certain are you that you can. . .", with responses falling along a seven-point Likert scale ranging from Not Certain at All to Absolutely Certain. The Norwegian

External Control scale took a different approach, asking teachers the degree to which a set of statements were False or True on a 6-point Likert scale.

Within the TSES, two subscales asked about more specific areas of teaching self-efficacy:

The **Efficacy in Instructional Strategies** scale asked teachers the extent to which they could use a variety of assessment approaches and alternative teaching strategies, among other teaching tasks. The **Efficacy in Student Engagement** scale asked teachers about their ability to engage students, including the degree to which they could motivate students and help students value learning.

From the NTSES, the subscale for **Adapting Instruction to Individual Needs** was particularly relevant to the adaptive use of data, with items asking teachers their level of certainty for adapting assignments to students' needs and for organizing classrooms to adapt to both low-ability and high-ability students. The **External Control Scale** (NECS) asked teachers the degree to which a teacher or a school can impact student achievement above and beyond the influence of students' home environment or innate abilities. Table 25 summarizes the type, number of items, and Cronbach's Alpha for all subscales.

Table 25

Summary of Survey Scales

Subscale	Cronbach's Alpha
Survey of Educator Data Usage (SEDU) (Wayman et al., 2009b) (4-point Likert scale)	
Computer Data Systems (4 items)	.90
Attitudes Toward Data (4 items)	.89
Data Use Practice (5 items)	.96
Data's Effectiveness for Pedagogy (5 items)	.95
Instructional Resources (8 items)	.95
Support for Data Use (6 items)	.92
DBDM Efficacy and Anxiety (3D-MEA) (5-point Likert scale)	
Efficacy for Application of Data to Instruction (6 items)	.92
Efficacy for Data Analysis and Interpretation (3 items)	.81
Efficacy for Data Identification and Access (3 items)	.84
Efficacy for Data/Technology Use (3 items)	.91
Teacher Self-Efficacy Scale (TSES-Short Form) (9-point Likert scale)	
Efficacy in Instructional Strategies (4 items)	.86
Efficacy in Student Engagement (4 items)	.81
Norwegian Teacher Self-Efficacy Scale (7-point Likert Scale)	
Efficacy to Adapt Instruction to Individual Needs (4 items)	.87
Norwegian External Control Scale (6-point Likert Scale)	
External Control Scale (5 items)	.79

Data Sets. In addition to the survey data collected by the researcher, retrospective data from the 2015-16 school year was provided by the school, including the log file exports analyzed in Study One, classes taught, and professional development received. All data provided by the school was stripped of personally-identifiable information, with unique identifiers used for all teachers. Three main data sets were used in the current study.

Data Set 1, Online Data Usage Only, included 40 core content Teachers with active accounts in the Benchmark Data system, out of a possible total of 42. A subset of 39 teachers who used the system at any point in the semester—active users—is also used for several analyses.

Data Set 2, Survey Only Analysis, included data from 37 of the total 42 core content teachers.

Data Set 3, Online and Survey Analysis Combined, included data from the 35 core content teachers with active accounts in the Benchmark Data system and survey participation.

Missing data. Missing data were treated in various ways depending on the analyses employed. In a very small percentage of cases where one item to be included in a survey scale went unanswered, the scale score was still calculated with the average of the remaining items. If more than one item was missing, the scale was not calculated and considered missing in later analyses. For all descriptive statistics, correlation matrices, and Cronbach's Alpha calculations, pairwise deletion was employed.

Analysis. A variety of analyses were conducted in response to the research questions of Study 2:

(R2) Descriptive statistics (mean and standard deviation), correlation matrices, and Cronbach's Alpha were calculated on all survey scales and subscales regarding (R2) **"**What attitudes do teachers have towards data use?" Line charts comparing average survey responses across school level and content area were generated, as well as HCA heatmaps clustering teachers along survey dimensions. HCA heatmap methods follow those described in Study 1. Row numbers identifying users are not comparable between Study 1 and Study 2.

(R3) To investigate (R3) "To what extent are teachers' data use attitudes, technology acceptance, self-efficacy, and roles related to their online use of student testing data?" exploratory Pearson correlations were calculated between teacher survey scores (all scales and subscales) and online data usage features generated in Study 1. Additionally, HCA heatmaps were generated clustering teachers on their self-reported frequency of data use, the usefulness of data types, the usefulness of data activities, general data use attitudes, self-efficacies, PEOU, and online usage metrics.

Unfortunately, the small data set of 35 core content teachers does not allow for sufficient power to conduct two-tailed T-tests for such a large number of correlation coefficients. In a study with a larger data set, significance tests would be conducted, along with the Benjamini-Hochberg Procedure (Benjamini and Hochberg, 1995) to control for the overall rate of false discovery with multiple correlations. Considering the limitations of this school-specific study, a more exploratory approach was taken in examining correlations. Such an approach, even with its faults, still has the potential to richly describe complex relationships in context, providing useful descriptive analysis (Loeb et al., 2017) or what Bowers et al., (2017) term a quantitative phenomenology.

Results

Given the large number of factors, results of analyses are organized into three sections:

1. A school-specific data use profile by reported use and usefulness of data types
2. Relationships and patterns across data-use attitudes and efficacies
3. An exploration of the Technology Acceptance Model (TAM) for understanding teachers' online data use.

The first section focuses on creating a rich profile of school data use through survey responses to two basic questions: How often do you use specific types of data? And, how useful are they? Results are presented in the form of descriptive tables, exploratory correlations, line graphs, and HCA Heatmaps.

The second section explores relationships between attitudes and self-efficacies related to teachers' use of data. Specifically, general data use attitudes from the SEDU are related to teachers' self-efficacies for data-driven decision making (3D-MEA) and teaching self-efficacy (TSES, NTSES). Descriptive tables, exploratory correlations, and HCA Heatmaps are used to investigate patterns across these psychological factors.

The final section explores the TAM factors of perceived ease of use (PEOU) and perceived usefulness (PERUSE) as possible determinants of teachers' online use of student data. Exploratory correlations and HCA Heatmaps are used to investigate PEOU and PERUSE in relation to teachers' roles and online use.

The two basic research questions remain the same across the three sections, with relevant results generated in each:

(R2) What attitudes do teachers have towards data use?

(R3) To what extent are teachers' data use attitudes, technology acceptance, self-efficacy, and roles related to their online use of student testing data?

Population and Sample. As shown in Table 26 and Table 27, survey respondents in data set 2 represented 37 (88%) of 42 core content teachers at Progress Secondary. Data Set 3, combining active Benchmark Data Users and Survey respondents, included 35 (83%) of 42 core content teachers.

Table 26

Teacher Sample by School Level

	Core Content Teachers		Survey Respondents		Had Benchmark Data Access	
	n	%	*n*	%	*n*	%
Middle School	22	52	20	54	20	50
High School	20	48	17	46	20	50
Total	42	100	37	100	40	100

Table 27

Sample by Online Access and Survey Response

		Responded to survey		Did not respond to survey		Total
		n	%	*n*	%	*n*
Access to Benchmark Data	Yes	35	95	5	100	40
	No	2	5	0	0	2
Total		37	100	5	100	42

A School Data-Use Profile by Reported Use and Usefulness of Data Types. On the whole (see Table 28) teachers at Progress Secondary considered themselves to be either average or expert in terms of overall comfort with technology, use of the student information system, and use of the Google Suite for Education platform. Benchmark Data had the lowest average user proficiency (2.35) of the four systems included in the survey. Roughly equal numbers of users reported themselves as average users of Benchmark Data as reported themselves non-users and novices combined. No teachers reported being expert users of Benchmark Data. The lower reported proficiency for Benchmark Data is not surprising given its first-year implementation and the complexity of the system.

Table 28

Survey Responses: Technology Comfort and User Level

Survey Item	(1) Non-user (%)	(2) Novice user (%)	(3) Average user (%)	(4) Expert user (%)	Mean (SD)	n
General Tech Comfort	0	0	19 (56)	15 (44)	3.44 (0.50)	34
Level User: SIS	0	0	12 (32)	25 (68)	3.68 (0.48)	37
Level User: Google Apps	0	0	11 (30)	26 (70)	3.70 (0.46)	37
Level User: Benchmark Data	5 (14)	14 (38)	18 (49)	0	2.35 (0.72)	37
Level User: Google Classroom	5 (14)	5 (14)	16 (43)	11(30)	2.89 (1.00)	37
Level User: Avg. All Systems	0	1 (3)	24 (65)	12 (32)	3.16 (0.41)	37

As shown in Table 29, Benchmark Data also ranked lowest in perceived usefulness among teachers, below both the Student Information System and the Google Apps Platform. A combined 71% reported the program as either somewhat, very, or extremely useful, with the other 29% reporting the program as only slightly useful or not at all useful.

Table 29

Survey Responses: Perceived Usefulness of Data Systems

Data System	(1) Not at all useful (%)	(2) Slightly useful (%)	(3) Somewhat useful (%)	(4) Very useful (%)	(5) Extremely useful (%)	Mean (SD)	n
SIS	1 (3)	3 (8)	6 (16)	16 (43)	11 (30)	3.89 (1.02)	37
Google Apps	1 (3)	1 (3)	5 (14)	14 (39)	15 (42)	4.14 (0.96)	36
Benchmark Data	2 (6)	8 (24)	8 (24)	12 (35)	5 (12)	3.24 (1.13)	34
Average All Systems	1 (3)	2 (5)	9 (24)	17 (46)	8 (22)	3.79 (0.85)	37

Reported frequency of data use. Results for teachers' reported frequency of data use are summarized in Table 30, with data types ranked from the most frequently used to the least. Teachers reported using their own assessments most frequently, with 65% reporting weekly use or use a few times a week. Discipline records and IEP information followed, with 57% and 46% reporting weekly or greater use. Surprisingly, the reported frequency of use for formal assessments, administered 3-4 times a year, was only slightly higher than that of State tests, administered only once a year. However, this difference in frequency may be underestimated based on the limited number of response options. Since interim testing, administered every nine weeks, and State tests, administered annually, are both offered "less than once a month," the same response was likely selected for both. The lack of a frequency response option for use occurring "Once or Twice a Quarter" (or similar) may collapse the difference between these two sources of data.

Table 30

Survey Responses: Reported Frequency of Use for Data Types

Type of Student Data	(1) Less than once a month (%)	(2) Once or twice a month (%)	(3) Weekly or almost weekly (%)	(4) A few times a week (%)	Mean (SD)	n
Own Assessments	2 (5)	10 (27)	16 (43)	8 (22)	2.83 (0.85)	36
Discipline Records	12 (32)	3 (8)	9 (24)	12 (32)	2.58 (1.27)	36
IEP Goals	8 (22)	12 (32)	6 (16)	11 (30)	2.54 (1.15)	37
Software Data	18 (49)	5 (14)	10 (27)	3 (8)	1.94 (1.07)	36
Formal Assessments	20 (54)	13 (35)	3 (8)	1 (3)	1.59 (0.76)	37
State Testing	26 (70)	7 (19)	4 (11)	0	1.41 (0.69)	37

Along with these summary tables, two types of figures are used to explore results related to teachers' reported use and usefulness of data types. First, simple line plots are used to compare group means between middle and high school and content areas. Second, HCA heatmaps are used as a form of visual data analytics (Bienkowski et al., 2012) to facilitate pattern recognition across a disaggregated view of user behaviors, factors, and outcomes. In each case, the HCA heatmap visualization builds on the simple line graph by providing a more fine-grained view of possible subgroups of teacher attitudes in relation to teacher roles and usage. While the underlying data of these paired analyses is the same, they provide complementary descriptions, one focusing on overall trends in school and content area and the other facilitating the search for patterns across individual users.

Line graphs of reported frequency of data use. The line graphs below compare the average responses for school level and content area, visualizing patterns in how teachers report their frequency of usage. Each graph includes a dotted line indicating the combined average responses, along with separate lines representing teacher groups. Line graphs representing middle school include 20 participants. High school graphs represent the responses of 17 participants, and graphs combining middle and high school represent 37 participants total. Data types are listed along the *x*-axis, from the highest reported frequency on the left to the lowest frequency on the right. Several interesting group-level differences in reported frequency are suggested by these profiles of school and content-area.

In comparisons between middle and high school teachers (Figure 26), for example, reported frequency is similar for teachers' own assessments, IEP information, formal assessments, and State testing data. Since high school teachers in Study 1 demonstrated higher average use of testing data than middle school teachers, it is interesting that their perceived frequency of use in these key testing categories is still similar to that of the middle school. Some of these similarities could again be due to lack of sensitivity in the survey, where "Less than once a month" is the lowest frequency response option.

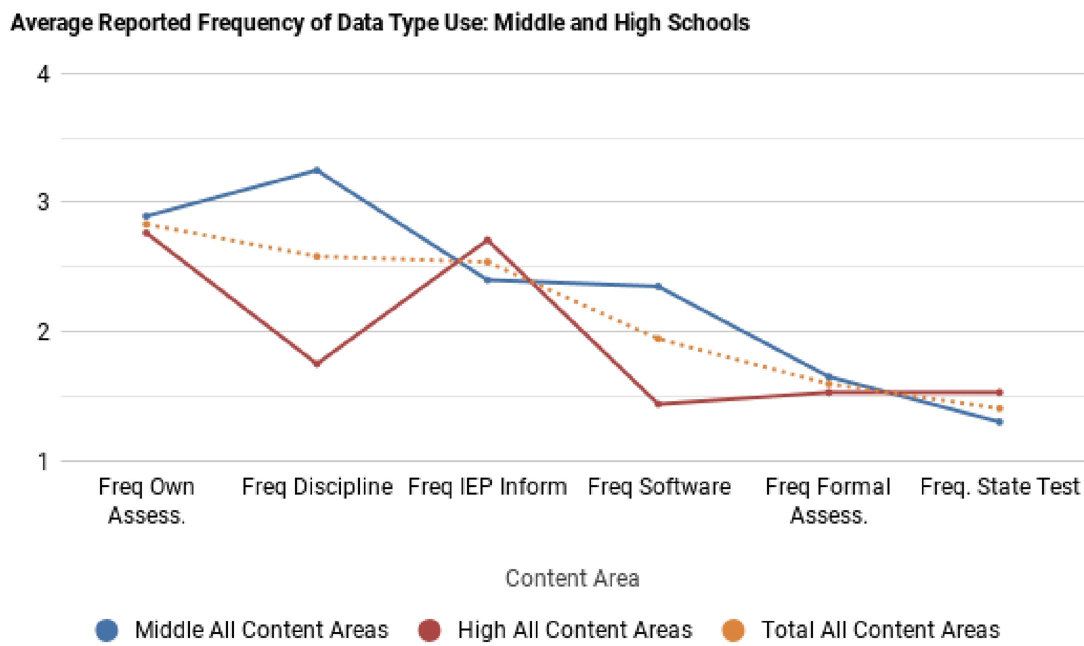


Figure 26. Average reported frequency of use by data type: middle and high schools

On the other hand, Figure 26 indicates different reported frequencies for Discipline Data and Data from Instructional Software. While there are many possible reasons for this difference, from fewer discipline incidents in high school than middle school to the use of different instructional software between schools, the differing patterns may also indicate different schoolwide approaches to the use of information. Middle school teachers, for example, report accessing a wider range of information more frequently: their own assessments, discipline data, IEP information, and software data are all consulted on average more than once or twice a month. The high school, in contrast, reports using only their own assessments and IEP information at the same high rate, with all other sources of information used less than once a month.

This portrait of data use is very different from the one painted when considering only online usage of the Benchmark Data system, a system in which the high school demonstrates higher rates of access. Given that the sources of information that the middle school reports using most frequently were not accessible in Benchmark Data, this survey profile of reported data use in middle and high school offers a complementary view and a reminder of the limitations of considering online usage on its own, without factoring in other possible sources of student data.

As compared to middle/high school differences, the profiles of reported use by content area (Figure 27) appear remarkably similar, with the exception of ELA teachers' reported frequency of use. While Math, Science, and Social Studies show similar trajectories of reported use, ELA teachers reported higher levels of usage for IEP information and lower levels of usage for data for instructional software. The larger number of IEP goals related to literacy and language needs, as opposed to specific content area knowledge, may provide one possible explanation for the higher reported usage of IEP information among ELA teachers, since the monitoring and reporting of literacy goals may require more extensive collaboration from a student's ELA teacher.

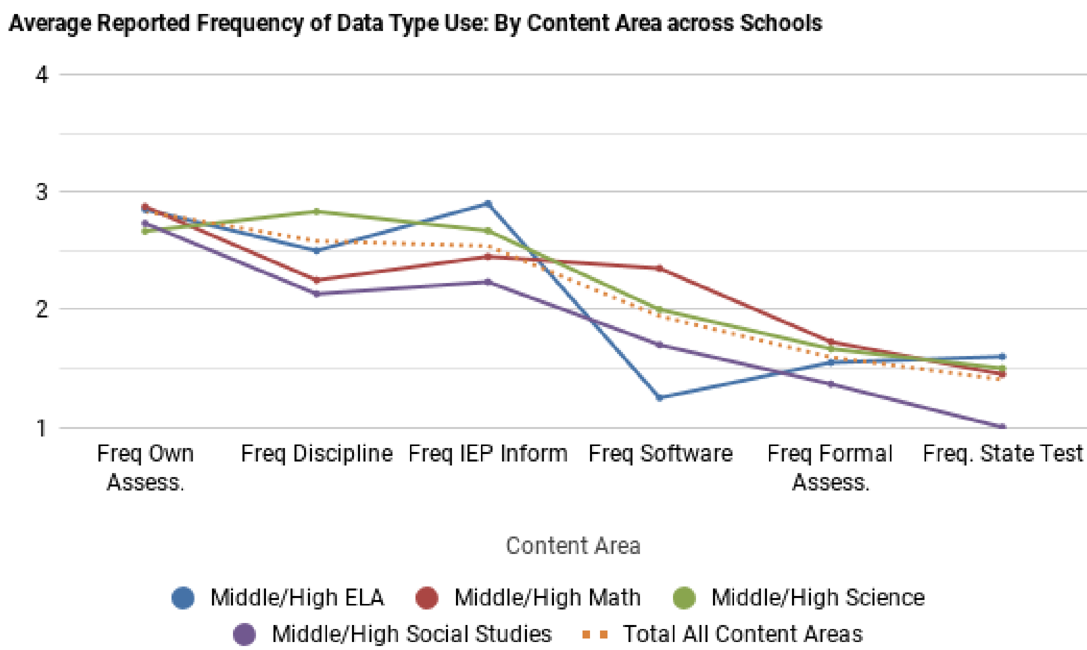


Figure 27. Average reported frequency of data type use: by content area across schools

Viewing reported frequency of use divided by school and content area (Figure 28 and Figure 29) it becomes clear that the differences in ELA access to IEP information and software information are largely driven by high school responses, which follow a more varied pattern of reported use than middle school.

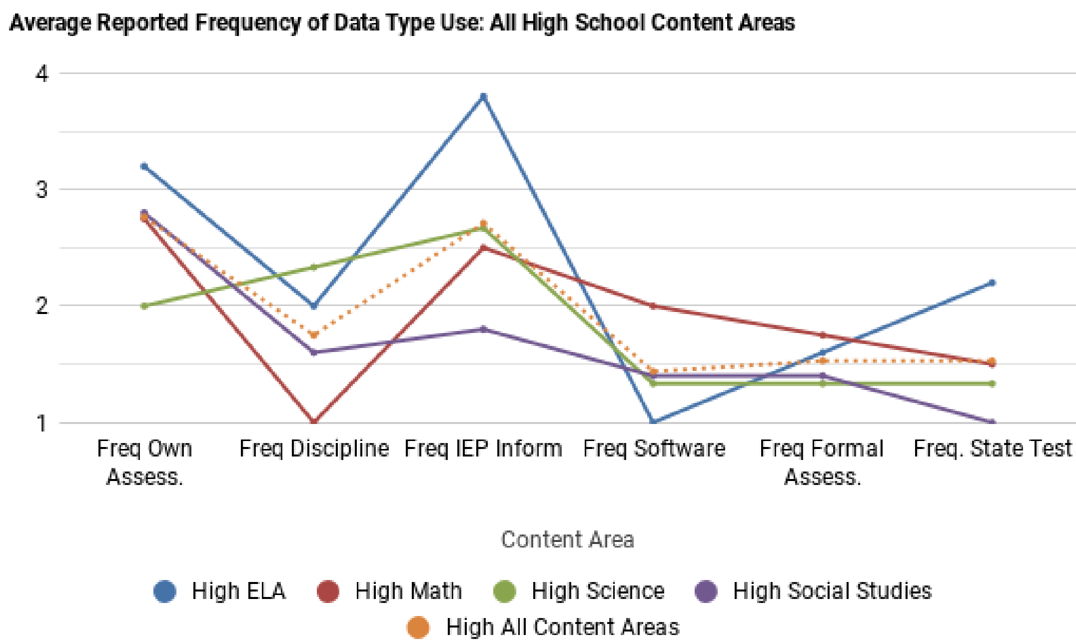


Figure 28. Average reported frequency of data type use: all high school content areas

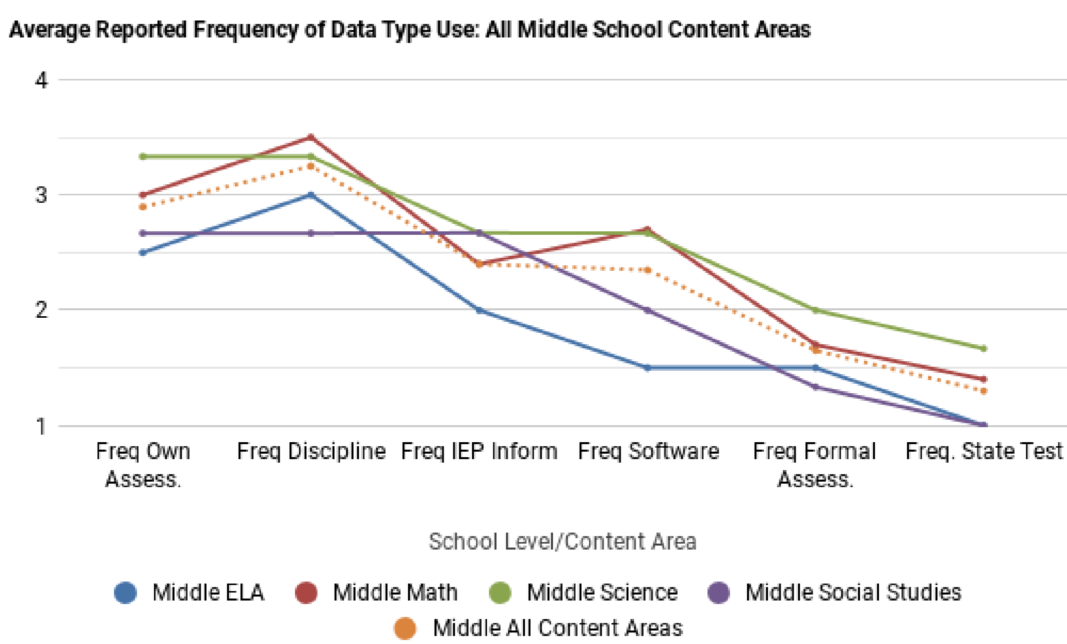


Figure 29. Average reported frequency of data type use: all middle school content areas

Figure 28 and Figure 29 compare content areas across middle and high School. While middle school content areas follow basically similar profiles of reported use, high school reported use varies quite a bit from content area to content area. In ELA, for example, there are striking differences in the reported frequency of access to IEP information and access to state test data. While it is not immediately clear why the two schools should differ so much regarding use of IEP information, one reason may be that the semester-based grading periods of high school and twice-yearly state testing require more frequent access to IEP information in order to verify modified grading criteria or testing accommodations.

While additional profile plots by content area are included under Appendix B, an interesting observation comes with examining Figure 30, comparing ELA teachers across middle and high school. The lower values for software use in both schools raise the possibility that ELA teachers are not viewing the use of Google Docs for student writing as a form of data from Instructional Software. While there is no clear reason why teachers should have conceived of Google Docs as “instructional software,” they do access large amounts of student writing through the system. Alternatively, ELA teachers may not have felt that student writing itself should be considered data. In either case, the lack of inclusion for this large category of student work/instructional information speaks to the constraints that teachers may feel as to what constitutes “data.”

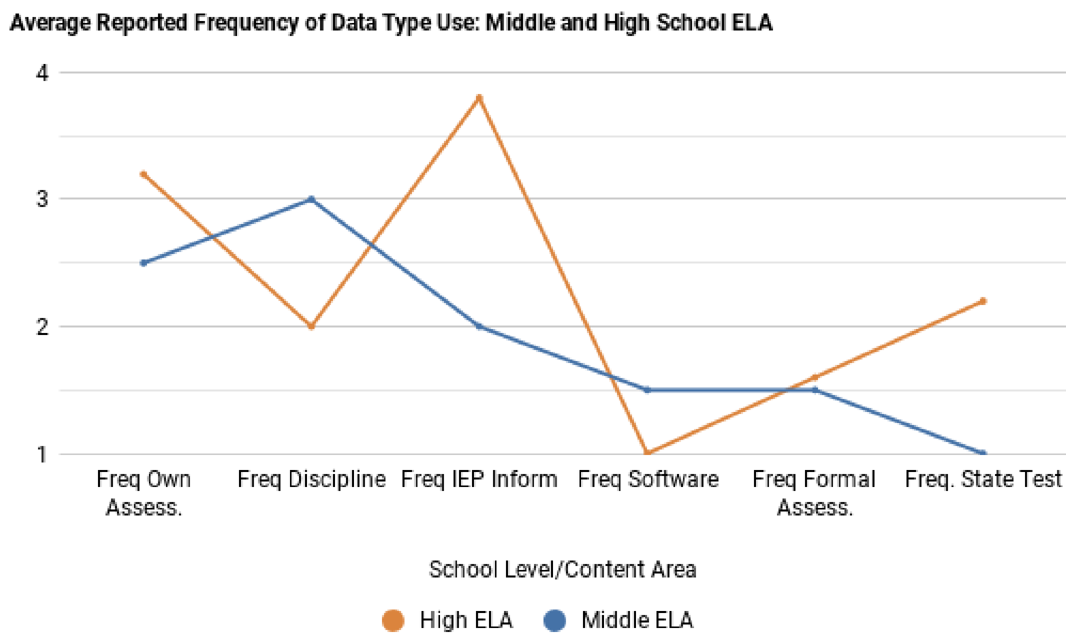


Figure 30. Average reported frequency of data type use: middle and high school ELA

The next section of analyses applies the HCA heatmap methodology discussed in Study 1 to the reported frequency of data use. Some of the same patterns emerge as in the previous profile plots, along with other possible subgroups of teachers cutting across school level and content area.

HCA heatmap: reported frequency of use by data type. Figure 31 clusters teachers in rows and data types in columns according to their reported frequency of use. Dendrograms on the top and left describe the hierarchical clustering, while annotations to the right of the heatmap indicate school level, above average duration of use, and content area.

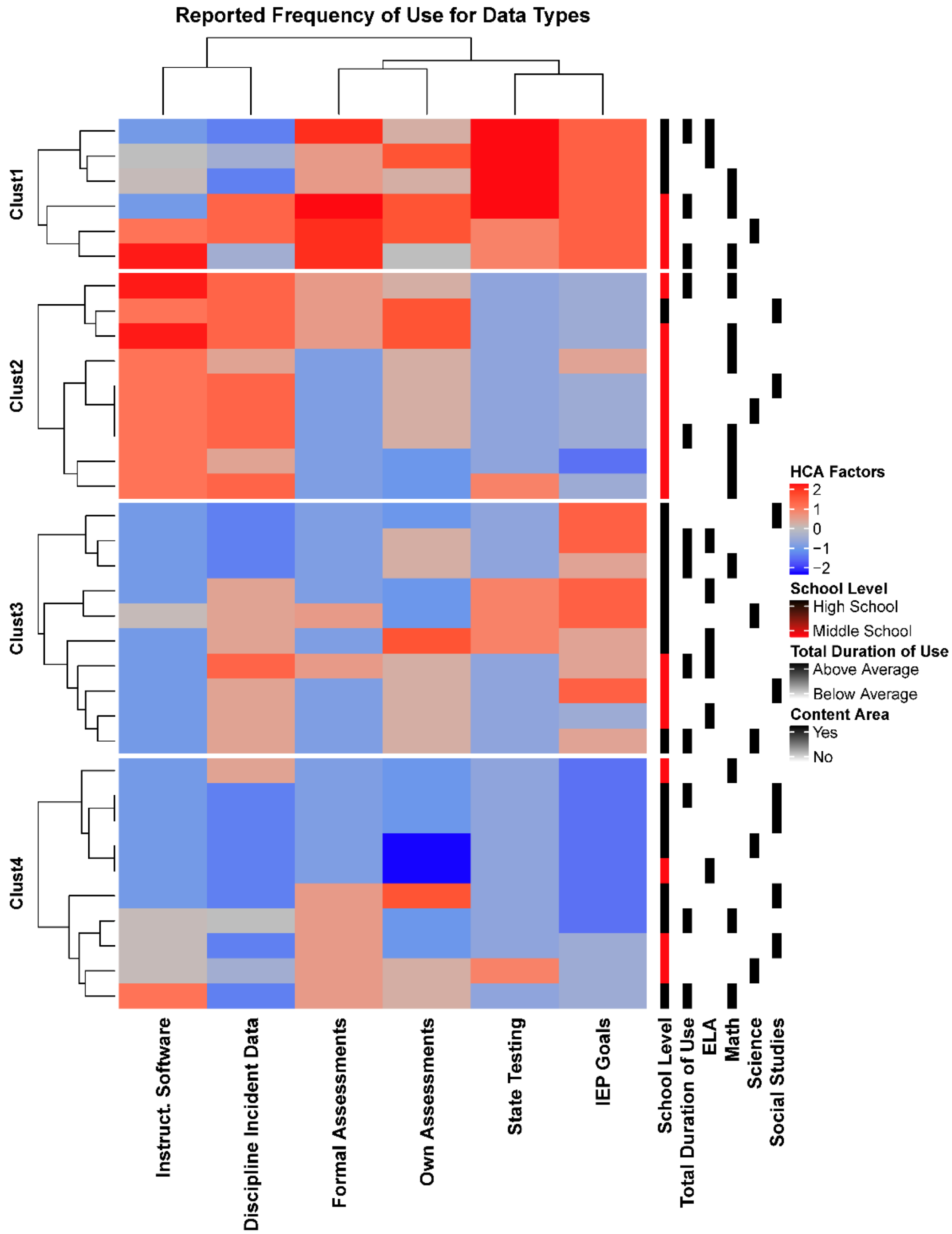


Figure 31. HCA heatmap: reported frequency of use for data types. In NbClust, 8 methods proposed 2 clusters for reported frequency of use factors, while 6 methods each proposed 3 and 4 clusters. Visual Inspection suggested 4 clusters. See Appendix C for all heatmap dendrograms with distance scales and defined clusters.

The four clusters identified in Figure 31 include: users with high reported frequency of use for most types of data (Clust1), higher use of instructional software and discipline incident data (Clust2), higher than average frequency of access to IEP goals (Clust3), and lower overall frequencies of reported access (Clust4). In the cases of Clust2 and Clust3, heatmap clustering identifies similar patterns as the profile plots above, where middle school teachers report more frequent access to instructional and discipline data and make up the majority of Clust2. Similarly, high school teachers make up the majority of Clust3, a cluster characterized by higher access to IEP Goals.

The final Clust4 reported generally much lower rates of use across the board, even regarding their own assessments. Users reporting high frequency access of multiple data types (Clust1) are distributed across middle and high school but consist almost entirely of Math and ELA content teachers. Consistent with Study 1, Social Studies teachers appear concentrated in Clust4, reporting the lowest frequency of data use, even regarding their own assessments.

An interesting contrast to how often teachers report using different *types of data* is how often they report using data to complete *different instructional tasks*. The frequencies of data-based instructional actions (as opposed to frequency of access to data type) are visualized in Figure 32, by viewing responses to individual items of the Instructional Resources Scale of the SEDU. Items in this scale include data-use activities such as identifying students' learning needs/goals, tailoring instruction, recommending tutoring, assigning students to classes and small groups, and identifying which content to teach.

HCA heatmap: reported frequency of data use functions. Figure 32 clusters participants in rows and items from the Instructional Resources subscale of the SEDU in columns.

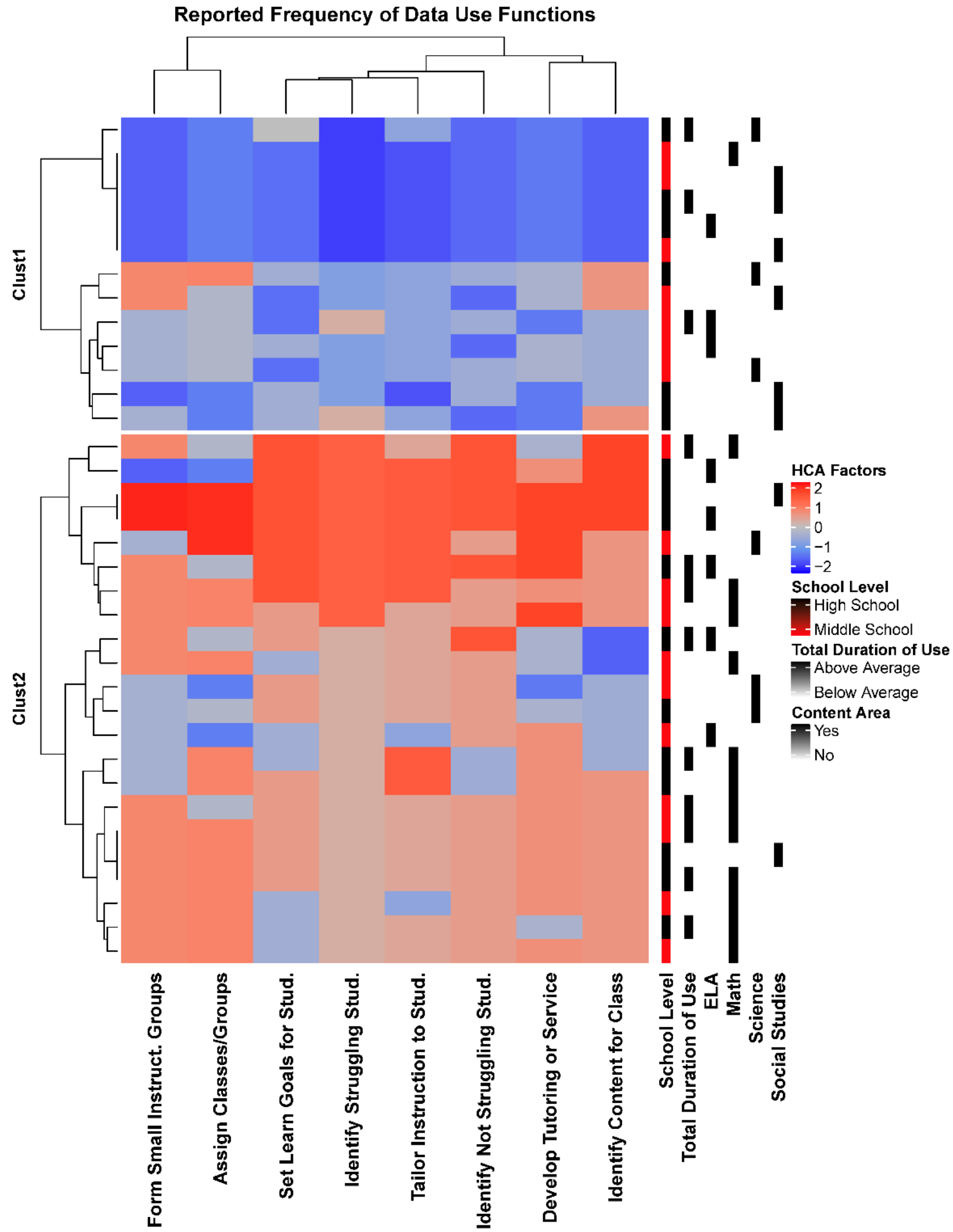


Figure 32. HCA heatmap: reported frequency of data use functions. 13 methods from the NbClust package propose 2 clusters, with 5 methods proposing 4 clusters. Visual inspection suggests 2 clusters.

In contrast to Figure 31, where only a small cluster of teachers indicated that they use multiple data sources more frequently, Figure 32 shows the majority of teachers indicating higher frequency of data use actions, such as forming small groups for instruction and identifying struggling students. This contrast merits further investigation: more than half of teachers indicate they are frequently applying important data-based strategies in the classroom, but they do not report drawing from many of the possible sources of student performance data to implement those strategies.

Above the cluster of teachers reporting frequent classroom use of data strategies (Clust2), a sizeable group report lower frequencies of data use actions (Clust1). In both Clust1 and Clust2, however, teachers report their frequency of data use fairly consistently across categories of action; teachers respond either more highly for almost all data use actions, or they respond more negatively for actions. While these actions could hypothetically be completed by one teacher at very different frequencies, perhaps forming small groups more frequently than identifying content for class, teachers' responses indicate similar frequencies of actions across the group.

Also interesting, the two large clusters seem to avoid concentrations of middle or high school teachers, but do appear related to online usage, with teachers demonstrating above average online usage clustering in the high frequency group for data use actions. In terms of content area, Science and Social Studies cluster in the low frequency group (Clust1). This pattern holds despite Science teachers reporting more frequent access to data types than do Social Studies teachers. Math teachers and some ELA teachers tend to dominate the high frequency group (Clust2) for data use actions.

Perceived usefulness of data types. Table 31 summarizes teacher responses about how useful several types of data are for understanding students and targeting instruction. The perceived usefulness of data types is one way in which this study operationalizes the TAM construct of PERUSE (Davis, 1989). PERUSE will also be considered later in these results in relation to the data system itself. The percentages in the table are reported as a percent of the total who responded to each item and who were aware of the type of data in question. Data types are listed in the table from highest perceived usefulness to least.

As with reported frequency of use, teachers' own assessments, as well as their own observations and gradebook were ranked highest, followed by IEP, Instructional Software, and Cognitive Skills data. The high level of usefulness for software-related data is interesting given how outside of teachers' control such systems often are, with metrics generated from students' unseen interaction with the software system. Various types of standardized testing were ranked lower for usefulness, with interim testing ranked the lowest. Students' Past Grades, though generated by known colleagues and fellow teachers, ranked only slightly higher than state testing for usefulness. Also of note was that the usefulness of discipline records fell below that of state testing. This ranking is surprising given the high frequency with which teachers reported using these records. Such a mismatch between usefulness and use may indicate a more bureaucratic than adaptive use for discipline data; in other words, teachers may be required to use discipline data frequently for some reporting or reward process but may not perceive that process as useful to their instruction. It is also intriguing that state standards and Lexile levels, despite the complexity of the data they represent, were considered some of the more useful types of data. Two other complex metrics—predictions of students' risk of dropping out of high school and predictions of students' future state scores—fared less well, with teachers indicating a lower usefulness for these data types.

Table 31

Survey Responses: Perceived Usefulness of Data Types

	(1) Not at All Useful (%)	(2) Slightly Useful (%)	(3) Somewhat Useful (%)	(4) Very Useful (%)	(5) Extremely Useful (%)	Mean (SD)	n
Own Observations	0	1 (3)	0	13 (35)	23 (62)	4.57 (0.65)	37
Own Assessments	0	2 (5)	0	14 (38)	21 (57)	4.46 (0.77)	37
Own Gradebook	0	1 (3)	1 (3)	15 (42)	19 (53)	4.44 (0.70)	36
IEP	1 (3)	2 (5)	4 (11)	15 (41)	15 (41)	4.11 (0.96)	37
Software	1 (3)	1 (3)	12 (34)	15 (43)	6 (17)	3.69 (0.90)	35
Cognitive Skills	1 (3)	4 (12)	6 (18)	17 (50)	6 (18)	3.68 (1.00)	34
State Standards	1 (3)	4 (11)	13 (35)	14 (38)	5 (14)	3.49 (0.96)	37
Lexile	2 (6)	6 (18)	8 (24)	11 (32)	7 (21)	3.44 (1.19)	34
Attendance	1 (3)	7 (19)	13 (35)	9 (24)	7 (19)	3.38 (1.09)	37
Past Grades	1 (3)	6 (16)	14 (38)	12 (32)	4 (11)	3.32 (0.97)	37
State Testing	2 (6)	10 (30)	7 (21)	9 (27)	5 (15)	3.15 (1.20)	33
Discipline	2 (6)	9 (27)	11 (33)	7 (21)	4 (12)	3.06 (1.12)	33
Regents Testing	5 (14)	7 (20)	10 (29)	8 (23)	5 (14)	3.03 (1.27)	35
Drop-out Predict	5 (14)	8 (23)	8 (23)	9 (26)	5 (14)	3.03 (1.29)	35
Progress Quizzes	4 (11)	8 (23)	12 (34)	8 (23)	3 (9)	2.94 (1.14)	35
Comp. Adapt. Test	5 (14)	9 (26)	11 (31)	6 (17)	4 (11)	2.86 (1.22)	35
State Test Predict	6 (17)	8 (23)	14 (40)	4 (11)	3 (9)	2.71 (1.15)	35
Interim Testing	7 (19)	10 (27)	12 (32)	6 (16)	2 (5)	2.62 (1.13)	37

In the same way that Figure 26 through Figure 30 visualize results for reported frequency of use, Figure 33 through Figure 36 use line graphs to create average response profiles by school level and content area. In contrast to the large differences between middle and high schools in reported frequency of use, remarkably similar profiles are reported by the two schools for the perceived usefulness of data types. In other words, reported frequency of use tends to differ more across school level, while reported usefulness of data types differs more across content areas. Data types are arranged on the *x*-axis in the plots below from highest reported usefulness to lowest. If there is a difference between middle and high school, it would appear to be in their perceptions of formalized and standardized testing, with high school teachers perceiving these tests and quizzes to be slightly more useful than do middle school teachers. However, given the small sample size the importance of this difference may be negligible.

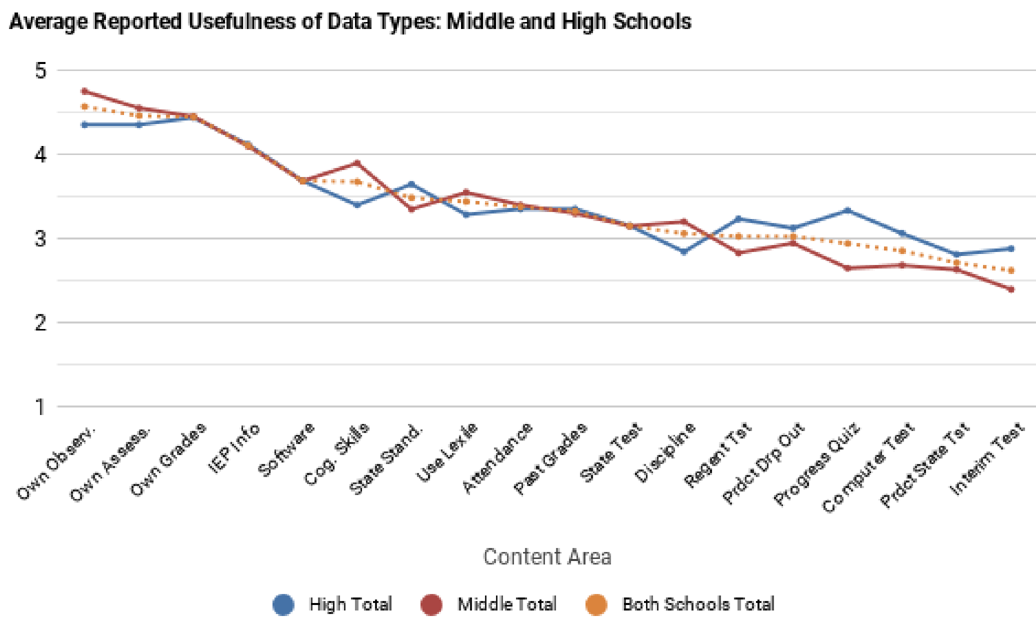


Figure 33. Average perceived usefulness of data types: middle and high schools

Unlike the school-based view in Figure 33, the content area view in Figure 34 indicates greater differences for the perceived usefulness of data types. Social Studies teachers, for example, in keeping with their lower reported frequency of use and lower observed usage in the Benchmark Data system, reported lower levels of usefulness for almost all data types. With this persistent negative trend, it seems extremely important to better understand the relationship of

Social Studies teachers to the instructional use of information, as well as the additional sources of information used by Social Studies teachers in making instructional decisions. Science teachers, in contrast, indicated higher usefulness for formalized and standardized testing.

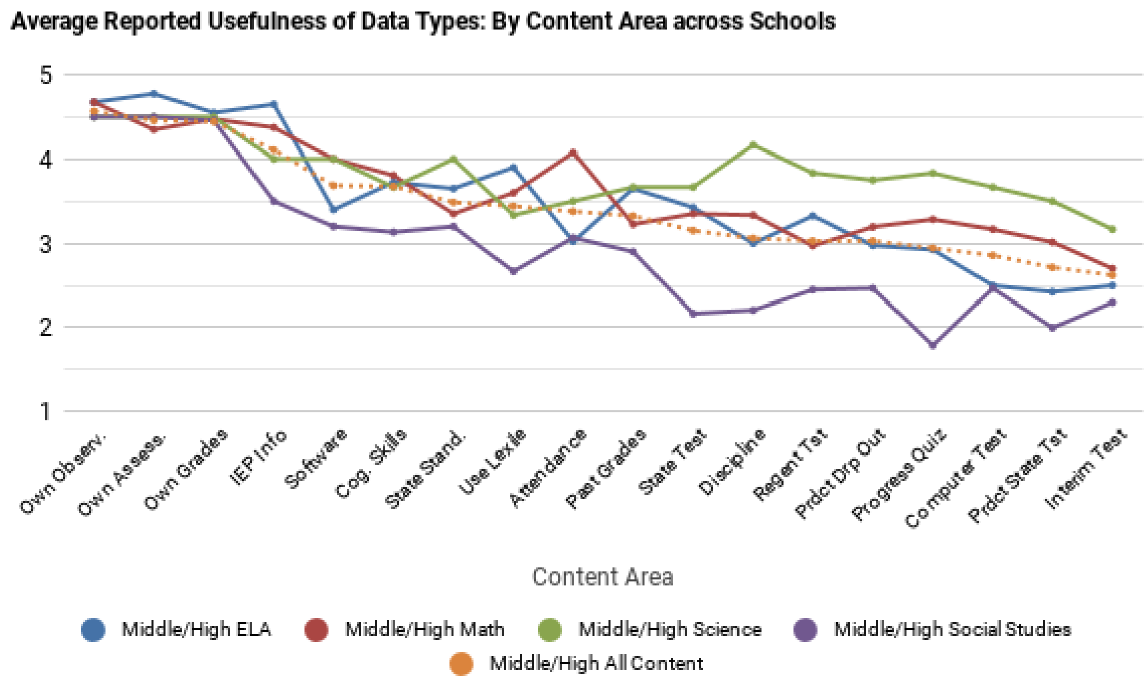


Figure 34. Average perceived usefulness of data types: by content area across schools

When comparing profiles of perceived usefulness in the middle and high school, the middle school profile, as with reported frequency of data use, appears more unified across content areas, while the patterns of usefulness reported across high school content areas are much more varied. That said, middle school Science teachers do appear to have a distinct profile for usefulness of data types (Figure 35), one which indicates higher usefulness for standardized and formal testing than the profiles for other middle school content teams. High school content teams, in contrast, show several strong differences in perceived usefulness (Figure 36): the usefulness of Lexile Levels, for example, is perceived quite differently between Science and

Social Studies on the one hand and ELA and Math on the other, with Math and ELA reporting much higher usefulness of Lexile data. This is an interesting contrast, given that the purpose of Lexile data is to help match texts to students at an appropriate level of difficulty. With the nonfiction reading demands of Science and Social Studies classrooms, Lexile Level would seem to have strong potential for matching classroom readings to students' reading levels. On the other hand, if Science and Social Studies classes are constrained in their choice of texts or have limited ability to leverage Lexile data in instructional decisions, a potentially useful metric like Lexile level may not be considered particularly useful given constraints on decision making.

Teachers' perception of IEP Data in the high school is another area where the variation in reported usefulness is surprising. At first glance, the individualized data of an IEP might seem equally useful across content areas, yet high school ELA and Math teachers record it as more useful than do Science, and especially Social Studies teachers. Another interesting difference is the degree to which high school Math teachers found attendance data more useful than did teachers in other content areas. While difficult to know the reasons for such a preference, it is easy to imagine a scenario where systematic learning progressions in Math demand a higher awareness of student attendance. Alternatively, higher usefulness of attendance data might indicate a more policy-based pattern, as with the minimum attendance policy in effect at Progress Secondary as a course passing requirement.

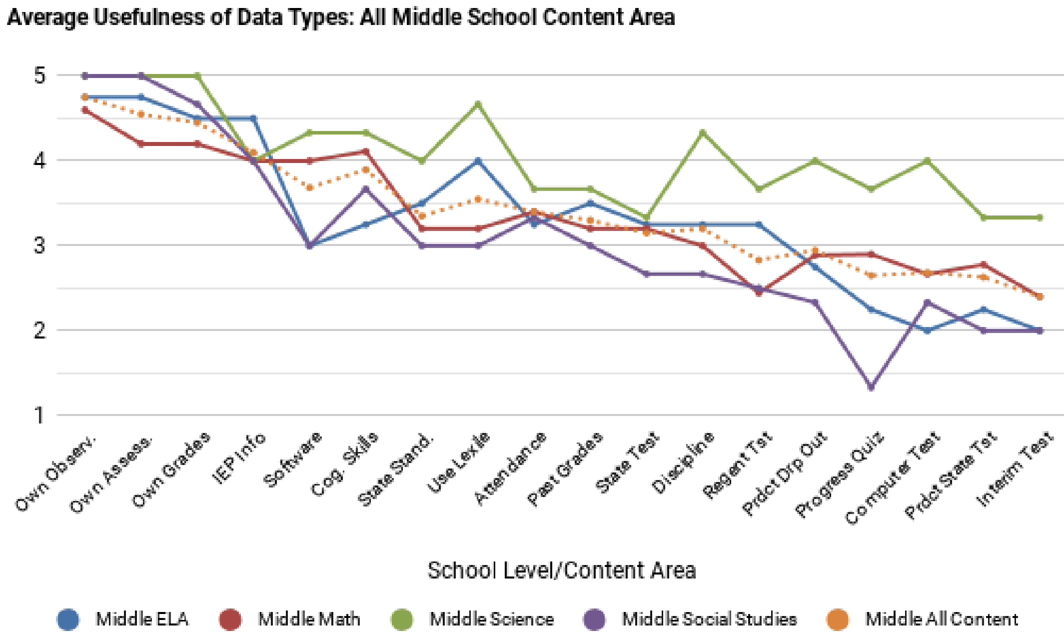


Figure 35. Average perceived usefulness of data types: all middle school content areas

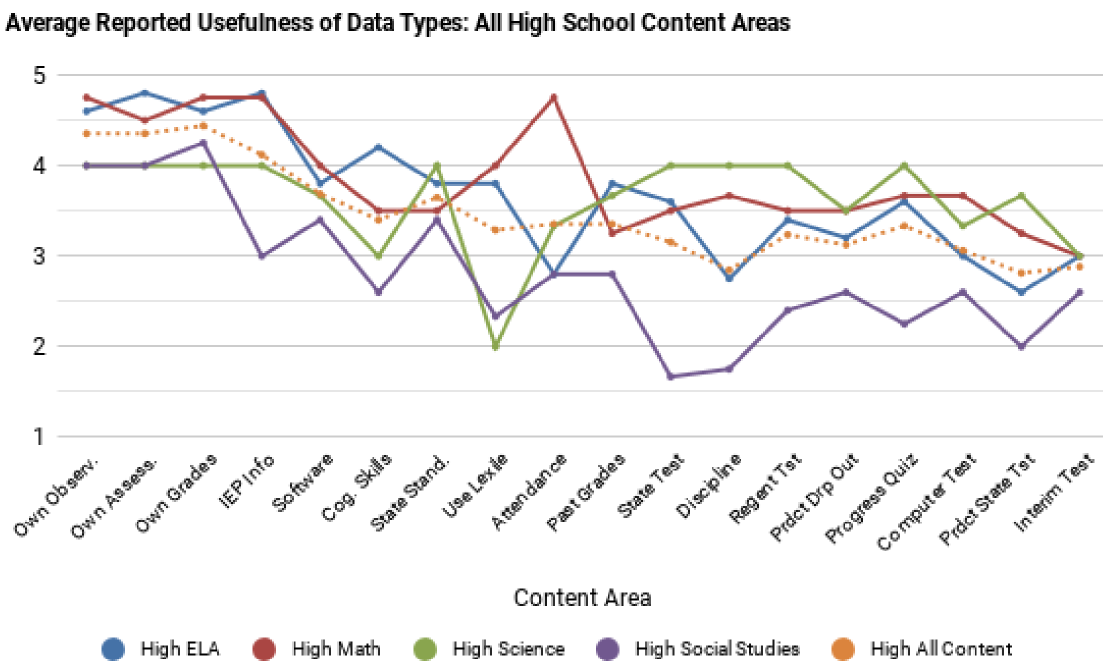


Figure 36. Average perceived usefulness of data types: all high school content areas

Figure 54, Figure 55, Figure 56, and Figure 57, included in Appendix B, compare each content areas' reported usefulness of data types between the middle and high school. Generally, Math and ELA present similar patterns across the two school levels with higher usefulness reported in high school for data from both standardized and formal testing. Profiles for Science and Social Studies, on the other hand, demonstrate larger differences between how middle and high school teachers perceive the usefulness of data types.

Overall, profile plot visualizations of teachers' reported frequency of use and of their perceived usefulness of data types suggest multiple differences between teachers based on both school level and content area. School-level differences include greater middle school access to discipline and software data. Content area differences include greater usefulness of formal testing for Science teachers, greater usefulness of attendance data in Math, and lower usefulness for many data types for Social Studies teachers. From an applied perspective, visualizations such as these offer school administrators actionable data for departmental or school-level inquiry. Schoolwide trends are easily identified in these line graphs, such as teachers' preference for either their own data (own assessments, observations, and grades) or for data tailored to individual students in complex profiles, such as IEP information, cognitive skills, instructional software, and State standards. These overarching trends in data preference have planning ramifications for how the data from formalized assessments might be better structured and presented to provide use to teachers. Alternatively, the low-ranked usefulness of formalized and state testing should call for more honest, internal inquiry into whether and how such sources of instructional data can be made more useful.

Along with schoolwide trends, these visualizations can pinpoint not only specific data types, such as Lexile level and IEP information, which stand out as underutilized, but also identify which instructional teams are best positioned to share their uses of data with colleagues. ELA teams, for example, are quickly identified through the profile plot as users of Lexile data, who might be interested in encouraging less-enthusiastic Science and Social Studies teachers to use that data.

These findings clearly suggest that a teacher's role, at the intersection of school level and content area, appears to impact the perceived value and use of many specific types of data. What is less clear from these comparisons is how each of these differences in frequency and perceived usefulness arise. Something as basic as how the statewide testing structure differs between middle and high school or a department's decision to implement an instructional software program can have ramifications for data usefulness and frequency that overwhelm any individual attitudes that teachers may have regarding data use. Additionally, the small-scale nature of this study makes it impossible to speak to any trends in data use between middle and high schools or content areas more generally.

At the same time, however, these varied data profiles for content areas speak to content knowledge, and more specifically pedagogical content knowledge (Shulman, 1986), as a possible determinant of data use attitude and use. Pedagogical content knowledge no doubt demands specific varieties of instructional feedback in the form of pedagogically-relevant and content-specific forms of student data. Generalized practice and training in assessment literacy and in formal and standardized testing may support but ultimately have difficulty capturing such content-specific demands for information. While multiple subject areas can be assessed through formalized multiple-choice tests to increase assessment reliability, arranging assessment tasks that validly inform the pedagogy and disciplinary skills specific to each content area is a more challenging task, a task which may have been under-explored in an accountability-based framework for data-based decision making.

The next HCA heatmaps in Figure 37 and Figure 38 cluster user responses based on the perceived usefulness of data types and data tasks. While these heatmaps represent the same underlying data as the line graphs above, their disaggregated view of participants allows for a more nuanced search for patterns between each of these PERUSE factors, teachers' roles, and observed use of the Benchmark Data System.

HCA heatmap: perceived usefulness of data types. Figure 37 clusters participants in rows and data types in columns, according to survey responses evaluating their perceived usefulness of data types.

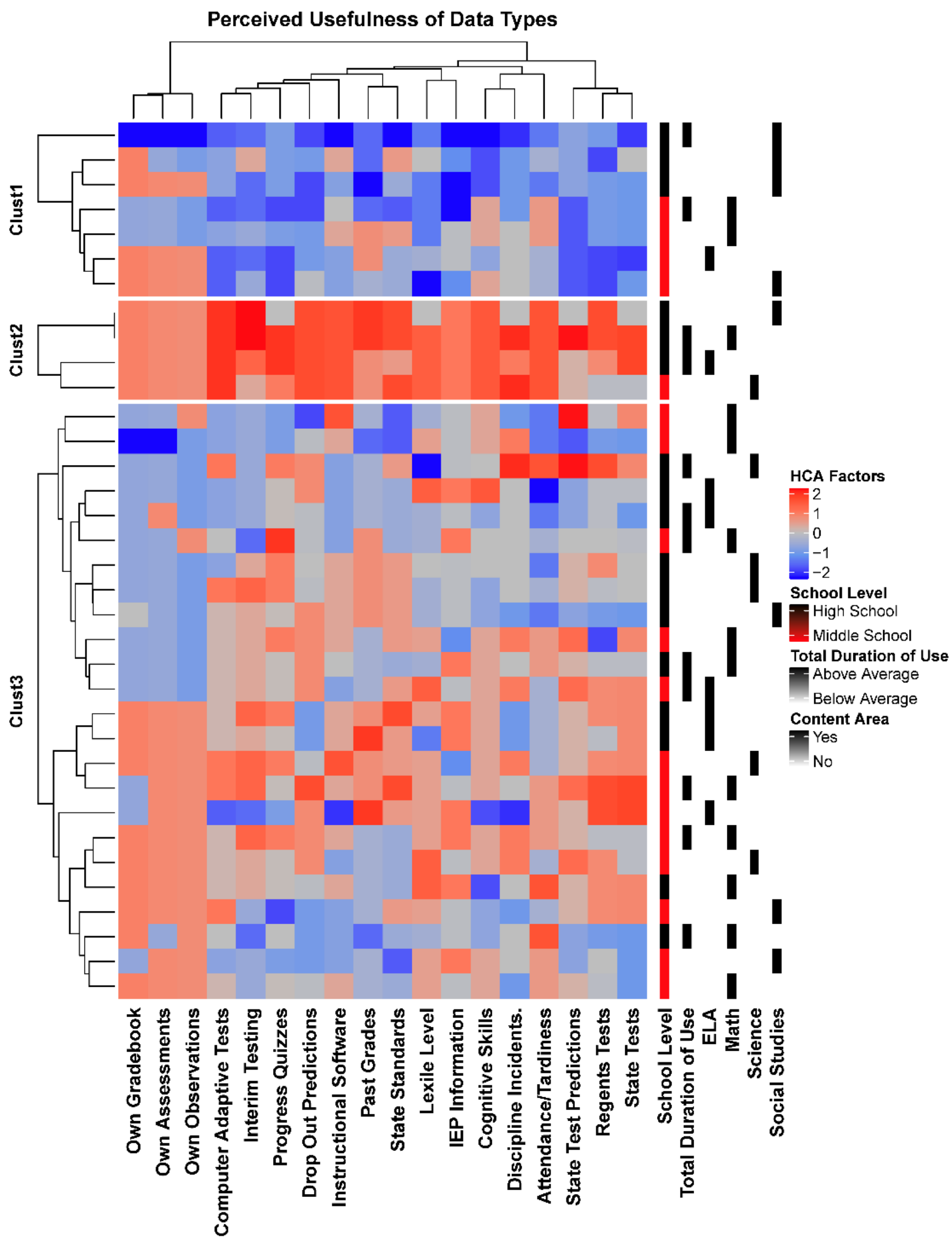


Figure 37. HCA heatmap: perceived usefulness of data types. NbClust was unable to generate results, suggesting that some factors may be too highly correlated. Visual inspection of the dendrogram, however, suggests three or possibly five clusters for interpretation.

Figure 37 groups teacher responses into three clusters: Clust1 reporting lower usefulness of data types, a small Clust2 reporting consistently higher usefulness of data types, and a larger Clust3 reporting mixed impressions of usefulness. Two subgroups of Clust3 appear defined by the greater or lesser perceived usefulness of teachers' own assessments, observations, and gradebook. On the other hand, in absolute, as opposed to standardized, terms, the range of responses for survey items on the usefulness of teachers' own materials was highly constrained to the top two options, with almost all teachers selecting either Very or Extremely Useful for these types of data. That said, a small minority of teachers did rank these data types as relatively less useful, and that pattern may merit further inquiry, if only to support teachers in making the most of these critical sources of student information.

The annotations of Figure 37 fail to indicate strong patterns regarding online use, school, or content area, except in the case of high-usefulness Clust2, which consists largely of high school teachers. While according to the Technology Acceptance Model (TAM), perceived usefulness might be predicted to have a stronger relationship with data system use, it is important to remember that in this section many of the data sources listed in the survey were not available within the data system, making the relationship between these survey items and observed usage much more tenuous. In the final section of the results, in a more specific analysis, the PERUSE of data types will be filtered to include only those available within the Benchmark system.

HCA heatmap: perceived usefulness of data activities. Another perspective on the usefulness of data was provided by the items of the Data's Effectiveness for Pedagogy subscale of the SEDU. Including these items separately, as columns in an HCA heatmap (Figure 38), allows for examination of patterns in the perceived usefulness of data use activities, as opposed to the usefulness of data types. Some items in this scale asked about the general effectiveness of data—whether students benefit from data use (Students Benefit), whether data use benefits the teacher's students personally (My Students Learn More), and whether students benefit schoolwide from data use (School Students Learn More). Other items asked about the effectiveness of particular types of data, such as state assessments (State Test Valid), formal assessment (Formal Test Valid), or local assessments (Teacher Made Valid). More action-oriented items asked about the effectiveness of data for particular tasks, such as planning instruction (Plan instruction), gaining new information about students (Learn New Information), knowing the concepts that students are learning (Identify Learned Concepts), and identifying learning goals (Set Learning Goals).

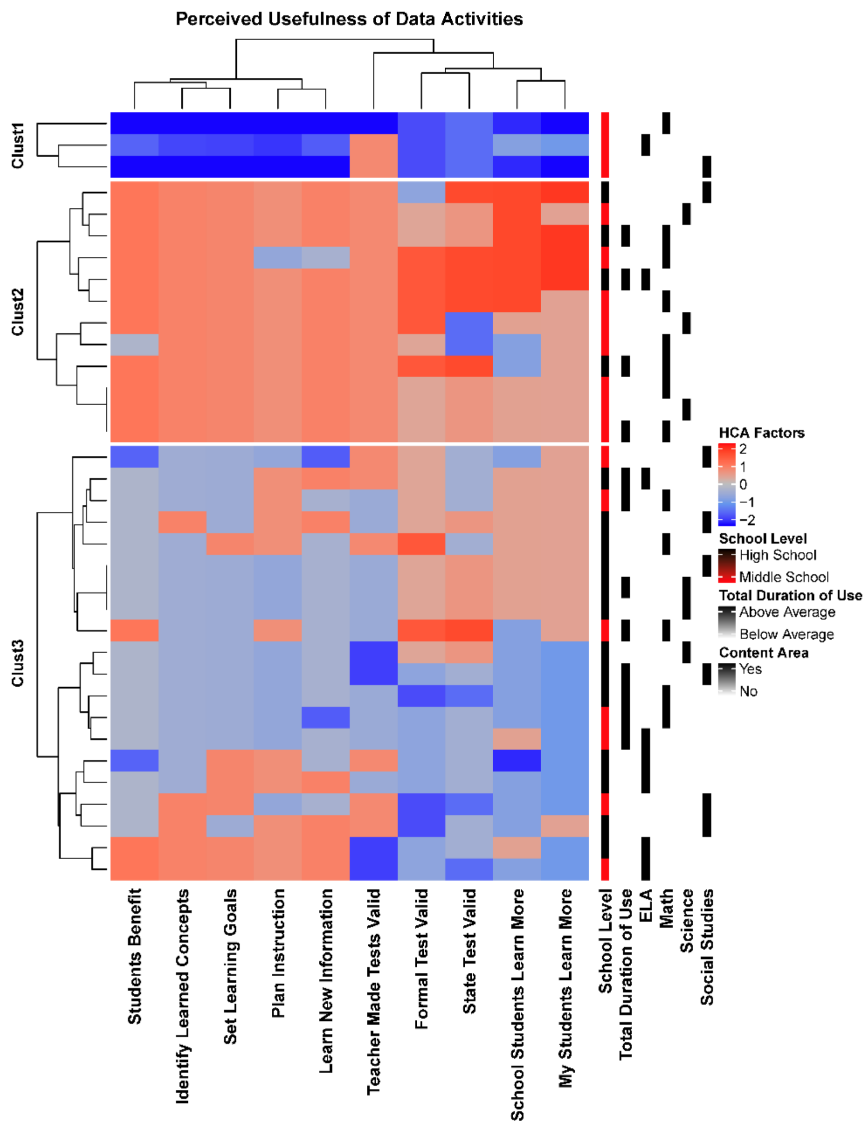


Figure 38. HCA heatmap: perceived usefulness of data activities (SEDU). NbClust proposed three as the best number of clusters. Visual inspection confirmed three groups.

In Figure 38, three main clusters are visible based on teachers' perceptions of the effectiveness of data use activities. In the small topmost Clust1, teachers consistently ranked data activities low for usefulness. The somewhat larger Clust2 consistently reported higher usefulness for data activities, and the largest Clust3 appears divided between responses that preferred state and formal assessments over classroom data use and those that preferred the reverse, classroom use over state and formal assessments. Relationships to school level and content area are also apparent. Math teachers, for example, appear to cluster in either high usefulness Clust2 or in a subcluster of Clust3 which finds formalized testing more useful. A cluster of ELA and Social Studies teachers, in contrast, appears in another subcluster of Clust3 that prefers classroom data use to formalized testing.

Two easily interpretable column clusters appear as well in Figure 38. The cluster to the left includes mainly specific classroom strategies, while the cluster to the right includes items related to formalized testing and the school-specific effectiveness of data use.

Overall, as with reported frequency, data activities are considered more useful than data types. While more teachers are skeptical about the available types of data, many fewer teachers are skeptical about the instructional activities impacted by data.

Relationships between reported frequency of data use and perceived usefulness. Correlations reported in Table 32, between the reported frequency of data use and the reported usefulness of data types, while present, were not as strong as might be expected from a simple hypothesized relationship where useful data types are used more frequently. The usefulness and frequency of use for software data, for example, were correlated at r = 0.34 and for discipline data at r = 0.31. Reported frequency of use for formal assessments, on the other hand, demonstrated positive correlations to the usefulness of several formal assessments: to the usefulness of Lexile levels (*r* = 0.41), drop out predictions (*r* = 0.42), computer adaptive testing (*r* = 0.47), and interim testing (*r* = 0.39). The highest correlation between reported frequency and usefulness was found for IEP information at *r* = 0.51.

Weaker correlations (or even negative relationships) between frequency and usefulness may suggest that the frequency of data use is often determined by factors beyond its perceived usefulness, such as mandated use, or school-policy based tasks. Alternatively, lower correlations may indicate information that is useful but available only infrequently, as with standardized tests administered annually or cognitive testing batteries administered every three years to special education students. Infrequent administration alone may not impact the perceived usefulness of these data sources. At least with survey responses these disjunctions between the perceived use of data types and their reported frequency of use raises concerns for how the TAM can account for mandated usage of information technology and data sources. If technology acceptance is motivated by obligatory professional demands, then relationships between PERUSE and usage outcomes may weaken.

Possible future work may wish to tease apart the opposite case as well, where data types are reported as less useful but more frequently accessed, as with an earlier description of discipline incident data: middle school teachers ranked their frequency of use for such data relatively high (Figure 29) but its usefulness much lower (Figure 33). Such a contrast may indicate areas where record-keeping feels out of balance with utility of information. It might also be useful to drill deeper on the types of data judged indispensable, but which are only available on an infrequent basis. There may be openings in these inconsistencies to increase the usefulness of less useful (but more frequently used) data, either by altering its reporting context or alternatively, by limiting bureaucratic demands for less useful data types in order to reduce unnecessary reporting for teachers.

Table 32

Intercorrelations of Reported Frequency and Perceived Usefulness

	Reported Frequency					
	1	2	3	4	5	6
Reported Frequency						
1 Own Assess.	—					
2 Discipline	.41	—				
3 IEP Info	.39	.05	—			
4 Software	.21	.45	-.10	—		
5 Formal Assess.	.43	.08	.35	.25	—	
6 State Test	.26	-.06	.60	-.05	.59	—
Perceived Usefulness						
7 Own Observe	.23	.06	.36	.17	.25	.28
8 Own Assess.	.12	-.11	.41	-.11	.19	.27
9 Gradebook	.26	-.06	.29	-.07	.20	.26
10 IEP Info	.06	.07	.51	-.05	.28	.38
11 Software	.13	.21	.30	.34	.28	.31
12 Cognitive Skills	.32	.44	.38	.25	.19	.17
13 State Stand.	-.12	-.15	.26	-.03	.20	.11
14 Lexile	.03	-.13	.28	.11	.41	.34
15 Attendance	-.09	-.18	.32	-.01	.22	.16
16 Past Grades	-.18	-.05	.29	.02	.07	.21
17 State Test	-.17	-.14	.24	.07	.26	.36
18 Discipline	.37	.31	.24	.21	.38	.09
19 Regents	-.26	-.32	.35	-.02	.25	.25
20 Drop Out Predict.	.12	-.06	.33	.11	.42	.21
21 Test Quiz	.08	.01	.34	.10	.38	.33
22 CAT	.17	-.06	.49	.16	.47	.28
23 State Test Predict.	-.17	-.13	.18	.08	.20	.27
24 Interim Testing	-.02	-.04	.42	.08	.39	.38

Note. Correlations higher than $r = 0.3$ or lower than $r = -0.3$ are highlighted in green

Relationships and Patterns Across Data-Use Attitudes and Efficacies. Description of results now turns from teachers' opinions on the frequency and usefulness of school-based data types to more general data use attitudes and self-efficacies. Basic descriptive survey results for General Data Use Attitudes (SEDU), Data Use Self-Efficacy (3D-MEA), and Teaching Self-Efficacy (TSES) are presented in Table 33, including minimum, maximum, mean, standard deviation, and Cronbach's Alpha calculated on non-standardized items. Because of low reliability, the Norwegian External Control Scale was calculated in two ways, both with and without a single reverse-coded item. Reliability was generally strong for all scales, higher than 0.8, except for the External Control Scale, which had a Cronbach's Alpha of 0.58, including a reverse-coded item, and 0.79 without. Any inferences based on the External Control scale should consider this lower reliability.

Table 33

Descriptive Statistics and Cronbach's Alpha of Survey Scales

Variable	n	Min	Max	Mean	SD	Cronbach's Alpha	Composite Variable
Survey of Educator Data Usage							
Data's Effectiveness for Pedagogy	37	1.00	4.00	3.34	0.75	0.96	Q1-5; (1) Strongly Disagree
Data Use Practice	37	1.40	4.00	3.25	0.70	0.92	Q1-5; (1) Strongly Disagree
Attitudes Towards Data	37	1.00	4.00	3.20	0.76	0.94	Q1-4; (1) Strongly Disagree
Computer Data Systems	37	1.50	4.00	3.11	0.67	0.91	Q1-4; (1) Strongly Disagree
Support for Data Use	37	1.50	4.00	3.01	0.64	0.88	Q1-6; (1) Strongly Disagree
Instructional Uses of Data /Instructional Resources	37	1.00	4.00	2.50	0.89	0.95	Q1-8; (1) Less than once a month
DBDM Efficacy and Anxiety and Survey (3D-MEA), Self-Efficacy for. . .							
Data Analysis and Interpretations	36	1.33	5.00	3.90	.74	0.92	Q1-3; (1) Strongly Disagree
Application of Data to Instruction	37	1.00	5.00	3.86	.78	0.95	Q1-6; (1) Strongly Disagree
Data Identification and Access	37	1.67	5.00	3.77	.79	0.81	Q1-3; (1) Strongly Disagree
Data Technology Use	36	1.00	5.00	3.66	.90	0.90	Q1-3; (1) Strongly Disagree
Teacher Self-Efficacy Scale – (TSES – Short)							
Efficacy in Instructional Strategies	36	5.50	9.00	7.48	.90	0.82	Q1-4; (1) Not at all
Efficacy in Student Engagement	36	5.00	9.00	6.80	1.07	0.81	Q1-4; (1) Not at all
Norwegian Teacher Self-Efficacy Scale (NTSES)							
Adapt Instruction to Individual Needs	37	3.50	7.00	5.60	.98	0.89	Q1-4; (1) Not certain at all
Norwegian External Control Scale							
External Control	35	1.80	5.40	3.08	0.82	0.58	Q1-5; (1) False
External Control (No Reverse-Coded Item)	35	1.00	5.50	2.94	0.53	0.79	Q1-4; (1) False

Correlational analysis. Relationships between all survey scales are presented in the form of a correlation matrix (Table 34). As stated in the Methods section, the small sample precludes any statistical significance tests with post hoc corrections. The correlations below are merely exploratory and await confirmation in future work.

Overall, subscales from the same survey are highly correlated with each other, such as the r = 0.82 between the Data's Effectiveness for Pedagogy and the Data Use Attitudes subscales of the SEDU. However, some subscales demonstrate interesting independence, such as the SEDU Computer Data Systems and Supports for Data Use scales which demonstrate lower relationships with other SEDU scales.

3D-MEA scales appear to be less inter-correlated than SEDU scales, indicating that these constructs of data-driven decision making self-efficacy (Technology Use, Identification and Access, Analysis and Interpretation, and Application to Instruction) are tapping into related but substantially separate areas of efficacy in teacher practice. The three subscales of general teaching efficacy—Efficacy for Instructional Strategies, Student Engagement, and Adapting Instruction to Student Needs—are also only moderately correlated, supporting their ability to distinguish between different forms of teaching self-efficacy.

Correlations across surveys indicate a few areas of overlap, as with the strong correlations between the SEDU scales of Effective Pedagogy, Data Use Practice, and Data Use Attitudes and the 3D-MEA scale of Efficacy for Applying Data to Instruction. Otherwise, the 3D-MEA scales for Efficacy in Analysis and Interpretation, Identification and Access, and Technology Use demonstrate only weak to moderate relationships with SEDU subscales. Since as far as can be determined the SEDU and 3D-MEA survey scales have not been previously compared, it is interesting to note their potential for capturing separate dimensions of data use.

General teaching efficacy scales in Instructional Strategies and Student Engagement show little correlation with either data attitudes (SEDU) or data efficacy (3D-MEA) scales, emphasizing the differences between general teaching efficacy and data use constructs. However, the Efficacy for Adapting Instruction subscale of the NTSES demonstrates some weak to moderate correlations with subscales of the SEDU and the 3D-MEA. Some relationship between these subscales would make intuitive sense given that the primary uses of student data are for adapting instruction. The weakness of the relationship, however, is worth noting in that teachers' sense of efficacy in adapting instruction appears to be largely independent of their attitudes or efficacy towards data use.

Table 34

Intercorrelation Matrix: All Survey Scales

	SEDU						3D-MEA				Teaching Self-Efficacy				
Survey Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
SEDU															
1 Effective Pedagogy	—														
2 Data Use Practice *(n=37)*	.80	—													
3 Data Attitudes *(n=37)*	.83	.72	—												
4 Comp. Data Syst. *(n=37)*	.17	.11	.41	—											
5 Supp. for Data Use *(n=37)*	.03	.05	.19	.35	—										
6 Instructional Res. *(n=37)*	.58	.73	.62	.16	.18	—									
3D-MEA															
7 Analysis/Interpret. *(n=36)*	.03	.19	.39	.50	.18	.39	—								
8 Apply to Instruct. *(n=37)*	.75	.68	.84	.36	.06	.65	.39*	—							
9 Identify/Access *(n=37)*	.28	.37	.41	.38	.20	.37	.54	.37	—						
10 Tech Use *(n=36)*	.23	.30	.52	.47	.14	.35	.61	.58	.62	—					
Teacher Self Efficacy (TSES, NTSES)															
11 Instruction. Strat. *(n=36)*	.00	.07	-.11	.18	.33	.08	.11^	-.14	-.05	-.20^	—				
12 Student Engage *(n=36)*	.15	.13	.05	.11	.08	.17	.01^	.10	.10	.16^	.46	—			
13 Adapt Instruction *(n=37)*	.08	.20	.24	.33	.60	.23	.30*	.23	.18	.18*	.41*	.43*	—		
14 Ext. Control *(n=35)*	-.19	-.32	-.17	-.12	.24	-.13	.14~	-.18	-.15	-.24~	.07	-.10	.18	—	
15 Ext. Control, No Rev *(n=35)*	-.28	-.38	-.25	-.12	.25	-.21	.10~	.21	-.10	-.16~	.01	-.09	.14	.95	—

Note. * *n=36, ^n=35, ~n=34*

The three following heatmaps explore possible subgroups of teachers' general data use attitudes (SEDU survey scales, Figure 39), data-use self-efficacy (3D-MEA survey scales, Figure 40), and combined data-use and teaching self-efficacies (3D-MEA, TSES/NTSES survey scales, Figure 41).

HCA Heatmap: General data use attitudes (SEDU). Figure 39 clusters participants in rows and subscales of general data-use attitudes in columns. Subscales capture teacher attitudes towards support for data use and relevant computer systems, along with more general attitudes towards whether data use is interesting and effective for instruction.

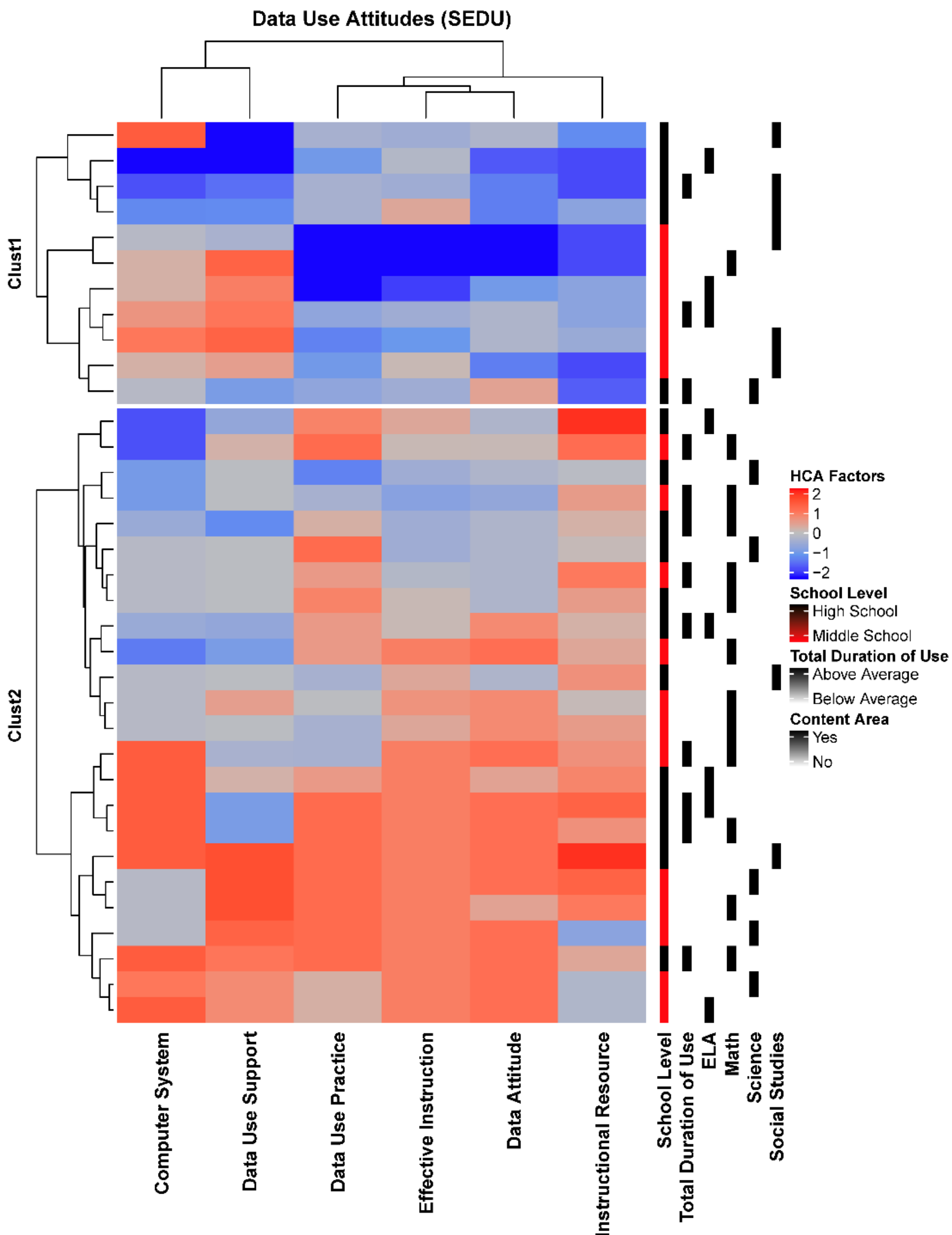


Figure 39. HCA heatmap: survey of educator data use (SEDU) Scales. NbClust suggests 2 clusters (10 Methods) and 5 clusters (7 Methods).

In Figure 39, teacher responses to SEDU subscales cluster into two main groups with interesting possible subgroups. Clust1 appears more negative to data use, while Clust2 appears more positive. Along with these two main clusters, an interesting subgroup of Clust1 appears to respond more favorably towards the school's data use systems and support, yet still respond negatively on scales of overall attitude towards data use. An opposed subgroup appears in Clust2, with more negative responses towards data support and systems, but with generally more positive responses towards data use attitudes. Clust2 also includes a subgroup with a positive view of both data support and data use attitudes.

In relation to teacher factors, higher levels of usage, along with a greater concentration of Math and Science teachers appear in Clust2 with more positive attitudes towards data use. Clust1, with lower data use attitudes, contains a larger concentration of middle school and Social Studies teachers.

While one might guess that higher perceptions of data-use support would go hand in hand with more positive perceptions of data use in general, Figure 39 indicates that this relationship may apply only some of the time and that teachers' perceptions of the quality of support may operate independently from their general attitudes towards data use. Teachers, in other words, may feel supported in their efforts to use data, but that level of support may not alter an existing negative stance towards DBDM.

HCA Heatmap: Data-driven Decision Making Efficacy scales (3D-MEA). Figure 40 clusters participants in rows and four subscales of data-driven decision making efficacy in columns. Subscales assess confidence in data identification and access, data technology use, data analysis and interpretation, and applying data to instruction.

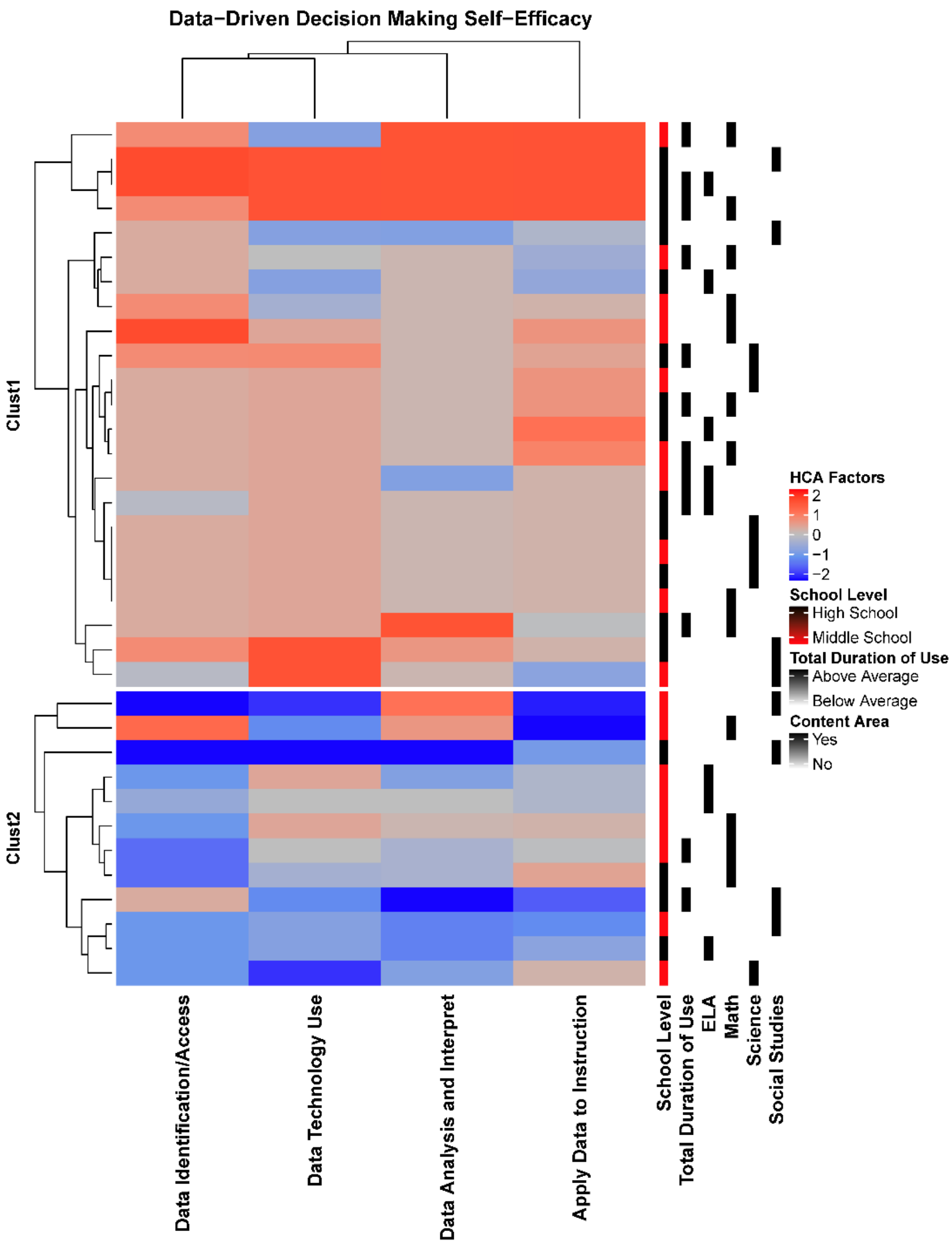


Figure 40. HCA heatmap: data-driven decision making (DDDM) efficacy scales (3D-MEA). NbClust was unable to arrive at a solution, possibly because of highly-correlated values. Visual inspection suggests two clusters.

Clustering four scales of the 3D-MEA survey suggests two major clusters (Figure 40): Clust1 of users who indicate higher levels of DDDM self-efficacy and Clust2 of responses indicating lower levels of DDDM self-efficacy. While only two major clusters are evident, a subgroup of particularly high self-efficacy also appears in Clust1. Annotations suggest a strong relationship between higher levels of DDDM self-efficacy and higher than average online use. Math and Science content areas also have a stronger relationship to higher DDDM Efficacy. This heatmap suggests that data-use self-efficacy may have a stronger relationship to usage outcomes and content area than did SEDU scales for general data use attitudes.

HCA Heatmap: Data-Driven Decision Making and Teaching Self-Efficacy (3D-MEA, TSES). The final HCA heatmap in this section of the results combines participants in rows with data use, self-efficacy, and teaching self-efficacy in columns. This combined view of a range of self-efficacies explores the ways in which teacher confidence in general instructional skills relates to confidence in data use.

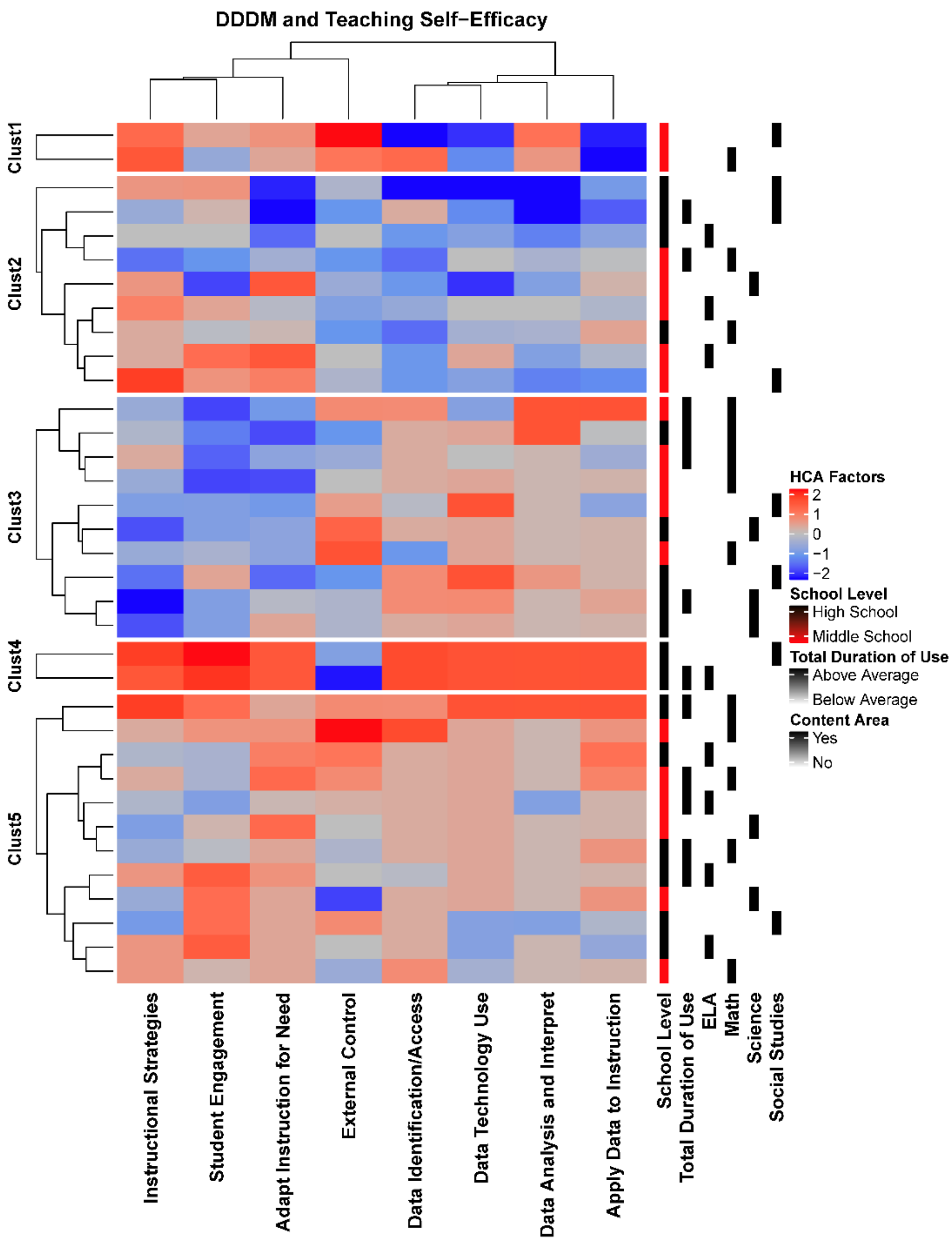


Figure 41. HCA heatmap: data-driven decision making and teaching self-efficacy. NbClust suggests 5 clusters (12 methods).

Overall, columns in Figure 41 group neatly into a teacher self-efficacy cluster on the left (including all scales from the TSES and NTSES) and, on the right, a data-use self-efficacy cluster of subscales from the 3D-MEA. Across rows, clustering general teaching self-efficacy along with data-use self-efficacy identifies several subgroups based on positive or negative trends in the two broad column clusters. Clust1 and Clust2 responses indicate higher levels of general teaching self-efficacy and lower levels of data use self-efficacy. Clust3 responses show the opposite pattern: higher responses for DDDM self-efficacy and lower for general teaching self-efficacy. Clust4 and Clust5 include participants ranking themselves higher in all areas of self-efficacy. This range of clusters highlights the separation of teachers' efficacy in fundamental areas of instructional strategies and student engagement, from confidence in data use skills, such as access and interpretation of data.

Strong patterns of content area and online use are also visible across these clusters. Interestingly, above average use is more present in Clust4 and Clust5 where users are confident in both teaching and data use than it is in Clust3 where users are more confident in data use only. Clust3 users tend to be Math teachers, while ELA teachers tend to group in Clust2 and Clust4/5, not unconfident with data use, but only in combination with efficacy in general teaching skills.

Exploratory Correlations of Data Use Attitudes and Efficacies. The same data use attitudes and efficacies explored above through visual data analytics are also investigated through exploratory Pearson correlations. Table 35 reports relationships between teacher survey scales and more specific metrics of online data use developed in Study 1. While most of these metrics for online use, such as total actions, total sessions, total duration, and maximum weekly duration/total duration are explained in Study 1, an additional factor of total duration (view results only) requires introduction. Total duration (view results only) separates out the duration of time that users spent specifically viewing test results—the main instructionally-focused use of the system—in order to explore relationships to that specific category of use.

Several interesting correlations are found between the various survey scales (SEDU, 3D-MEA, and TSES) and observed online behaviors. Perhaps most interesting are the correlations (Table 35) related to general data use attitudes (SEDU subscales) and data use self-efficacy (3D-MEA). Neither factor produced strong correlations with usage outcomes, but multiple moderate-strength correlations were found between DDDM self-efficacy scales (3D-MEA) and usage outcomes, while consistently weaker correlations were found between general data use attitudes and usage factors. Efficacy for Applying Data to Instruction, from the 3D-MEA, for example, correlated at $r = 0.42$ to total sessions of online use, and Efficacy for Interpretation and Analysis correlated at $r = 0.48$ to total duration (view results only). Overall, this same scale, Efficacy for Interpretation and Analysis, appeared to have the strongest relationship to usage outcomes of any variable outside of teachers' perceived usefulness of the Benchmark Data system itself. Items in this scale asked about teachers' confidence in (a) understanding assessment reports, (b) interpreting student performance from a scaled score, and (c) interpreting scores to determine student strengths and weaknesses. Since these interpretation activities are all central to the use of the Benchmark Data platform this stronger correlation seems a reasonable fit between a specific area of user self-efficacy and the intended use of the system.

The same 3D-MEA scales also demonstrated the strongest negative correlation with teachers' tendency to focus usage in mandated professional development sessions (max weekly/total duration); the higher a teacher's efficacy in Interpretation and Analysis of data, the weaker their tendency to focus their usage in one week of the school year ($r = -0.41$). While previous studies involving the 3D-MEA have found a significant relationship between Efficacy for Application of Data to Instruction and self-reported data use behaviors, this study may be the first to tentatively relate DDDM Self-Efficacy to teachers' DBDM behaviors in the form of usage in an online data system.

Table 35

Intercorrelation Matrix: Survey Scales and Usage Outcomes

Survey Scale	Total Actions	Avg. Session Duration	Weeks Used	Max Weekly/ Total Dur.	Total Session Count	Total Duration (Results Only)	Total Duration of Use
Survey of Educator Data Use (SEDU)							
Computer System	.17	-.19	.14	.06	.11	.07	.13
Data Use Attitude	.25	-.24	.22	-.14	.24	.20	.23
Data Use Practice	.14	-.08	.15	-.25	.17	.08	.14
Effective Instr.	.08	-.16	.08	.00	.09	.07	.09
Instructional Resource	.02	-.00	.25	-.26	.16	.16	.07
Data Use Support	-.08	-.13	-.09	.24	-.13	-.19	-.16
Data-Driven Decision Making Efficacy (3D-MEA)							
Apply to Instr.	.35	-.31	.37	-.35	.40	.33	.35
Ident. & Access	.27	.17	.22	-.19	.29	.35	.27
Interpret. & Analysis	.39	.04	.45	-.41	.47	.48	.44
Technology Use	.32	-.05	.25	-.30	.28	.24	.30
Teacher Self-Efficacy Scales (TSES, NTSES)							
Student Engagement	-.15	.23	-.30	.12	-.32	-.27	-.15
Instructional Strategy	-.08	.21	.02	.17	-.10	-.08	-.03
Adapt for Instruction	-.07	.00	-.14	.11	-.18	-.19	-.17
External Control	-.02	.03	.01	.12	-.03	-.01	-.05

Exploration of the Technology Acceptance Model (TAM) for understanding teachers' online data use. Shifting from a consideration of teachers' attitudes and efficacies, the final section of the results explores the core factors of the technology acceptance model: perceived usefulness (PERUSE) and perceived ease of use (PEOU) as determinants of teachers' online data use. The concept of determinants or a determinant framework (Nilsen, 2015), in this case, does not imply any causal relationships, but rather a set of factors that may be related to implementation success. In this section, PERUSE factors are operationalized from survey items asking, "How useful do you find" the Benchmark Data system, and items asking, "How useful do you think the following kinds of data are. . .", for those types of data included in the online system. PEOU was operationalized through items asking about participants' general comfort level with technology and their level of use for the Benchmark Data System. Also, in order to tailor the PEOU construct to the DBDM context, additional items for PEOU were drawn from the 3D-MEA subscales for data use self-efficacy. Given the exploratory nature of Study 2 and the strong relationship found in previous studies between PEOU and self-efficacy (Davis, 1989, Venkatesh and Davis, 1996, Venkatesh, 2000) these data use self-efficacy items are included as a proxy for teachers' PEOU for the online DBDM system.

HCA heatmaps of PEOU, PERUSE, and online use. The three following HCA heatmaps relate the constructs of PEOU and PERUSE to teachers' online use and professional roles. Figure 42 begins by clustering factors related to the PEOU of the Benchmark Data System, including survey scales for data-use self-efficacy, general comfort level with technology, and level of use for Benchmark Data. Figure 43 continues by clustering these PEOU factors alongside factors related to the PERUSE of the Benchmark Data System, combining the major factors of the TAM into one visual data analysis. Finally, Figure 44 clusters PEOU and PERUSE alongside several metrics for online use developed in Study 1 in order to explore more fine-grained relationships between data use attitudes and specific usage behaviors.

HCA Heatmap: Perceived Ease of Use (PEOU). Figure 42 clusters users in rows and system-specific PEOU factors in columns. Though similar to previous clustering around self-efficacy for DBDM, the PEOU construct used here differs by including users' level of expertise with the data system and their general level of comfort with technology.

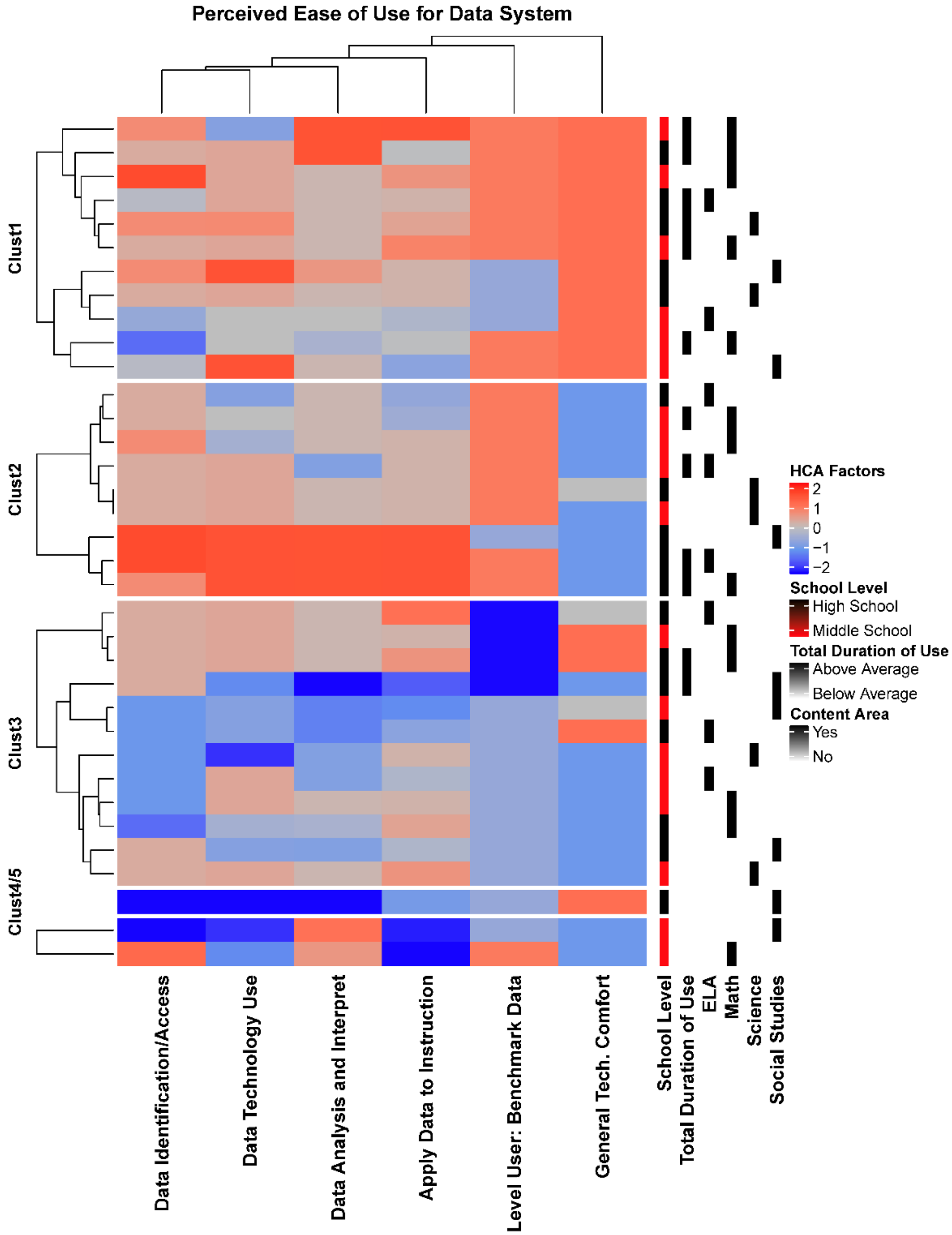


Figure 42: HCA heatmap: perceived ease of use for Benchmark Data system. Nine NbClust methods proposed 5 clusters; 8 methods proposed 2 clusters.

Overall, the top two clusters of Figure 42 (Clust1 and Clust2) group responses with higher PEOU for the Benchmark system. Clust2 differs from Clust1 in that Clust2 responses were slightly lower for general comfort level with technology, while still high for system-specific ease of use. Both Clust1 and Clust2 relate strongly to above average usage, as with the strong relationship to online use found in Figure 40 for DDDM self-efficacy alone. Unsurprisingly, high responses to level of use for the Benchmark system appear strongly related to above average usage. Many of the cases where higher levels of DDDM self-efficacy do not indicate above average use belong to Science and Social Studies teachers, who have been found throughout to have lower levels of usage. Of the 23 teachers clustered with higher levels of DDDM efficacy, 11 indicate above average usage of the Benchmark system. Of the remaining 12 teachers, half are either Science or Social Studies teachers. While many Social Studies teachers indicated low PEOU for the online data system, a factor which might help explain their more limited usage, even those Social Studies teachers with high PEOU appear to have made more limited use of the system.

The lack of clustering across columns is also interesting to note in Figure 42. Instead of finding subclusters of factors, the HCA appears to find a pattern of increasing distance from the concrete access and use of technology to the left of heatmap to the more general construct of comfort with technology.

HCA Heatmap: Perceived ease of use and perceived usefulness. Combining PEOU and PERUSE, Figure 43 clusters along key elements of the TAM.

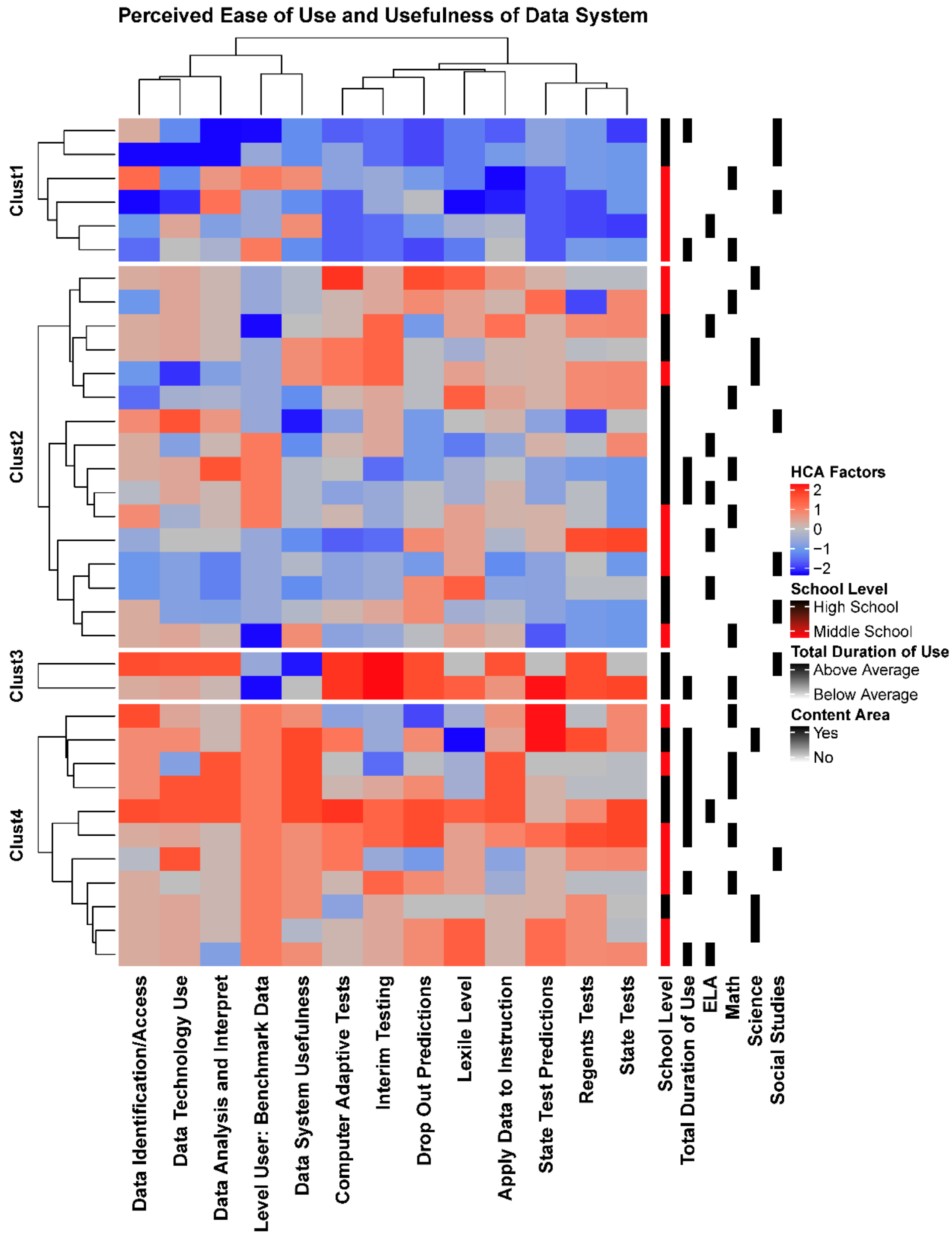


Figure 43. HCA heatmap: perceived ease of use and usefulness for the Benchmark Data system. 11 methods of NbClust propose 2 clusters, 5 propose four clusters. Visual inspection suggests four clusters.

Supporting the expectations of the TAM, combining PEOU and PERUSE in Figure 43 creates more precise clustering in relation to online data use. In the previous cluster for high PEOU (Figure 42), 50% of users (10 of 20) demonstrated above average usage. Looking at Figure 43, in clusters identified for both high PEOU and high PERUSE (Clust3 and Clust4), 62% of users (8 of 13) demonstrated above average usage. Of the five users in Clust3 and Clust4 without above average usage, three were middle school Science or Social Studies teachers, without access to interim testing data for their content areas. Overall visual analysis of PEOU and PERUSE factors suggests they are highly relevant to teachers' online data use.

While generally clustering along the lines of PEOU and PERUSE factors, two factors break this pattern in an interesting way. First, the item asking about the perceived usefulness of the data system itself falls into the cluster of items assessing the system's PEOU. On the other hand, one item measuring self-efficacy for applying data to instruction is grouped not with other items intended to capture PEOU, but with items assessing the PERUSE of data types within the system. Generally, though, there is a marked degree of homogeneity to responses across both PEOU and PERUSE responses.

HCA heatmap: PEOU, PERUSE, and online use. In a final HCA Heatmap for these results, Figure 44 clusters the previously analyzed TAM factors of PEOU and PERUSE with a range of metrics for online use, developed in Study 1.

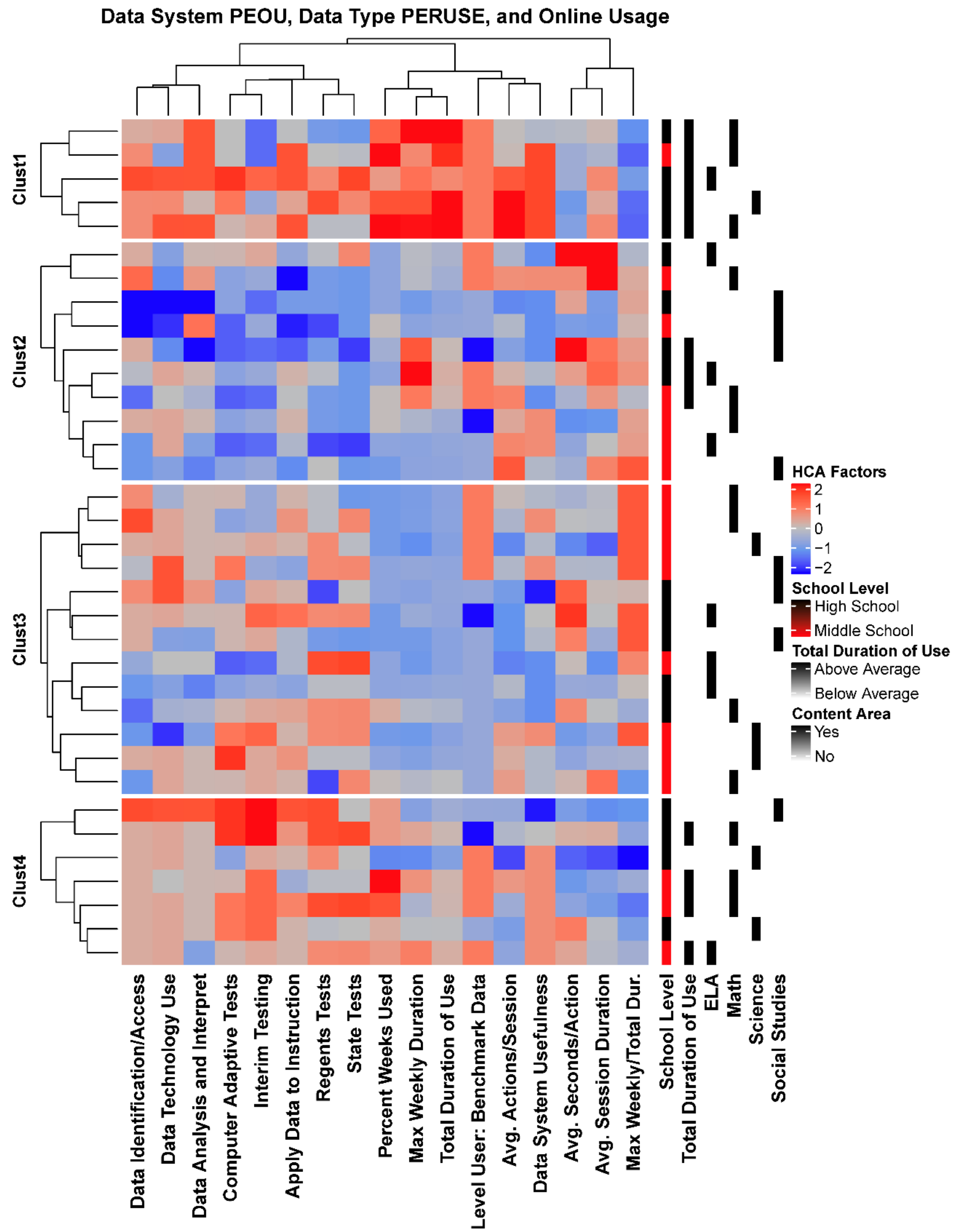


Figure 44. HCA heatmap: PEOU, PERUSE, and online usage metrics. 11 NbClust methods propose 2 clusters; 7 methods suggest 3 clusters. Visual inspection suggests 4 clusters for interpretation.

Of the four clusters identified for interpretation, Clust1 and Clust4 appear to group users with higher PEOU, higher PERUSE of the Benchmark system and data types, and higher online use. The difference between Clust1 and Clust4 users appears to be one of degree, where Clust4 users demonstrate slightly lower levels for factors such as percentage of weeks used, total duration of use, and average actions per session. While users in Clust4 demonstrate somewhat higher levels for these factors, Clust1 users demonstrate the highest levels across all clusters. Most of the users in these two clusters demonstrated higher than average total usage of the online system.

In contrast, only three of the 23 users in Clust2 and Clust3 demonstrated above average usage. While the online behaviors of users across Clust2 and Clust3 appear similar, they differ more strongly along the lines of the first cluster of columns, which includes scales for data use self-efficacy and the perceived usefulness of testing data. Clust2 responded more negatively across all these factors, demonstrating lower PEOU and PERUSE than Clust3, while still demonstrating similar patterns of usage.

Inspection of the HCA clustering of columns in Figure 44 reveals that most items related to the PEOU and PERUSE of the Benchmark system itself clustered together on the left, while usage metrics clustered on the right. Self-efficacy for access, interpretation, and analysis clustered more tightly with the PERUSE of specific types of testing data, such as state testing, interim testing, and computer adaptive testing. It is important to note that higher levels of this combined PEOU and PERUSE factor, while highly related to above average usage, are not sufficient to guarantee it. In fact, a large group of users appear to have both higher PEOU and PERUSE for a set of extremely relevant data use factors, without manifesting higher than average use of the Benchmark system. Conversely, a small subset of Clust2, with lower PEOU and PERUSE, does demonstrate above average use. While difficult to tease apart in this limited analysis, content area and school level may play a critical role in these contradictory cases of higher PEOU and PERUSE, but lower usage of online student testing data.

Exploratory Pearson correlations between PEOU, PERUSE, and online usage metrics. Results for additional analysis of these factors are presented below in the form of exploratory Pearson correlations. Table 36 reports correlations between technology acceptance factors and a set of metrics related to online use. While overall correlations between PEOU, PERUSE, and observed online usage factors presented in Table 36 were low to moderate, some interesting trends in correlation are summarized below.

In keeping with the TAM, teachers' PEOU and PERUSE of the Benchmark Data system had the strongest correlations to usage outcomes. These included Level of Benchmark Data User ($r = 0.39$ to Total Sessions), perceived usefulness of the Benchmark system ($r = 0.54$ to Total Actions), and perceived usefulness of progress monitoring quiz data ($r = 0.50$ to Total Duration, View Results Only). While not surprising, these stronger relationships are reassuring in that they leave open the possibility that by increasing teachers' perceived ease of use, as well as their perceptions of system usefulness, long term teacher usage might be increased as well. Though clearly causation could act in the opposite direction as well, with increased system usage leading to improved perceptions of usefulness and ease of use.

The degree to which teachers found Attendance/Tardiness and Progress Quiz data useful also correlated positively at a moderate level with a range of usage outcomes ($r = 0.44$ with Weeks Used). While stronger correlations with progress monitoring quizzes ($r = 0.50$ to Total Duration, View Results Only) match expectations—since progress quiz results were only available through Benchmark Data and available at more frequent intervals—a stronger relationship with Attendance/Tardiness data and usage outcomes is more difficult to explain, since that type of data was generally unavailable in Benchmark Data. Additionally, since Benchmark Data was a main source for state testing results, computer adaptive testing results, regents testing results, and state standards information, it is surprising that correlations between the perceived usefulness of these data types and usage outcomes are not stronger. It may be that even for the most dedicated user of state testing results, infrequent annual and biannual state testing simply does not enough offer enough time to substantially increase overall usage outcomes. On the other hand, interest in more frequently administered progress monitoring quizzes appears to be more closely related to actual online use.

The degree to which teachers find interim testing useful might also have been hypothesized to have a strong relationship to Benchmark Data usage. One of Benchmark Data's major functions, after all, was the dissemination of interim testing results. However, no such relationship was found. It may be that required training sessions specifically directed at accessing interim testing results clouded the relationship between the perceived usefulness of interim testing and actual online usage in a way that does not occur for other data sources. This lack of relationship on a key source of instructional data clearly merits further investigation.

On the other hand, in terms of negative relationships, teachers' tendency to focus their online usage in mandated sessions, captured by the outcome for Max Weekly/Total Duration, is only weakly negatively correlated to teachers' PEOU and PERUSE for Benchmark Data overall. Stronger negative correlations were found between max weekly/total duration and the usefulness of progress monitoring quizzes (r = -0.48), drop out predictions (r = -0.44), and discipline incidents (r = -0.43). These stronger negative relationship between perceived usefulness of progress quizzes and actual online use may be due to teachers logging in more consistently to access information they find useful. The other two relationships are more difficult to interpret: Since discipline incidents were only available in Benchmark Data in a limited form and drop out predictions available only at one point in the semester, it is more difficult to imagine how these factors might demonstrate a stronger negative relationship to teachers' tendency to restrict their online usage to mandatory training sessions. Some of the implications for these relationships will be addressed in the Discussion section.

Table 36

Intercorrelation Matrix: Perceived Ease of Use, Perceived Usefulness, and Online Use

Variable	Total Actions	Avg. Session Duration	Weeks Used	Max Weekly/ Total Dur.	Total Sessions	Total Duration (Results Only)	Total Duration of Use
Perceived Ease of Use (PEOU)							
General Tech Comf.	.11	-.23	.09	-.08	.22	.20	.14
Lev User: Benchmark	.37	.11	.28	-.25	.39	.39	.36
Perceived Usefulness of Data Types and Systems (PERUSE)							
Benchmark Data Sys.	.54	.01	.42	-.25	.50	.49	.47
Attend/Tard.	.35	.15	.44	-.38	.36	.34	.38
Comp. Adapt Test	.21	-.02	.34	-.31	.30	.40	.22
Disc. Incident	.34	-.12	.38	-.43	.33	.32	.29
Drop Out Pred.	.24	-.30	.38	-.44	.26	.19	.16
Interim Test	-.04	-.06	.15	-.21	-.03	-.05	-.11
Lexile	-.28	-.30	-.07	.09	-.17	-.16	-.24
Progress Quizzes	.31	-.01	.39	-.48	.42	.50	.35
Regent Test	.19	-.24	.20	-.17	.21	.14	.09
State Standards	-.03	-.05	.04	-.27	.00	.06	-.08
State Test	.10	-.02	.20	-.26	.10	.15	.06
State Test Prediction	.26	-.06	.25	-.22	.23	.22	.19

Summary of Results. Study 2 had the overarching goals of expanding the methods and theory of DBDM research by combining visual data analytics and the TAM to investigate determinant frameworks for online data use. Results were organized into three main sections:

1. A school-specific data use profile by reported use and usefulness of data types;
2. Relationships and patterns across data-use attitudes and efficacies; and
3. An exploration of the Technology Acceptance Model (TAM) for understanding teachers' online data use.

Analyses produced a range of descriptions for local data use and attitudes.

The first section generated a school data use profile based on teachers' reported frequency of use and usefulness of specific types data. These results suggest foundational differences in data use across school level and content areas that may be helpful in framing variation in data use for future studies.

The second section searched for relationships across a range of data use attitudes, from general attitudes towards data (SEDU), to data use efficacy (3D-MEA) to general teaching efficacy (TSES/NTSES). Overall, general attitudes towards the effectiveness and practice of data use were highly intercorrelated, but distinct from data use self-efficacies, which in turn were distinct from teacher self-efficacy, even a teacher self-efficacy directed towards adapting instruction to students' needs. These findings highlight the multiple attitudes and efficacies that must be considered separately when planning around the determinants and outcomes of data use.

The third section explored teachers' PEOU and PERUSE in relation to their online data use. As suggested by TAM, these two factors were more indicative of online use than were attitudes towards the general effectiveness of data. PEOU and PERUSE were possibly more effective in clustering high online usage when used together than when considered separately.

Along with results from these three sections, several subgroups of both user behaviors and factors were also generated from HCA heatmap analyses. These are summarized in Table 39 and interpreted in the Discussion section.

Applied Results: Data Dashboards for Educational Leadership. Turning from analysis to action, a final set of results includes a collection of data use dashboards designed for school leadership. I present several hypothetical dashboards, both at the school- and teacher-level, each integrating some of the most important results discussed throughout Study 2. School-Level dashboards, in Figure 45 and Figure 46, offer a faceted, multiform view of related data. Such views can exploit multiple visual encodings of data to support multiple abstractions and their related tasks. A single more complex view, such as an HCA heatmap, might present more information in one view, but the visual encodings of more complex single displays are less able to prioritize specific factors for analysis or action (Munzner, 2014). School-level dashboards combine treemaps, heat maps, and bar charts as separate facets of the display to more directly inform specific areas of decision-making around data use.

The factors included in both school- and teacher-level displays (Figure 45 and Figure 47) were selected based on the findings of Study 2. For example, the treemap, in the top-left position of the school-level views (Figure 45 and Figure 46), visualizes both overall online use and the factor of data use self-efficacy (3D-MEA), found to be more highly related to online use than general data use attitudes. Positioning the treemap as an initial view at the top left follows the common visual pattern of Overview First, Zoom and Filter, and Details on Demand (Shneiderman, 1996), where the treemap provides a school-specific Overview of critical factors (Online Use, Ease of Data Use, and Content Area). The heatmap and bar chart facets that follow provide the Zoom and Filter aspect of presentation, while the teacher-level display provides Details on Demand for specific teachers.

The use of a treemap display leverages the strength of this visualization idiom to show complete information about hierarchical school structure through containment (Munzner, 2014), allowing school leaders to quickly identify relationships between content area, teachers' frequency of online use, and teachers' self-efficacy with data use. Each sub-rectangle of the treemap represents one teacher, with the size of the rectangle indicating a teacher's relative weeks of use and its color indicating their relative self-report for data use efficacy. On a slightly larger scale, the grouping of rectangles by content area allows for easy comparison across content groups on these same dimensions.

Treemaps are particularly useful for the current situation, where the goal of analysis is to better understand attributes at the leaves of a shallow hierarchical tree (Munzner, 2014), in this case, the attributes of teachers within the shallow hierarchy of a school—teacher in content area in school. Treemaps are also helpful for identifying outliers in the displayed attributes, in this case identifying teachers with particularly small or large values for data use or data use self-efficacy. While human judgements comparing the rectangular areas used in treemaps are more prone to error than those comparing bars in a bar chart (Munzner, 2014), the intended task for the treemap view is not to make specific ranking decisions about teachers' online use or self-efficacy, but rather to diagnose more general issues related to data use within the organizational hierarchy, quickly identifying content areas or teachers in need of support.

The treemap of Figure 45, for example, efficiently communicates the degree to which Math teachers' usage dominates middle school online access, while at the same time demonstrating how the large majority of Math teachers feel stronger in their efficacy for data use. Overall trends in content area use and efficacy are visible along with individual cases meriting further attention, such as a middle school Social Studies teacher with relatively low data use efficacy but higher levels of online use. Or, in the case of high school (Figure 46), members of the ELA and Social Studies team with lower levels of self-efficacy, along with lower levels of overall use.

Once an area of need is identified in the treemap view, the heatmap view to the right provides additional context for supporting teachers or the content team. The heatmap display allows for comparison of PERUSE under the title "How Useful," PEOU or data use self-efficacy with the title, "How Easy," and the percentage of weeks of online data use. Since measures of overall online frequency were found to be highly correlated, Percentage of Weeks Used was selected for its greater interpretability, based on a more direct mapping onto the school year. Teachers' perception of how supported they felt was included as well, not because it was found to be directly related to online use, but because it most directly makes the connection between the actions of leadership and teachers' efficacy and online use. Including perceptions of support in the heatmap more directly supports ownership by school leadership in teachers' data use.

Matrix re-ordering is also employed in the heatmap view, as rows within each content area are ordered from the highest Percentage of Weeks Used to the Lowest Percentage of Weeks Used. This ordering has a subtle effect of prioritizing the outcome of access to online data. The heatmap view itself is employed for its high density of information, displaying information for multiple users and factors in a compact space (Munzner, 2014). The limitation of the heatmap view is a lack of specificity in the values presented due to the inability to distinguish between subtle variations in color (Munzner, 2014), however, because the goal of the view is not to draw fine comparisons between users, this limitation does not impede the overall goal of providing support for teacher data use.

In the particular case of Progress Secondary, the heatmap view provides a quick comparison across content areas and schools, indicating large differences in perceived support between middle and high school, while also providing a method for diagnosing teacher needs. ELA teachers in the middle school, for example, demonstrate lower ease of use alongside lower weeks of usage, suggesting a starting place for professional development. Or, in the high school, the heat map views suggest different approaches for support to ELA and Social Studies teachers: While both content areas indicate lower levels of support for data use, the concerns of Social Studies teachers appear to extend to strong concerns over the usefulness of the data and their ease of use for data systems. ELA teacher responses are less negative in both these areas, suggesting that their perceived lack of support may be related to other, more technical or logistical concerns.

The bar charts that follow offer middle and high school instructional leaders further specificity into each school's perceptions of the usefulness of data types and activities and the opportunity to respond based on these schoolwide priorities. Color is used in these views to group different types of data on the left and reported use versus frequency of use on the right. These views of data's usefulness also offer insight when viewed at the level of individual teacher profiles in Figure 47, Figure 48, Figure 49, discussed next.

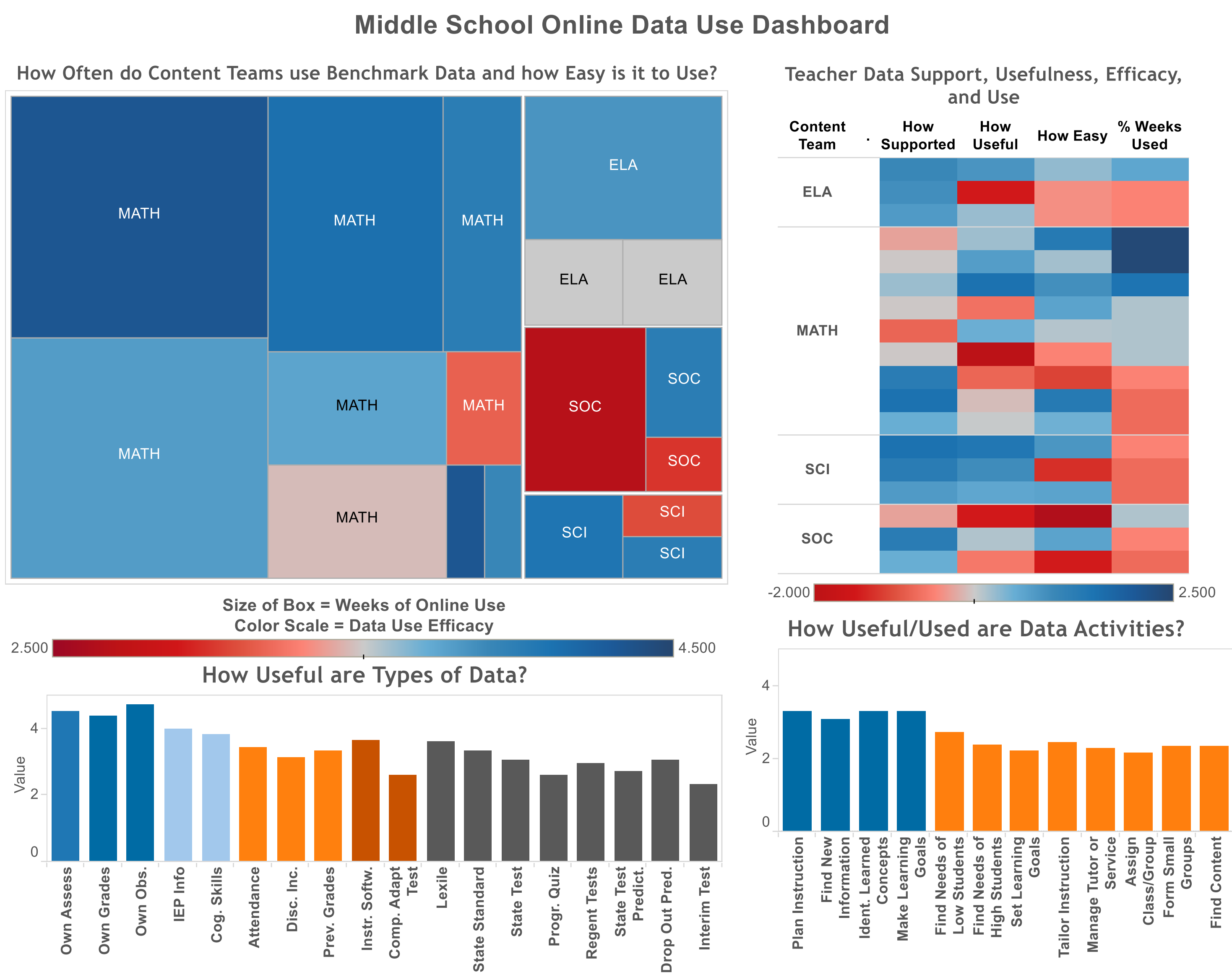


Figure 45. Data use dashboard (middle school)

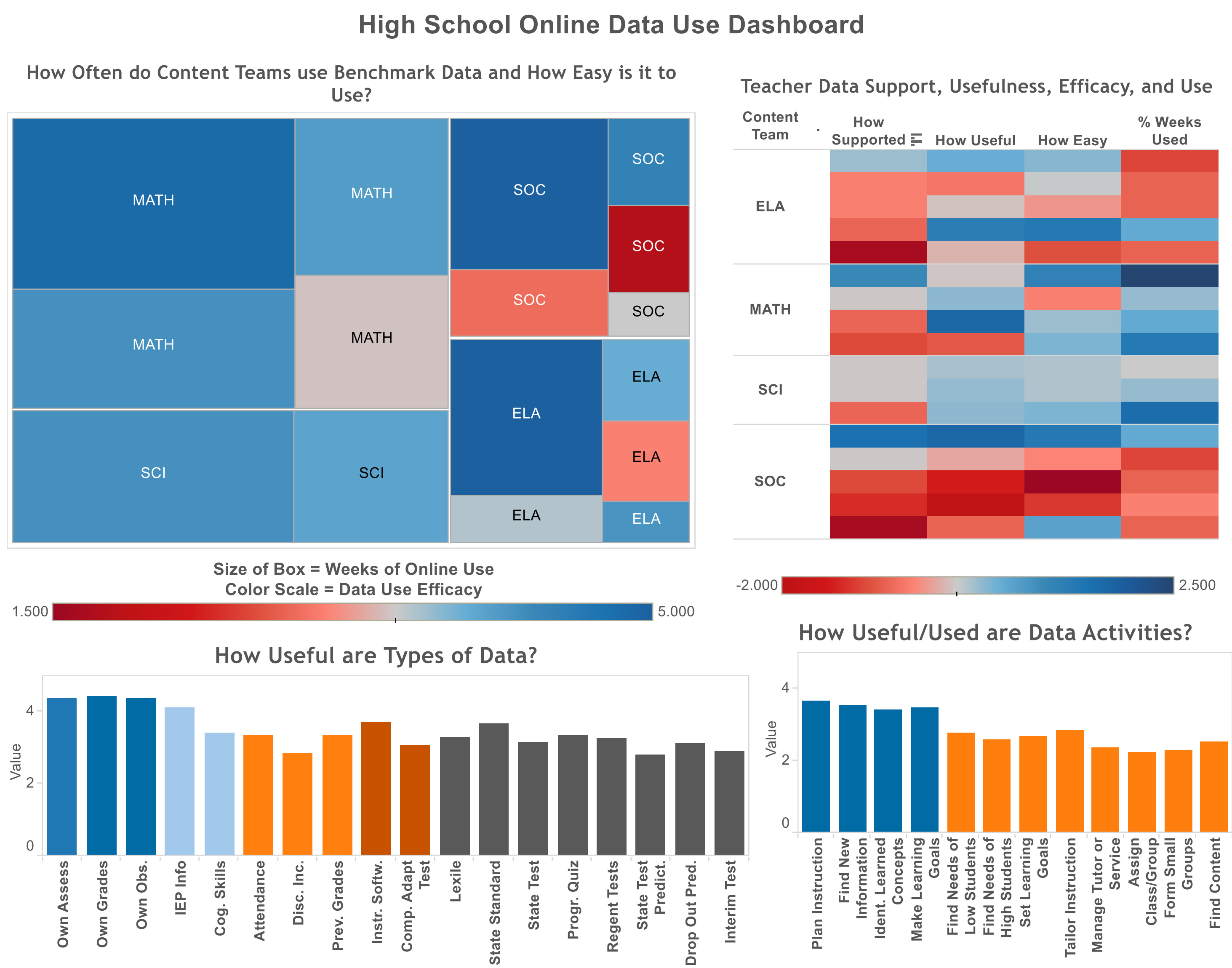


Figure 46. Data use dashboard (high school)

In addition to schoolwide dashboards, three examples of teacher profiles of data use and attitudes are presented. Figure 47 presents the profile of Teacher 1, a high school ELA teacher with above average overall use; Figure 48 presents Teacher 2, a middle school Math teacher with approximately average use; and Figure 49 describes Teacher 3, a middle school ELA teacher with online access far below the average. Each of these teacher profiles indicates the complex nature of attitudes and usage when it comes to evidence use in schools.

Teacher 1 (Figure 47), for example, indicated strong levels of usefulness for most data types and activities, as well as demonstrating above average use. These strong perceptions and online access, however, appear to have been achieved in spite of Teacher 1 indicating much lower levels of school support for data use. It is interesting to consider how more advanced users may perceive levels of support as low, not in an absolute sense, but to the degree that support fails to add to their already substantial skills. The same teacher also reports lower frequency of data use for forming small groups and selecting content. These specific data points might be helpful as an entry point for discussion about how this teacher might be better supported.

Teacher 2 (Figure 48) poses a sharp contrast with Teacher 1. As a middle school Math teacher, Teacher 2 comes from a team with some of the highest levels of use and most positive attitudes towards data use yet indicates strong negative feelings about the usefulness of multiple types of data. While Teacher 2 accesses the online system at an average level, his or her perceptions of the usefulness of data types and lower data use self-efficacy may be prompts for further discussion around how data types could be made more useful, particularly in the context of a content team with generally more positive attitudes towards data types and activities.

Teacher 3 (Figure 49), a middle school ELA teacher, offers an interesting contrast to Teacher 2. Where Teacher 2 had low perceptions of usefulness for many types of data and average online use, Teacher 3 has high perceptions of usefulness for many relevant data types, but well-below average access to online data. Teacher 3 does, however, indicate lower levels of data use self-efficacy, which may be related to his or her overall lack of use. While Teacher 3 indicates stronger school support, his or her lower levels of data use efficacy indicate that school support was not sufficient to encourage online access.

The uses and audiences for such school-level and individual profiles are many. At their most basic, they demonstrate how multiple organizational and individual determinants feed into decisions to engage with evidence. While school leaders might effectively use similar tools on their own for planning and making decisions, another possibility is to create teacher-facing versions of these tools in order to generate useful discussions between teachers and school leadership on feasible and impactful approaches to evidence use, particularly in regard to adapting data types and activities to the particular needs of each content area. For instance, the observation that a Social Studies teacher does not find certain data types useful should lead to discussion with a content team of whether that data is, in fact, useful, and if so, by what routes it impacts classroom decisions.

While these hypothetical profiles are already dense, a useful next step may be to include information summarizing each teachers' *patterns* of usage in order to provide feedback on the system functions teachers choose to prioritize. Another indicator to include might be the degree of teachers' connection to state-level accountability demand. Middle school ELA and Math, along with high school teachers of courses ending in state tests, would be identified at a higher level of articulation with the state accountability system. Given the exploratory nature of these studies, these dashboards are generated as a starting place for future collaborative work creating visualizations to improve the nature and use of evidence for school decisions.

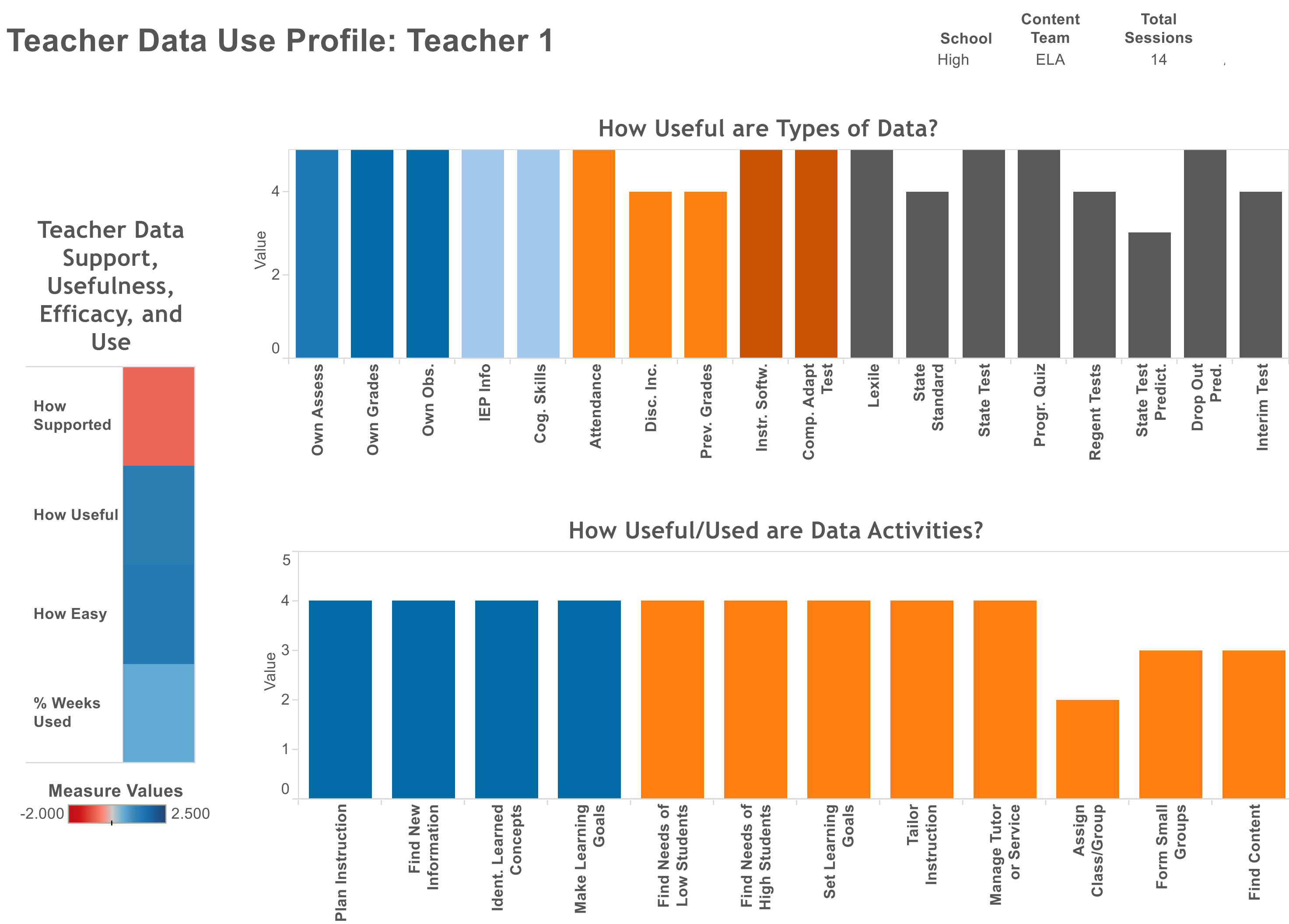


Figure 47. Teacher data use profile (Teacher 1)

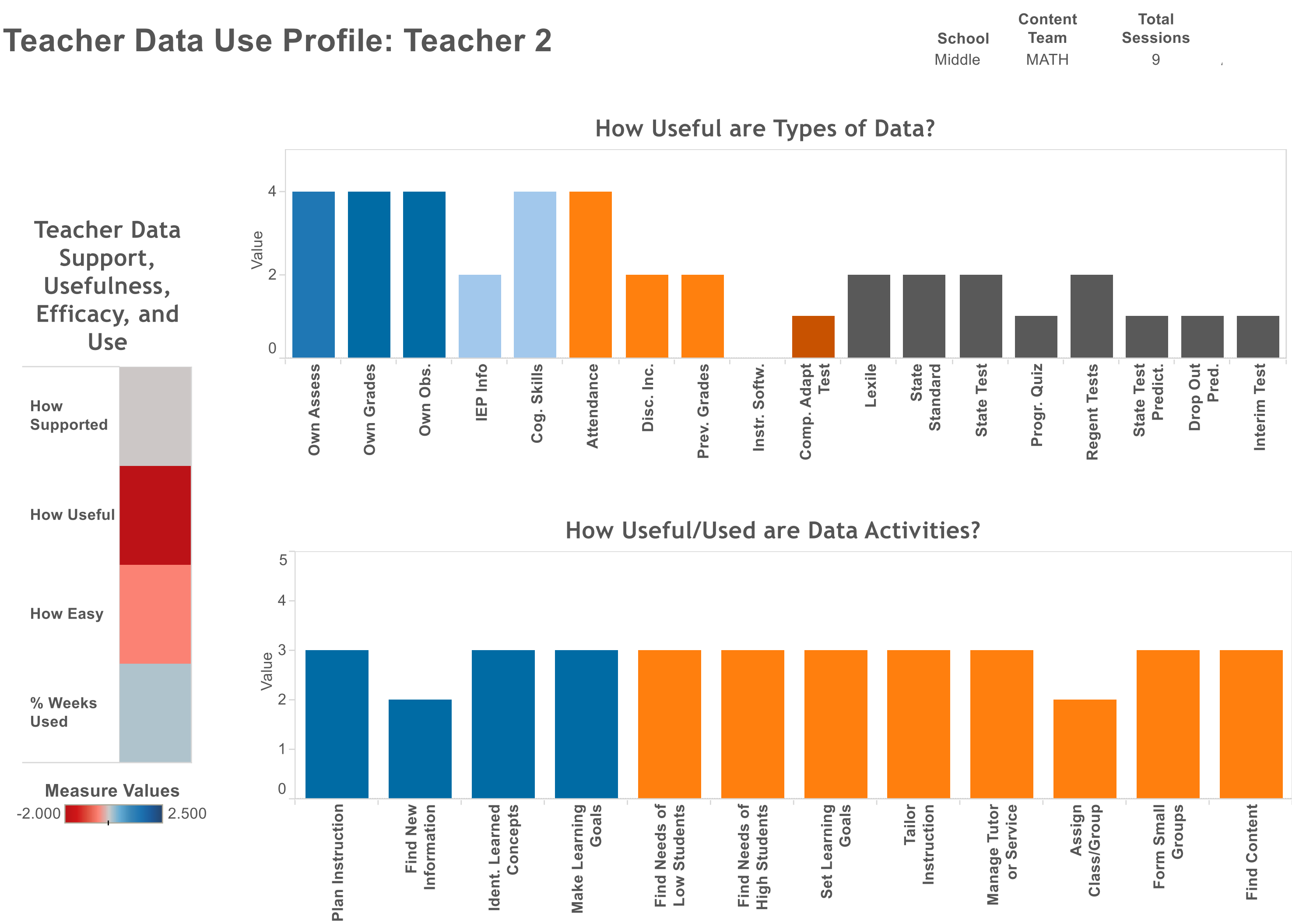


Figure 48. Teacher data use profile (Teacher 2)

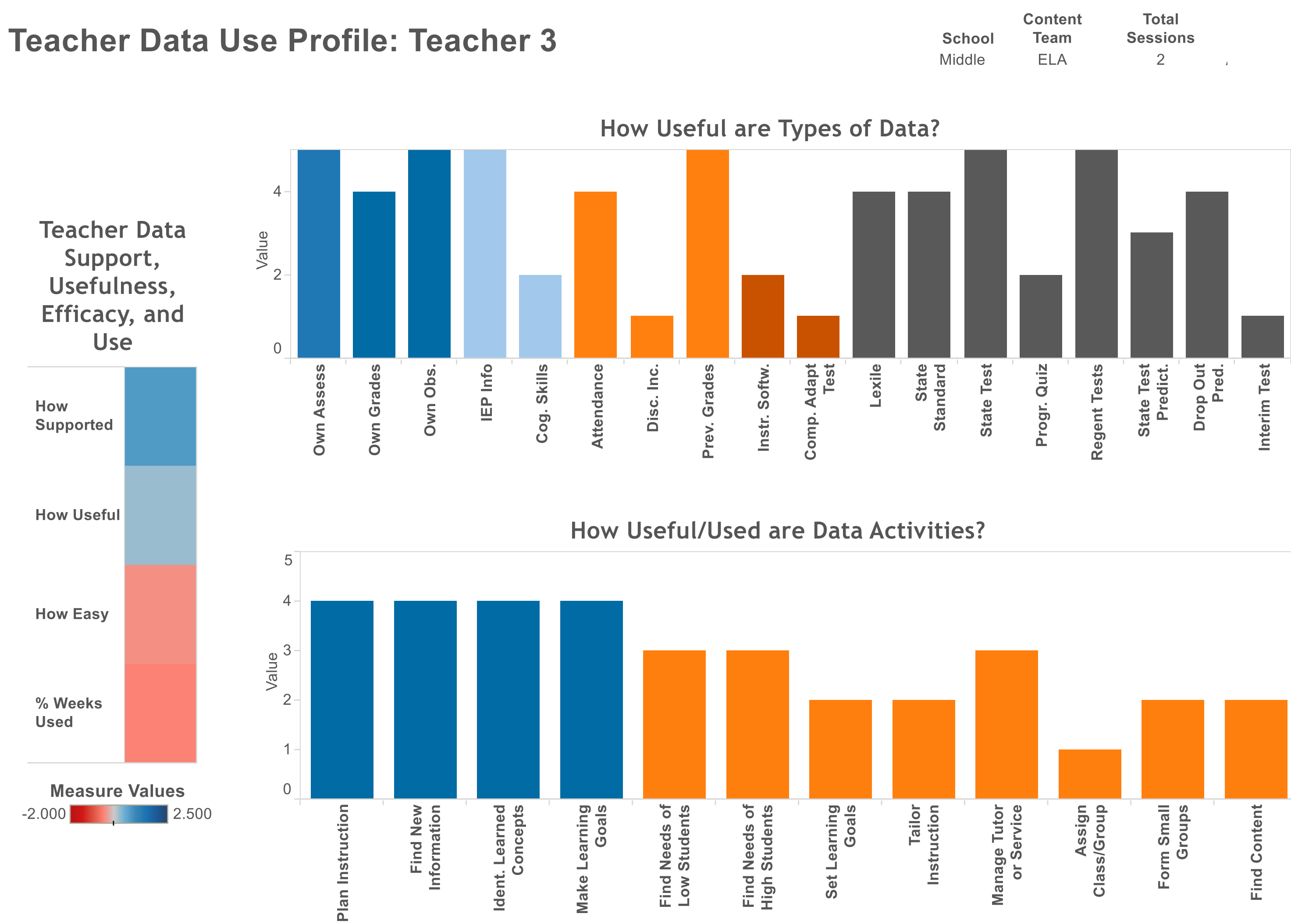


Figure 49. Teacher data use profile (Teacher 3)

Discussion

Study 2 tells the story, albeit a quantitative and empirical story, of one school's attitudes and interactions with online student data. In telling that story I hope to demonstrate how an expanded toolkit of methods and theory can inform central goals of DBDM research, such as improving determinant frameworks of teacher data use and expanding avenues for practitioners to participate in inquiry with instructional evidence. EDM/LA methods of log file analysis (Krumm et al., 2018; Rodrigo et al., 2012) and visual data analytics (Bienkowski et al., 2012; Bowers, 2010) support the search for patterns in teachers' access and response to student data, while theoretical frameworks of use diffusion (Shih and Venkatesh, 2004), self-efficacy (Bandura, 1977; Dunn et al., 2013a; 2013b), and the TAM (Davis, 1989); guide the interpretation of those patterns.

Overall, the application of these methods and theories provides evidence for clusters of data use attitudes, relationships to online data use, the impact of teachers' professional roles, and possibilities for improving determinant frameworks of online data use. This discussion is organized into four areas, regarding:

1. Relationships between data use attitudes, self-efficacies, and teachers' roles;
2. Subgroups of attitudes, self-efficacies, and technology acceptance in relation to teacher roles and online use of student data;
3. Possibilities for organizing teacher roles, self-efficacies, and technology acceptance into a determinant framework for online use of student data; and
4. Possibilities for guiding school practice in ways that improve teachers' use of student data for instructional decisions.

Teachers and schools approach the use of data and evidence based on complex attitudes and organizational constraints, and no data set or analytics can provide the whole picture of student success or teacher impact. Study 2 is an attempt to disentangle and prioritize some of these factors and should in no way be construed as an evaluation or judgment of teacher behavior or performance. If anything, this study suggests that the data-driven systems created to increase teachers' responsiveness to students, could be much more sensitive and responsive to teachers' own needs in the classroom, their attitudes towards data, their hard-earned pedagogical content knowledge, and their considerable organizational constraints.

Relationships between data use attitudes, self-efficacies, and teacher roles. The results of Study 1 suggest foundational, individual differences in data use attitudes and efficacies, along with complex relationships between these factors, school level, and content area. While focusing on attitudes and self-efficacy, as distinct from online use, this section discusses connections to past studies, the usefulness of data types, the usefulness of data use activities, possible clusters of data use attitudes and efficacies, and finally, organization-level relationships to school level and content area.

Comparing survey results to past studies. Survey results on data use attitudes from Study 2 are comparable to those from past studies. Previous administrations of the SEDU subscales related to teachers' attitudes towards data use found similar mean values to the current study. Wayman (2009a), for example, found an average and standard deviation of 3.41 (0.57) for Data's Effectiveness of Pedagogy and a lower 2.36 (0.79) mean for teachers' Instructional Uses of Data. Study 2 found similar means of 3.34 (0.75) and 2.50 (0.89) for the same measures. Both studies follow a pattern of higher scores for abstract belief in data and lower scores for specific instructional uses of data. Values for Support for Data Use were found to be slightly higher in the current study, with a mean of 3.01 (0.64) as opposed to 2.79 (0.66) in Wayman et al. (2009a). Table 37 summarizes results from Mesa Public Schools in Wayman et al. (2009a), as well as results from three school districts presented in Wayman, Cho, Jimerson and Spikes (2012).

Table 37

SEDU Results Comparison

	Study 2	Boyer	Gibson	Musial	Mesa
Attitudes Toward Data	3.20	3.05	3.13	3.12	NA
Data Effect. for Pedagogy	3.34	3.34	3.42	3.45	3.41
Computer Data Systems	3.11	3.11	3.07	2.94	NA
Instructional Uses of Data	2.50	NA	NA	NA	2.36
Support for Data Use	3.01	NA	NA	NA	2.79

Usefulness of data types. In terms of their usefulness to teachers, HCA clustered data types broadly into (a) teachers' own sources of data, including observations, assignments and assessments; (b) State Testing data; and (c) all other types of data (Figure 37). Line graphs presenting data types in order of preference add further context (Figure 33 and Figure 34), where teachers' own data received the highest average ranking, interim testing one of the lowest average scores, and state testing a ranking in between.

The results in Table 31 of values for perceived usefulness of data types produce an interesting instance where data from the same source is considered both more and less useful. Teachers were asked about the usefulness of both Lexile levels and computer adaptive testing. In this particular case, however, the administration of the same computer adaptive reading test generated both a Grade Level Equivalent (GLE) and a Lexile score of reading. A student might score, for instance, a 3.5 GLE in reading, indicating 3rd to 4th grade performance, and a 700 Lexile Measure (700L), both from the same testing administration. Though generated from the same computer test, teachers found the Lexile Level much more useful than the GLE score.

While reasons for this require further investigation, one possibility is that the close association of a Lexile score with a classroom decision (selecting texts at an appropriate level of difficulty) makes the usefulness of such a measure much more apparent. In contrast to the GLE score, the Lexile is both a representation of the students' reading ability and of text difficulty (Lennon and Burdick, 2004). This option for teachers, of interpreting the Lexile as a measure of texts and not of the students, may help avoid concerns around minimizing students to scores and numbers.

Usefulness of data activities. The HCA heatmap in Figure 38 presents interesting subgroups of teachers' perception, not of data types, but rather of data use activities. Two main clusters can be interpreted, one including beliefs about how useful data is for performing classroom-level activities, such as setting learning goals and planning instruction, and the other including beliefs about the validity of formal, statewide, and teacher-made tests. Separation into these two factors mirrors a typology of data use from Schildkamp et al. (2017) distinguishing *data use for instruction*, such as setting learning goals and tailoring instruction to student needs, from *data use for accountability* and *school improvement*, such as using assessment results for external reporting or internal evaluation. While not a perfect match, the HCA grouping of instructional and testing aspects of data use (Figure 38) provides additional support for this typology.

Possible clusters of data use attitudes and efficacies. Additional clusters of factors are apparent from Study 2 analyses of general attitudes towards data (SEDU), data use efficacy (3D-MEA), and general teaching efficacy (TSES/NTSES). Overall, general attitudes towards data use effectiveness and practice were highly intercorrelated, yet distinct from data use self-efficacies, which in turn were distinct from teacher self-efficacies, even teacher self-efficacy directed towards adapting instruction to students' needs (Table 34). A few exceptions to these general separations are apparent when interpreting both the correlation matrix in Table 34 and the clustering of factors in the HCA heatmap of SEDU scales in Figure 39. For example, self-efficacy for applying data to instruction is more highly correlated to several SEDU scales measuring data use practice and effectiveness (r = .68 - .84) than to other data use self-efficacy scales within the 3D-MEA (r = .37 - .58). Additionally, HCA clustering in SEDU scales suggests that perceptions of computer data systems and support for data use may form their own subgroup of perceptions distinct from other SEDU scales. Taking all these findings into account, Table 38 summarizes exploratory clusters of attitudes and efficacies from Study 2, clusters which may be useful for choosing factors for investigation in future DBDM studies.

Table 38

Summary of Factor Groupings for Data Use Attitudes and Efficacies

Category Topic	Survey Subscales
Data Systems and Support	Computer Data Systems (SEDU)
	Supports for Data Use (SEDU)
Self-Efficacy for Data Analysis	Efficacy for Data Identification and Access (3D-MEA)
	Efficacy for Analysis and Interpretation (3D-MEA)
	Efficacy for Data Technology Use (3D-MEA)
Attitudes towards Data and Data Use	Effectiveness for Pedagogy (SEDU)
	Data Use Practice (SEDU)
	Data Use Attitudes (SEDU)
	Efficacy for Applying Data to Instruction (3D-MEA)
Teacher Self-Efficacy	Efficacy for Instructional Strategies (TSES)
	Efficacy for Student Engagement (TSES)
	Efficacy for Adapting Instruction to Individual Students (NTSES)

These findings highlight the multiple attitudes and efficacies at play across the range of data use practices, from a general stance towards data, to school support, to specific efficacies for analysis, to general teaching skills. While sharing limited relationships in some cases, these categories seem to operate fairly independently, where higher values in any one of these areas fails to guarantee higher values in another. In other words, teachers' sense that data is useful may not relate to their self-efficacy for analyzing data, their practice of differentiation in the classroom, or, especially, their general confidence with instructional strategies or student engagement. With the lack of correlation between these categories, it seems important for schools and districts to separately consider organizational structures, training, and evaluation in regard to each cluster of factors when planning for implementation and evaluation of DBDM process.

Organizational impact: relationships to school level and content area. Along with insight into categories of data use attitudes, Study 2 offers evidence regarding the relationship of these attitudes to teachers' roles as content area specialists at a particular level of schooling. Relevant attitudes include perceptions of specific data types, as well as general data use attitudes and efficacies.

Study 2 suggests several differences in how often middle and high school teachers reported using data types and how useful they perceived them to be. Differences appeared between content areas as well, both when subject-area teachers were combined for middle and high school and when they were separated. Science and Social Studies teachers in particular followed very different profiles for perceived usefulness of specific types of data. One more general pattern worth further exploration was how teachers' reported frequency of use for data types was fairly similar across content areas (Figure 27), but different for middle and high school teachers (Figure 26). Teachers' perceived usefulness of data types, on the other hand, followed the opposite pattern: similar responses between middle and high school (Figure 33), but differing responses by content areas (Figure 34). Though further evidence is needed, this interesting contrast may indicate how differences in the frequency of use for particular data types are more related to differing expectations and requirements at different school levels, in this case middle and high school. Teachers' perceptions of the usefulness of data types, on the other hand, may be more closely tied to each teacher's needs based on curriculum and pedagogical content knowledge.

Several other attitudes and self-efficacies appear closely tied to school level and content area. Aside from the usefulness of data types, the usefulness of data activities appears related to school level, as middle school teachers tended to respond more favorably to data use across all SEDU scales, while high school teachers tended to cluster in a subgroup that assigned higher values to the validity of formal and state assessments than to classroom-based instructional uses of data (Figure 38). This high school preference for the validity of state testing may be due in part to the close relationship between high school state testing and course curriculum, as each high school state Regents exam assesses the accumulated content and skills from one year-long course. Other school-level differences included higher levels of perceived support, but lower self-efficacy for data use in the middle school (Figure 39, Figure 40).

In terms of content area differences, the most visible relationship was between lower usefulness of data types (Figure 34, Figure 37), lower usefulness of data activities (Figure 38), and lower values for data use support and attitudes (Figure 39), all for Social Studies teachers. Math teachers, in contrast indicated a higher usefulness for data use activities (Figure 38) and higher values for data use self-efficacy over general teaching self-efficacy (Figure 41). ELA teachers indicated a preference for instructional uses of data over the validity of state and formal testing (Figure 38).

While evidence from Study 2 does not support inferences about the cause of these differences, school level and content area have been found to be related to teachers' data use attitudes previously. Qualitative studies in particular, both large- and small-scale, have highlighted similar differences in how groups within organizations approach evidence (Coburn and Talbert, 2006; Horn, Kane and Wilson, 2015). Coburn and Talbert (2006), describe how the nature of work roles and the specific histories of instructional reform movements create pockets of teacher and leadership attitudes towards the validity of evidence about student learning and how that evidence should be used. For content areas in particular,

> Professional preparation programs provide role-specific and discipline-specific conceptions of valid evidence, professional associations embed meanings of evidence in their standards, and teacher associations promote views of evidence use consistent with practice in their professions. Thus, there are multiple and at times conflicting norms of evidence use that coexist in the environment of public schooling. (Coburn and Talbert, 2006, p. 490)

It seems likely that these ingrained differences in professional training and practice may extend to content teachers' perceptions of the usefulness of data types and activities and even to their self-efficacies for data analysis.

Datnow, Park, and Kennedy-Lewis (2013) also note that educational reform is mediated by content area, with different teams displaying different organizational cultures. The authors found similar content-area patterns in their qualitative study of data use in high schools, where the Math department demonstrated the furthest advancement in a data-driven curriculum. The authors suggest that this priority in Math may have been due to the tendency to view math "as a series of tasks or skills" that students learn and that teachers facilitate to mastery (Datnow et al., 2013, p.257). In other disciplines, particularly English and Social Studies, the relative importance of content-area skills and content was far less established (Datnow et al., 2013), impacting the departments' ability to structure DBDM efforts.

While on a smaller scale, Studies 1 and 2 may contribute to mapping the landscape of evidence-based practice and the differing perceptions of evidence underlying that practice. In particular, conceptualizing data use as specific to the intersection of school level and content area may provide a more solid foundation for future analysis of variation in teachers' data use and attitudes.

Connecting to online use: subgroups of attitudes, self-efficacies, and technology acceptance in relation to teacher roles and online use of student data. Now bringing in results related to online data use, this section first discusses correlations between attitudes and online use, then reviews school-level and content area relationships to online use, and finally discusses possible subgroups related to a wide range of factors: attitudes, efficacies, professional roles, and online use.

Correlations of attitudes and efficacies to online data use. As discussed more extensively in the results sections, moderate correlations were found between online usage and DDDM Self-Efficacy scales. For instance, the strength of the correlation between total sessions and Teacher's Efficacy for Interpretation of Data (r = .47), was second only to the correlation between total sessions and perceived usefulness of the Benchmark Data system itself (r = .50). Overall, teacher responses regarding the perceived usefulness of the Benchmark system and the usefulness of relevant types of data had the strongest relationships to online use, followed closely by the slightly more general data use self-efficacy scales of the 3D-MEA. Scales of abstract data use attitudes (SEDU) and measures of teaching self-efficacy showed weak to no correlation with usage outcomes. Even a survey scale related specifically to self-efficacy for adapting instruction to student need showed almost no correlation to online data access, while self-efficacy for student engagement demonstrated a moderate, negative correlation to total sessions. In the one comparable study, Shaw (2010) also found a slight negative correlation between teacher efficacy and online use (r = -.10).

Based on these exploratory findings, data use self-efficacy and technology acceptance factors, such as perceived ease of use and perceived usefulness of the data system and of data types, appear more closely related to online access of student data than more generalized judgements about the importance or effectiveness of data use in the classroom. While these findings may suggest possibilities for using the literature on technology acceptance to increase teachers' online access of student data, they should also raise concerns about the disconnects between teachers' online access, their strong desire to respond to students' needs, and their perceptions of their own efficacy as teachers. Even if successfully encouraged, increased online access may have little impact on evidence-based intervention in the classroom or student outcomes.

The usefulness of data types and interim testing. As noted above, teachers' perceived usefulness of data types demonstrated one of the relatively stronger relationships with online usage. Discussion of a few specific data types may provide useful insight into how teachers relate to these different sources of evidence. The definition and impact of formative and interim testing, for example, has provoked extensive discussion (Black and William, 1998; Dunn and Mulvenon, 2009; Filsecker and Kerres, 2009). Study 2 appears to highlight some of these same ongoing issues with interim testing as often practiced in relation to end-of-year standardized testing. In Study 2, teachers' perceived usefulness for interim testing was the lowest for all data types and was found, surprisingly, to be unrelated to teachers' online access to Benchmark Data, a system primarily used to distribute interim testing results. Figure 33 and Table 31.

Table 31 show strong agreement about the lower usefulness of standardized and formal assessments as compared to teacher-created assessments and observations. On the other hand, line plots in Figure 34, Figure 35, and Figure 36 show how opinions about data types can differ quite a bit by content area both within schools and between them. Math teachers, for instance, tended to rank the usefulness of data activities higher than other content areas and demonstrated higher actual use of the Benchmark Data system, a content-area trend mirrored in Wayman et al. (2017).

While one possible explanation for Math teachers' higher usage of online testing systems is their greater competency with quantitative reasoning, another possible explanation lies in the strong alignment for Math teachers between the items on interim assessments, state standards which categorize and rank the difficulty of these items, and classroom-practice centered around the frequent use of distinct math problems. For Math teachers, accessing online data on interim assessment may be particularly useful: knowing which items students got wrong allows for quick selection of similar items for classroom practice or homework, while knowing that an item matches a third or fourth grade standard gives Math teachers a general sense of pre-requisite skills and difficulty. Qualitative findings from Datnow et al. (2013) mentioned above may support this interpretation.

This level of alignment between interim testing, standards, and classroom practice is difficult to imagine in other content areas. In ELA, for example, classroom practice may focus on discussion, reading comprehension, and writing tasks on previously studied texts, not on multiple choice items and on-demand writing prompts about previously un-read texts. The interpretation of item performance according to ELA standards presents difficulty as well, when failure to answer an item correctly may stem from overall text difficulty of a passage and not from difficulty arising from the particular comprehension standard attached to an item. Social studies and Science teachers face their own difficulties translating students' interim test performance on a small sample of content knowledge items into more general classroom planning for content review. Student performance on items assessing a very specific element of content knowledge might serve as a useful, but rather limited jumping off point for adapting instruction. While making instructional decisions based on feedback of the standard alone, such as, "50% average performance on standard 9.2: Belief Systems, Rise and Impact," may have even less direct connection to instruction than feedback based on particular items. In disciplinary contexts where online data access is lower, the level or type of information available from interim testing may be simply out of step with the type of instructional decisions that teachers are able or interested in making.

These observations, while speculative, align with qualitative work on the role of pedagogical content knowledge in evidence-based practice. Multiple researchers have highlighted the role that teachers' pedagogical content knowledge (Shulman, 1986) plays in activating evidence for classroom use. Coburn, Honig and Stein (2009) and Coburn and Talbert (2006) have highlighted this need from a district-level perspective, while Mandinach (2012) approaches the conversion of data to instructional practice from a teacher training perspective, referring to the need for pedagogical data literacy. More recently, Horn et al. (2015) provides a framework for analysis of teachers' evidence-based practice, including teachers' epistemic stance, the representations of practice they employ, activity structures, and problem framing.

Such schema have great potential for qualitative work aimed at better mapping of discipline-specific practices related to evidence use.

Relationship of school level and content area to online data use. Both the results of Study 1 and previous work (Gold et al., 2012; Tyler, 2013; Wayman et al., 2009a; 2017) point to substantial differences in how different school levels and different content areas use online data. In fact, school level (Elementary vs. Junior High) is one of the only factors consistently significant across studies of online data use. Shaw and Wayman (2012) and Wayman et al. (2009; 2011) found teachers' online use significantly higher in the elementary grades (3^{rd} – 6^{th}), while Tyler (2013) found usage significantly higher for teachers of grades 6^{th}-8^{th}. Descriptive work by Gold et al. (2012) finds online use of the NYC ARIS system greater for middle schools than for both elementary and high schools. In contrast to Gold et al. (2012), Study 1 of this dissertation found substantially higher use of a testing and assessment system for high school than middle school (Table 13). While comparing across different types of data use systems and different school districts complicates these results, school level appears to play a key role in online data use.

Comparisons of online use between content area teachers are even fewer than those for school level. Wayman et al. (2017) found that junior high school Math teachers demonstrated roughly 1.6 times the prevalence of use of junior high school reading teachers and roughly twice the consistency of use. Study 1 found this same difference, if slightly more pronounced, with middle school Math teachers demonstrating more than twice the number of sessions and duration of use as middle school ELA teachers. Study 1's unique descriptions of high school usage, where ELA, Math, Science, and Social Studies teachers all faced state-level testing pressures, also found differences between content area teachers across multiple usage metrics (Table 18).

There are, however, alternative explanations for why the online use of one content area might exceed the usage of another. In one explanation, differences in online use might arise from the different perceptions of the usefulness that different content areas hold for the available student data. In another, the same differences might be explained by organizational factors. A principal with expertise in English might provide more individualized coaching for ELA teachers, accessing and synthesizing online assessment data on the teachers' behalf, while leaving Math teachers to access the data system on their own. Under these conditions, ELA teachers might be "accessing" online student data just as frequently through their principal as Math teachers, but their online use would appear much lower. In another possible scenario, different assistant principals overseeing different content areas might require their assigned teams to train and engage with student data differently. Such requirements would overwhelm the effects of any individual teachers' tendencies towards online data use. Identifying some of the most plausible of these possible explanations will be a fruitful area for future investigation.

Neither Gold et al. (2012) nor Tyler (2013) appear to address differences in how content area teachers access online systems. Further exploration of how school level and content area impact teachers' online data use are critical for shedding light on how teacher roles impact DBDM.

Identifying subgroups of teacher attitudes, efficacies, professional roles, and online use. Previous studies of teacher online data use have primarily suggested subgroups of teacher behavior based on overall frequency of use (Gold et al., 2012), but not on teachers' varying patterns of use or data use attitudes. Typologies of teachers' off-line data use have been proposed as well, based on qualitative analysis. Wayman et al., (2009a) describes three types of teachers: those completely opposed to data, those in favor of data as a supplement, and those who feel data is essential to their practice. Based on interview and survey data, the Gates foundation proposed a more complex typology of six types of teacher approaches to data and technology (Bill & Melinda Gates Foundation, 2015). More recently Graves and Bower (2018) use Latent Class Analysis of nationally-generalizable survey data to generate a four-class typology of technology-using teachers.

Review of subgroups from Study 1 based on online use. Study 1 found additional subgroups based on teachers' online behaviors when viewed through the lens of Use Diffusion theory (Shih and Venkatesh, 2004). Variety of online use was found to interact with frequency of use to produce overlapping, but distinct groupings of online behavior. For example, while many users from a cluster of high-frequency usage (Clust1 in Figure 21) overlap with a separate cluster of users who prioritize test administration functions (Clust1, Figure 22), other high frequency users clustered into groups that prioritized the creation and viewing of student reports (Clust2, Figure 22) or the viewing of test result views (Clust3, Figure 22). There is, in other words, more than one way to be a high-frequency user. Similarly, middle and low frequency users also prioritized different varieties of functions.

Subgroups of frequency/consistency, as well as variety, appear strongly related to organizational factors, particularly to content area, and possibly to specialized teacher roles. Overall usage itself also appears highly related to school-level training and testing schedules, as seen in usage timelines of Study 1 (Figure 18 and Figure 19). In future studies, basing inferences of teacher data use on overall frequency/consistency alone may overlook either important facets of teachers' variety of use or critical organizational factors.

While the use diffusion framework (Maull, 2013; Shih and Venkatesh, 2004) provided a useful starting place for the analysis of online use, the utility of the framework itself was rapidly replaced by the identification of specific subgroups of data use, each preferring access to a specialized mixture of system functions. It may be that a stricter interpretation of use diffusion, conceptualizing frequency and variety of use as single dimensions defining a four-quadrant typology, might have overlooked interesting subgroups of data use. In contrast, the clustering approach employed in Study 1 identified five possible subgroups of teachers' online data use related to (a) training-based assessment analysis, (b) assessment management, (c) student-centered analysis, (d) multiple measures analysis, and (e) intensive assessment management and analysis.

While Study 1 considers these five subgroups in relation to previous teacher typologies of data or technology use, also important to consider is how the same subgroups might relate to educators' goals in using data, whether their purpose is in using data for accountability, for school improvement, or for instruction (Schildkamp et al, 2017). In other words, how might these teacher groups defined by online access be using their online data? Without a more fine-grained analysis of teacher online usage their intentions remain largely opaque, but some tentative observations may be worthwhile.

For example, the largest identified subgroup of teachers—training-based assessment analysis—primarily accessed Benchmark data during training sessions designed to encourage the use of online reports presenting student-level interim testing data, such as summative test scores, correctness of items, and performance on standards. Teachers in this subgroup are focusing access on student data that could hypothetically be used to make a variety of instructional decisions, such as setting student learning goals, conducting item analysis, arranging small groups for focused instruction, or selecting standards for re-teaching. While the possibility for instructional use exists, it is unclear whether teachers in this category accessed the system intensively enough to translate the relevant information into specific lesson plans and, if they did so, whether these classroom interventions would be sufficient to substantially increase student learning.

Two other subcategories of teachers' online use, assessment management and intensive assessment management and analysis, define use in regard to managing assessments, as opposed to analyzing them. While such behaviors fall outside the bounds of online date use considered in previous studies (Wayman and Shaw, 2012; Tyler, 2013), some teachers appear to find value in assessment platforms functions for creating assessments or other classroom content. Unfortunately, the current level of analysis cannot determine whether teachers' use of assessment management functions is impacted by some preceding analysis of students' past test performance or is simply predetermined by the next unit of study. While not explicitly related to changes in daily instruction, online behaviors that integrate the analysis of previous assessments with the preparation of future assessments could indicate strong levels of adaptability and knowledge-seeking on the part of the classroom teacher. Data use in these cases would consist not just of analysis, but of the arrangement of inquiry to produce more reliable or actionable knowledge, a broader perspective that might encompass more stages of a data use cycle.

The final two subcategories of online use, student-centered analysis and multiple measures analysis may serve multiple purposes. Student-centered analysis, where the teacher views a profile of the available performance data for one student might serve an instructional purpose, such as creating individual learning goals across content areas, or, alternatively, a student-centered view might be leveraged in the accountability context of parent-teacher conferences (Schildkamp et al., 2017) or for a child-study team where teachers discuss one child's needs more holistically in relation to an individualized student plan that functions across classrooms. Such student-centered analysis might be less efficient, though, for making traditional classroom-level instructional decisions regarding student grouping or whole-class reteaching.

The final subcategory of use, multiple measures analysis, involves reports where individual students are represented in rows with their results in multiple assessments listed in the columns of the report. Such reporting might be easily applied to school improvement efforts, in an analysis of trends in students' performance across multiple assessments and classrooms. Another possible use for multiple measures analysis might be for assigning students to schoolwide programs for remediation or enrichment. This purpose, of selecting students for schoolwide programs seems to fall outside the category of data use for instruction. Though such an analysis has ramifications for students' instruction, it may not explicitly alter instruction in an existing classroom. Perhaps a more appropriate category for such data use might be that of data for analysis, proposed by Riehl, Earle, Nagarajan, Schwitzman, and Vernikoff (2018). Data for analysis describes the use of data to solve allocation problems, finding ways to "group and assign students to tracks, teachers, or supplementary services" (Riehl et al., 2018, p. 48).

Overall, the largest category of teacher online use—training-based assessment analysis—seems to fall, at least hypothetically, into the category of data use for instruction. The data use purposes of the other identified subgroups for assessment management, student-centered, and multiple measures analysis are less clear, suggesting some difficulty in aligning types of teacher data users with only one purpose of data use. The complexity and flexibility of roles and responsibilities in school settings may encourage user typologies that combine multiple purposes of data use as teachers aim for different analytical goals across their classroom, committee, or leadership responsibilities.

Additional subgroups based on role, attitude, and usage. While a subset of teacher attitudes and efficacy correlate moderately to online use, clusters based on these attitudinal factors do not map neatly onto online data usage. When clustered according to attitudes and efficacy, for example, high frequency users of Clust1 (Figure 21) are divided into multiple clusters and subclusters according to their differences in perceived usefulness (PERUSE) of data activities (Figure 38) and general data use attitudes (SEDU Scales, Figure 39). Considering HCA heatmap results for attitude and efficacy, the most effective factors for grouping high frequency users were those based on DDDM Self-Efficacy (Figure 40), perceived ease of use (PEOU) for the data system (Figure 42), and PEOU and PERUSE of data types (Figure 43). The last of these clusters, based on PEOU and PERUSE of data types, appears to be the most successful in tightly clustering high frequency users.

Table 39 attempts to summarize some of these complex relationships between data use factors, school level, content area, and online use. When viewed through the rich descriptive lens of Study 2, varied and conflicting motivations for teachers become apparent. ELA teachers may privilege classroom use of data over formal testing, but high school teachers may feel the opposite. Where, then, does a high school ELA teacher find value? Math teachers may have generally higher online use, but also feel less supported in data use and less effective in their general teaching practice. Why is a lack of support in this case related to higher online use? While higher PEOU and PERUSE appear strongly related to higher online use, PEOU and PERUSE are mixed for many teachers and appear to function within the context of an impactful set of organizational factors.

Middle school teachers, for instance, tended to express lower self-efficacy for data use, lower usefulness for data activities, and lower PEOU and PERUSE for the Benchmark system.

At the same time, however, they expressed more satisfaction with support and technology for data use. Middle school teachers also indicated frequent use of software and discipline data, neither of which were available in the Benchmark System (Table 39).

High school teachers were more supportive of formal testing and less supportive of classroom data use as compared to middle school teachers. They also tended to cluster around lower satisfaction with data use support and more negative general attitudes towards data use, while at the same time using the online system at higher rates than middle school teachers (Table 39).

Important to point out is the degree to which clusters of high frequency usage were unrelated to teachers' beliefs in the general importance of data use, the usefulness of classroom data use activities, and even beliefs in teachers' own teaching self-efficacy for adapting instruction to individual needs. While this disconnect makes explanations of online data use more manageable, it should caution schools and systems that successfully implementing systems for online data use is a far cry from implementing responsive interventions and successful classroom instruction.

Table 39

Summary of HCA Heatmap Analyses

Factors Clustered	HCA Column Clusters	HCA Row Clusters	School-Level Relationships	Content Area Relationships	Online Use
Reported Frequency of Use for Data Types (Figure 31)	>Software/Discipline >Testing/IEP Goals	Lower Frequency		Soc Stud +	
		Software/Discipline +	Middle +	ELA - / Math +	Lower Use
		Discipline/IEP +	High +	ELA +	
		Higher Frequency All		Soc Stud –	Higher Use
Perceived Usefulness of Data Types (Figure 37)	>Own Data >State Testing >Other Data Types	Lower Usefulness		Soc Stud +	
		Mixed Usefulness			
		Higher Usefulness	High +		
Perceived Usefulness of Data Activities (Figure 38)	>Classroom Use >Formal Testing	Lower Usefulness	Middle +		
		Classroom Use - / Formal Tests +	High +		
		Classroom Use + / Formal Tests -		ELA +	
		Higher Usefulness	Middle +	Math + / Soc Stud -	
General Data Attitudes (SEDU, Figure 39)	>Computers/Support >Attitudes	Support - / General –	High +	Soc Stud +	
		Support + / General –	Middle +		
		Support - / General +		Math +	Higher Use
		Support + / General +	Middle +		
Data Use Self-Efficacy (3D-MEA, Figure 40)		Low Efficacy	Middle +	Science -	
		High Efficacy			Higher Use
Data Use and Teaching Self-Efficacy (3D-MEA, TSES, Figure 41)	>Data Use Efficacy >Teaching Efficacy	Teaching + / Data use –			
		Teaching - / Data use +		Math + / ELA –	
		Teaching + / Data use +			Higher Use
PEOU for Benchmark Data (Figure 42)		Lower PEOU		Soc Stud +	
		Higher PEOU			Higher Use
PEOU/PERUSE for Benchmark Data (Figure 43)	>PEOU >PERUSE of Data Types	Lower PEOU & PERUSE	Middle +	Science –	
		Mixed PEOU & PERUSE		ELA +	
		Higher PEOU & PERUSE		Soc Stud –	Higher Use

Possibilities for organizing teacher roles, self-efficacies, and technology acceptance into a determinant framework for online use of student data. Correlational and HCA evidence suggests that DDDM Self-Efficacy and TAM factors of PEOU and PERUSE would make a strong contribution to existing determinant frameworks for DBDM, particularly in regard to online data use. The two TAM (Davis, 1986) factors were more suggestive of online use than were attitudes for the general effectiveness of data and were potentially more effective in clustering high online usage when used together than when considered separately.

Figure 50, following, lays out a simple diagram within the structure of the TAM, including factors relevant to Studies 1 and 2. While a full determinant framework would include a wider range of organizational and other factors, the framework presented here draws only on factors and findings from the current studies. To expand the outcomes of online use, Figure 50 collects results from Study 1 regarding usage metrics (Table 22) and typologies of user behavior (Table 23). To identify and categorize determinants of teachers' online data use, the same figure draws from summary Table 38 and Table 39, as well as correlations from Table 35 and Table 36. As in Study 2, PEOU includes elements of DDDM Self-Efficacy and system-specific ease of use. The factor for PERUSE, however, has been expanded beyond its basic application in Study 2 to include some of the organizational factors that may impact the perceived usefulness of data, as well as self-efficacy for applying data to instruction, a survey subscale which consistently clustered with factors related to perceived usefulness in HCA results. Facilitating conditions identified in Figure 50 relate to findings from Study 1 about the strong relationship between professional development and usage (Figure 19). Though Teacher Efficacy had almost no correlation to online use, it is included in the schema based on HCA results indicating a possible relationship between higher usage for teachers and the combined presence of higher teaching efficacy and DDDM efficacy.

Online usage outcomes, such as session- and action-level patterns, found relevant in Study 1 indicate that online use may differ in important ways beyond overall frequency and consistency. In fact, it is important to keep in mind that frequency itself is not a reliable proxy for more "successful" usage. For some users, more time spent in an online system may indicate uncertainty or inexperience rather than more efficient and successful use. Evaluation of "successful" usage outcomes may also differ according to some of the same school or content-area factors impacting the perceived usefulness of data and data systems. Successful online use may look markedly and justifiably different for a high school special education English teacher assigned to monitor students' Response to Intervention (RTI) progress than for a middle school Math teacher with and no additional responsibilities. A valid model of teachers' online data use should be able to incorporate a range of successful patterns, as opposed to implicitly judging success in terms of increasing frequency of use.

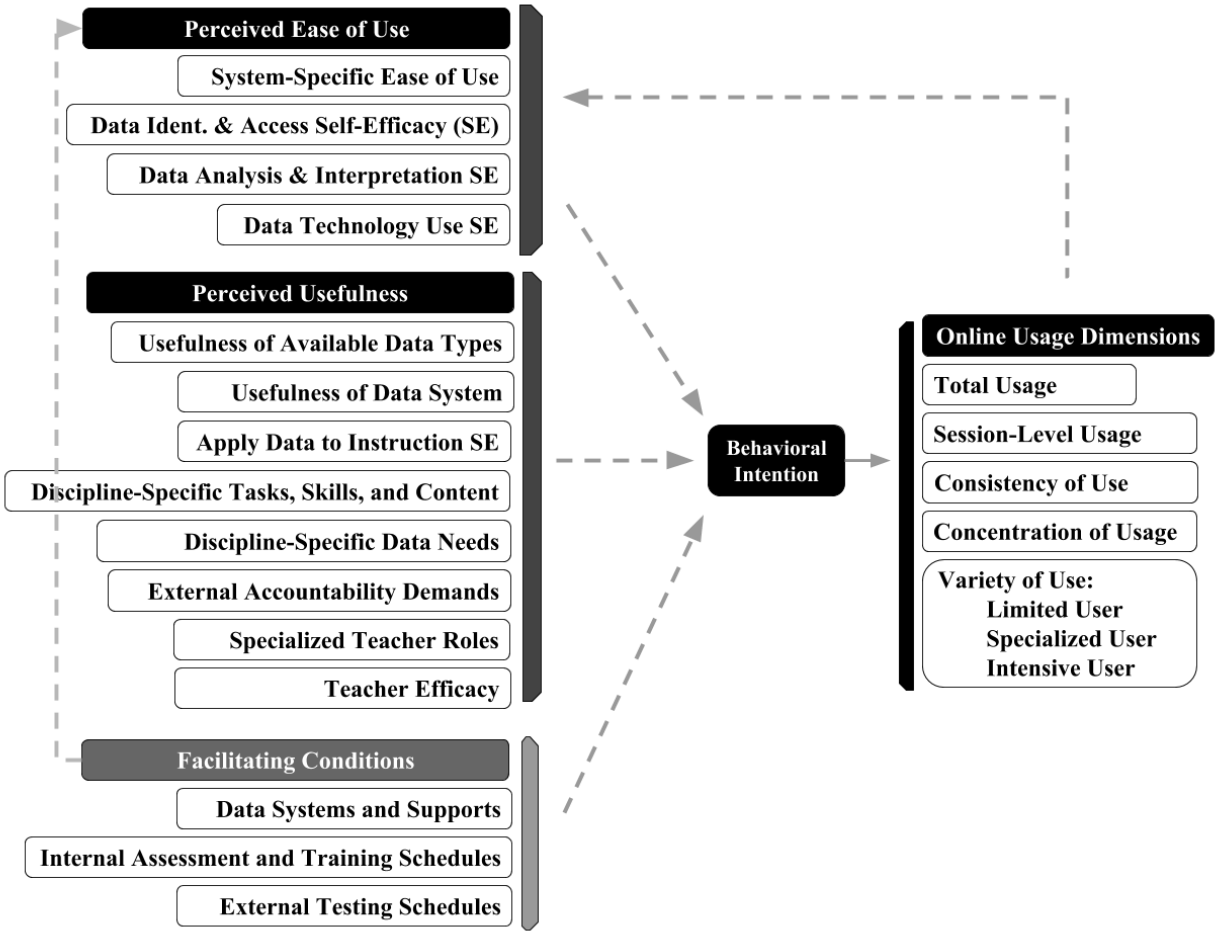


Figure 50: Summary of exploratory relationships between PEOU, PERUSE, and online usage *Note.* Since Study 2 failed to test any causal relationships or pathways, dotted lines are used to show proposed relationships between factors.

On the one hand, these results and discussion indicate the possible contribution that relevant theory and an exploratory, data-intensive approach can make to further understanding of teachers' online use of data. On the other hand, recognizing the importance of these technology acceptance and self-efficacy factors to teachers' data access only opens the door to more questions about how teachers' perceptions of ease of use and usefulness relate to local data systems and how these personal factors interact with the organizational factors that have played such a large role in these results. While these studies cannot provide clarity around the relative impact of individual versus organizational factors in online use, they do highlight a case where large variation in usage persists among users in the same school, responding to the same organizational structures. That such variation persists under similar organizational constraints serves to emphasize the importance of individual-level factors in data use decisions, as well as the need for more research into the relationships between technology acceptance, self-efficacy, and ubiquitous organizational factors, such as school level, content area, training, assessment schedules, and specialized responsibilities.

Possibilities for guiding school practice in ways that improve teachers' use of student data for instructional decisions. Studies 1 and 2 offer a range of insights and tools for the use of student data in local school contexts. Some of these insights apply specifically to the local context of Progress Secondary, while others may have value more broadly.

Implications for data use at Progress Secondary. Dashboards created for school leaders (Figure 45 and Figure 46) suggest several areas for inquiry and intervention. Of biggest relevance to the Benchmark Data system are the low rankings for the usefulness of interim and standardized testing across all content areas. Teacher leaders should engage in structured exploratory conversations with content teams, examining why interim testing is perceived as less useful and whether (or not) it can be made more useful. At the same time, the leadership team should re-examine the goals of interim testing and consider whether those goals might be met through the use of types of data that teachers consider more useful, such as teacher-made assessments, gradebooks, or instructional software.

Aside from these schoolwide concerns with interim testing, the middle school's lower online use of Benchmark Data and lower data use self-efficacy for ELA, Social Studies, and Science content teams (Figure 45) should prompt a dialogue centered around the data priorities of these teams, their content-specific decisions and needs for evidence. While the middle school Math team demonstrates overall higher data use efficacy and online use, several teachers indicate lower levels of support and usefulness of data. These needs should be examined with the team as well.

While high school teachers demonstrated higher levels of online use and generally higher data use efficacy, they also indicated the same concerns as middle school about interim testing (Figure 46), prompting a similar inquiry into how interim testing might be made more useful or whether its goals might be fulfilled through other sources of data. As opposed to the middle school, where perceptions of usefulness and efficacy appear to be the most pressing need, a general lack of support seems to characterize teacher attitudes in the high school. In particular, while both the ELA and the Social Studies team demonstrated much lower online data use, the needs of the ELA team appear to be more focused on a lack of support, while Social Studies team appears reticent about data use in terms of support, usefulness, and ease of use. As a starting place for better understanding the needs of the Social Studies team, it might be valuable to first talk with the one member of the team who demonstrates higher data use attitudes and efficacies in order to get a content-specific perspective on the team's needs and possibly to recruit that teacher for a leadership or mentoring role around data use in the Social Studies context.

Along with school-level and content team differences, Study 1 also identified several subgroups of specialized use within the Benchmark Data system: student-centered analysis, multiple measures analysis, and assessment management. Through conversations with these users, school leadership may want to identify the specific tasks and decisions they are seeking to inform and then assign internal, district, or vendor staff to iteratively design reports tailored to any regularly occurring evidence-based decisions.

General implications for evidence use in schools. The line graphs, HCA heatmaps, and data dashboards of Study 2 all offer the possibility for schools to analyze data use and attitudes across schools and content areas. Such profiles efficiently allow for the diagnosis of specific data use issues, as demonstrated in the section above. They also present a bird's eye view of overall program coherence and alignment, allowing leaders to quickly assess whether teachers' data use and perceptions of usefulness match leadership's prescriptive expectations. For example, if teachers are expected to use Lexile Levels when assigning texts to low-level readers, do visualizations suggest this use of data is occurring? If leadership perceives interim testing as useful, but teachers do not, where is the disconnect? Line graphs in Study 2 are also useful for assessing the degree to which teachers' perceptions align across content areas or other subgroups. Of course, content areas may feel differently about data types and activities for important and legitimate reasons. Acknowledging and encouraging discussion around these differences is part of the function of these visualizations.

Implications for teacher training. The literature of data-based decision making has had a substantial focus on how teacher training intersects with data and assessment literacy (Jimerson and Wayman, 2015; Mandinach, 2012; Mandinach and Gummer, 2016). Such work has made suggestions related to teacher training that align with observations from Study 1 and Study 2, particularly in regard to the critical importance of defining data broadly and of including pedagogical content knowledge as a core component of training in evidence-based practice (Mandinach 2012; Mandinach and Gummer, 2016). At the same time, Studies 1 and 2 raise some possible limitations to the focus on teacher education as a means of leveraging evidence-use for improvement in schools. Specifically, in Figure 39, Study 2 suggests that possible sub-clusters of teachers may feel supported in their use of data and confident in computer systems for using data, yet still have less than positive attitudes towards data use and feel more negatively about data use practice and effectiveness. Many teachers, despite scoring themselves higher on DDDM self-efficacy factors and ease of system use (Figure 42), still show low usage of online data, allowing organizational training requirements to determine the majority of their access to online testing data, often at the minimum level required. In these cases, more and better training may not be the solution.

Stated in terms of the TAM and teacher factors in Figure 50, increased training may partially address some of the self-efficacy factors related to perceived ease of use, but may be much less likely to impact factors related to the perceived usefulness of data systems and data types, especially in cases where teachers are unable to impact existing school systems or where convincing, successful examples of content-specific evidence-use are unavailable. Many organizational factors impacting teacher data use are outside of teachers' control, from the constraints of training, time, and protocols, to the management of external accountability pressures, to the development or purchase of online testing and data systems, to curricular decision-making and the implementation of schoolwide intervention structures. With these significant organizational impacts, any efforts at enhancing teacher data use through teacher education alone—without extensive efforts at understanding and improving evidence use at an organizational level—may see limited success. First understanding teachers' own role-based logics, representations, and practice of evidence-use and later, or simultaneously, designing or adapting systems to support these existing, practice-based systems may provide a more productive path for enhancing the usefulness, usability, adoption, and ultimately the effectiveness of evidence-use for student learning.

Overall, the results of Study 1 and Study 2 suggest that to facilitate teachers' engagement in DBDM, schools should shift from accountability-driven, "one-size fits all" data use to data use that engages a wider range of evidence and is differentiated by the professional structures of schools: less data-driven decision making, based on accountability-based testing, and more decision-driven data making based on the content- and context-specific needs of teachers.

Implications for data use software designers and companies. Along with suggestions related to the development of recommender systems discussed in Study 1, evidence from Study 2 may encourage system designers to prioritize development of methods to facilitate teachers' ease of use within the system, along with customization by teacher roles and responsibilities. Software designers may also decide to support embedded decision-making functions within data use systems to make the usefulness of data more relevant to users. However, findings from both studies suggest that a major priority for any data use system is the tailoring of the system to the particular roles, structures, and needs of the school, many of which may change rapidly and on an ongoing basis.

While one day data use systems may be intelligent enough and/or data science expertise available enough to satisfy the needs of schools, in the meantime, data system vendors might efficiently satisfy a range of school data use needs by providing ongoing, remote support from educational data scientists (Agasisti and Bowers, 2017; Piety, Hickey, and Bishop, 2014). While traditionally software vendors provide ongoing training in the use of their particular product, school's may benefit more from the ongoing support of an education professional who can knowledgably interface with a school about their structures and processes, while also efficiently building school-specific reporting capabilities or collaborating with software designers on the development of data-intensive or decision-targeted reports. Training in system use under these conditions becomes a natural extension of the needs of the school, as opposed to time spent learning to use yet another online system. While roles for educational data scientists might also be housed within districts (Agasisti and Bowers, 2017), independent schools or small or rural districts may struggle to hire the appropriate expertise. Educational data scientists engaged through software vendors would be positioned to anticipate school needs through review of their online practice (with appropriate permissions in place), inform their organizational structures and process, and rapidly leverage reporting capabilities developed across a network of schools. Given the complexities of local school structures and dynamics, the most impactful solutions for data use companies to integrate and maximize the use of their online tools may not be a software design decision, but rather the creation of new roles for supporting schools and efficiently communicating feedback between schools and a software provider.

The information market system: a shift in thinking about evidence-based decisions in schools. While admittedly beyond the scope of these results, this section reflects on possible theoretical frameworks that might better represent the complex nature of how schools use evidence for decisions. Supporting work by Piety (2013), the heterogeneity of results in these studies suggest that neither a technical view of evidence use, advocating for more data and better training, nor a social view of evidence, focusing on educators' perceptions and beliefs, sufficiently advances understanding of data use in schools. Particularly, the proposed typology of online data use (Figure 24 and Figure 25) based in both frequency and *variety* of teacher use suggests that teachers are searching for value in data systems in ways unanticipated by a monolithic understanding of "data," data literacy, or generalized data use attitudes. The framework that Piety (2013) suggests, and these studies support, is that of the marketplace, a marketplace of small, discrete decisions about educational information and its uses. The marketplace view acknowledges the complexity of the information system while giving a central place to the question of how users perceive value and make specific choices for information use. The marketplace metaphor readily handles the complexity of how different educational roles intersect with types of data and with goals for analysis. Sometimes these intersections occur in unexpected ways. Principals may find value in fine-grained gradebook data for improving instruction, while teachers may find valuable uses for state testing data to inform school improvement. Under a marketplace view these data are not good or bad but have "value for specific individuals and purposes" (Piety, 2013, The Value Exchange, paragraph 2**).**

> [At] every instance of information use, an exchange is made; value is perceived. That exchange, in many cases, may involve simply spending time in one way or with one set of tools versus another. . . When practitioners find value in the data for their teaching or managing, they will likely repeat the exchange. (Piety, 2013, The Value Exchange, paragraph 1)

Such market-based systems are already playing a growing role alongside district-based programs. A recent study finds that 97.6% of instructional software licenses purchased through districts are not used intensively (Baker and Gowda, 2018). Yet, other online products marketed directly to teachers, such as ClassDojo (Chaykowski, 2017), Kahoot!, or Gradecam appear to find widespread use. Such educational applications, along with online communities such as Instagram and teacherspayteachers.com are upending district and school hierarchies by appealing directly to teachers' sense of value. Such applications may be more successful than district-administered programs in tapping the potential of teachers' persistent attempts at sensemaking in the classroom (Riehl et al, 2018), attempts that are often distanced from analysis of standardized testing.

While not explicitly advocating a marketplace framework, Coburn and Talbert (2006) describe how in response to district level pressures for evidence in the form of standardized testing, reforms should foster a “system of evidence use that allows for and supports access to different kinds of evidence for different purposes at different levels of the system.” Creating this system involves “going beyond the sole use of standardized test scores to collect and make accessible to educators a broader range of data capable of answering different kinds of questions that people in different roles face in the course of their ongoing work.” (Coburn and Talbert, 2006, p. 491). While not as fine-grained as considering each data use transaction as a separate unit, Coburn and Talbert (2006) recognize the need for “different evidence” for “different purposes” at “different levels” within an educational system, suggesting a strong flexibility in defining value even for the same user, as they engage in different purposes for data use.

As exploratory visualizations, timelines of log file analysis in Study 1 and HCA Heatmaps from Studies 1 and 2 are well suited for identifying pockets of value and subgroups of usage behavior, where teachers’ exchange with the online data use system or another form of data has been “profitable”. In these cases, information needs are inferred from the actions individuals take to seek out and use information, actions which may be heavily influenced by job roles, ad hoc responsibilities, and organizational context. Visualizations of the search for information, such as those used in Studies 1 and 2, may be particularly useful when it comes to establishing segments of value to teachers or other users.

While the ramifications of this switch in perspective are far from clear, one impact would be to consider both the technological data systems that schools employ, along with their collaborative and organizational structures, all as part of one market system, the infrastructure of which, when well-designed, supports each educator in searching for the information they need in that moment and for that goal.

Limitations, Contributions, and Future Research

Several limitations of this work have been discussed throughout and a few will be briefly returned to here. This study examines evidence from a small number of teachers, in one school, over one semester. Given these limitations, any inferences or observations are tentative at best and cannot be considered broadly applicable. The Benchmark Data system was analyzed in its first year of adoption, so any identified patterns of use may be different from the stable use of experienced users. Survey measures were administered only once and at the end of the school semester, giving no indication of how attitudes might have changed over the course of the semester or in relation to use of the online system. Finally, even one school is a complicated place to study, and there is no doubt that many additional and confounding factors impacted either teachers' access to data or their attitudes about it over the course of the study.

Despite these limitations, this work makes contributions to the study of DBDM. In its analysis of log files, Study 1 expands the range of usage indicators relevant to teachers' use of online data and generates novel ways for visualizing and contextualizing teachers' usage of such systems over time. In exploring rich relationships between these online traces of evidence-use behavior and multiple facets of teachers' data use attitudes and efficacy, these studies expand the limited quantitative descriptions of teachers' online behaviors and make previously unmade, if tentative connections between online data use, DDDM Self-Efficacy, and the TAM. Analysis and clustering of teacher attitudes and online use both support the findings of previous qualitative scholarship on teacher data-based decision making and generate examples of possible tools for use in educational leadership decisions. By relating their descriptions to theory in self-efficacy, technology acceptance, and use diffusion, these studies allow for the increased impact of these and other theories into the inquiry and applications of data-based decision making and evidence-use.

Based on Studies 1 and 2, several strands of future inquiry into teachers' evidence-based practice can be envisioned as well. Building on the "market system" framework from the previous section, these suggestions for future inquiry attempt to capture, better understand, or crystallize the varying transactions in which educators find value for evidence. Complex relationships between individuals and organizations, only hinted at in this work, make clear that several levels of analysis, along with multiple methods of inquiry, may prove the most productive for future research. Suggestions for future studies are organized from the smallest scale to the largest.

Start to Finish Evidence-Use.

On the smallest scale, one potentially productive study would combine work by two researchers: a previously mentioned study by Horn et al. (2015) that implements a framework for analyzing teachers' logic, and practice for evidence-use and studies by Xhakaj, Aleven, and McLaren (2016; 2017) on teachers' use of student data from an intelligent tutoring system. Xhakaj et al. (2017) track the uptake and use of evidence from start to finish, from prior to the teachers' exposure to the eventual use of evidence in the classroom. At each stage they calculate the attrition of possible insights into student learning, from the evidence and insights that make it into a lesson plan, to the insights that result in classroom action. Capturing the same stages of use, but for a wider range of classroom uses of evidence might prove valuable in pinpointing the specific steps and barriers between receiving and using evidence. Horn et al.'s (2015) framework for analyzing teachers' logics and opportunities for evidence use—including teachers' epistemic stance, their representations of practice, the structure of the evidence-use activity, and the framing of the problem—would complement these analyses by providing factors for analyzing teachers' conceptions of value for different evidence use transactions.

Storyboarding Evidence-Use Practice.

A second line of inquiry, at a slightly larger scale, involves the need for better describing and disseminating teachers' existing, discipline-specific and individualized approaches to evidence use, instances where they currently find value for a wide range of data. While additional studies observing and recording teacher practice may be useful, there may be ways to leverage and amplify existing troves of qualitative data as both a practical resource for teachers and as a tool for deepening design-based research with schools. For example, the Spencer Foundation recently funded extensive qualitative work studying teachers with expertise in data-use, as a means to identifying promising evidence-use practices already embedded in teachers' everyday instruction (Barnes and Fives, 2018). While each separate study in Barnes and Fives (2018) makes useful points and contextualizes useful examples of evidence-use, perhaps specific anecdotal examples of teacher data-use could be extracted and summarized in a format commonly used in user-centered design: the storyboard (Aleven, Xhakaj, Holstein and McLaren, 2016; Kalbach, 2016).

Storyboards present a series of panels visualizing a process or scenario accompanied by short descriptions of each image. While the primary goal of storyboarding would be to capture the key elements of teachers' practice-based evidence use, it might be possible to again incorporate Horn et al.'s (2015) framework by referring where possible to teachers' epistemic stances, their representations of evidence, the structure of their evidence-use activity, and their framing of the problem, each within a separate frame of the storyboard, creating a structure that would allow storyboards to be easily sorted, arranged, or varied according to the factors of the framework. Capturing evidence-use scenarios in storyboards and cataloging them along dimensions from Horn et al. (2015) and/or others would serve several functions:

- Facilitate efficient understanding, use, and online sharing by teachers of practical evidence-use strategies.
- Provide a valuable and accessible resource for researchers to share and use in the context of professional development with schools, or for studies eliciting teacher feedback on evidence-use strategies, or more generally, in design-based implementation research (Fishman, Penuel, Allen, Cheng and Sabelli, 2013) and research-practitioner partnerships (Coburn, Penuel and Geil, 2013, Krumm et al., 2018).
- Generate content to integrate with existing or future online data use platforms to provide additional support for teachers' decision-making around evidence-use.

Recommender Systems for Evidence Use

Finally, as described in Study 1, much additional work can be accomplished working at scale with large online systems that either allow access to student data or provide decision support for evidence-based decisions. Data-intensive methods from the Learning Analytics community, web analytics, and log file analysis can help identify patterns of use and amplify the usability and effectiveness of such systems by piloting recommender systems to increase the capabilities of such systems to meet the needs of educators in different roles, whether leadership, content area, school level, or specialized teacher responsibilities. The development of simple recommender systems for evidence-use systems would also serve the purpose of continuing to explore undiscovered areas of value for evidence use.

Another related need for integration and testing of data use systems is with schools' existing classroom-based formative testing, response to intervention processes, and special education data systems. A more seamless interface for data access, intervention, and progress monitoring would be a valuable step towards more fully-fledged decision support systems, similar to those used in medicine to support clinical decision making (Alther and Ready, 2015).

Taking full advantage of new frameworks and methods for inquiry in learning analytics and educational data science, as well as the integration of theoretical perspectives from evaluation, implementation science, medicine, behavioral economics, and decision science would be important steps towards the future investigation of evidence-use in schools.

References

Abbott, D. V., Mandinach, E., and Honey, M. (2008). A functionality framework for educational organizations: Achieving accountability at scale. *Data Driven School Improvement: Linking Data and Learning*, 257–276.

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, 16(1), 3-9

Agasisti, T., Bowers, A.J. (2017) Data Analytics and Decision-Making in Education: Towards the Educational Data Scientist as a Key Actor in Schools and Higher Education Institutions. In Johnes, G., Johnes, J., Agasisti, T., López-Torres, L. (Eds.) Handbook of Contemporary Education Economics (184-210). Cheltenham, UK: Edward Elgar Publishing.

Aleven, V., Xhakaj, F., Holstein, K., and McLaren, B. M. (2016). Developing a Teacher Dashboard for Use with Intelligent Tutoring Systems. In *IWTA@ EC-TEL* (pp. 15-23).

Allensworth, E. M. (2013). The Use of Ninth-Grade Early Warning Indicators to Improve Chicago Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, *18*(1), 68–83.

American Recovery and Reinvestment Act of 2009: Title I, Part A Funds for Grants to Local Education Agencies -- 2009

Anderson, R. N., Greene, M. L., and Loewen, P. S. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student achievement. *Alberta Journal of Educational Research*.

Anderson, S., Leithwood, K., and Strauss, T. (2010). Leading Data Use in Schools: Organizational Conditions and Practices at the School and District Levels. *Leadership and Policy in Schools*, *9*(3), 292–327.

Anderson, T., and Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational researcher*, *41*(1), 16-25.

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., and Zellman, G. L. (1976). *Analysis of the school preferred reading programs in selected Los Angeles minority schools*.

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly*, *33*(1).

Arnold, K. E., and Pistilli, M. D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267–270). ACM.

Ashton, P. T., and Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. Longman Publishing Group.

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, M. M. T., and Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241.

Baker, R. S., Corbett, A. T., Roll, I., and Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, *18*(3), 287–314.

Baker, R. S. and Gowda, S. (2018). The 2018 Technology & Learning Insights Report: Towards Understanding App Effectiveness and Cost. BrightBytes.

Baker, R. S., Gowda, S., Wixon, M., Kalka, J., Wagner, A., Salvi, A., Aleven, V., Kusbit, G. W., Ocumpaugh, J., Rossi, L. (2012). Sensor-free automated detection of affect in a Cognitive Tutor for Algebra. In *International Educational Data Mining Society 2012*.

Baker, R. S., and Siemens, G. (2014). Educational Data Mining and Learning Analytics. In R. Keith Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences (Cambridge Handbooks in Psychology)* (2 edition, pp. 253–272). Cambridge University Press.

Baker, R. S., and Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, *1*(1), 3–17.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall, Inc.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: Freeman.

Barnes, N., and Fives, H. (Eds.). (2018). *Cases of Teachers' Data Use*. Routledge.

Beal, C. R., Qu, L., and Lee, H. (2006, July). Classifying learner engagement through integration of multiple data sources. In *Proceedings of the National Conference on Artificial Intelligence* (Vol. 21, No. 1, p. 151). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, *2*(2), 131.

Behrens, J. T., Piety, P., DiCerbo, K. E., and Mislevy, R. J. (2018). Inferential foundations for learning analytics in the digital ocean. *Learning Analytics in Education*, IAP.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *57*(1), 289–300.

Bertrand, M., and Marsh, J. A. (2015). Teachers' Sensemaking of Data and Implications for Equity. *American Educational Research Journal*.

Bienkowski, M., Feng, M., and Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, *1*, 1-57.

Bill & Melinda Gates Foundation (2015). Teachers know best: Making data work for teachers and students. Retrieved from https://files.eric.ed.gov/fulltext/ED557084.pdf.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, *83*(3), 377.

Bodily, R., and Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, *10*(4), 405-418.

Bowers, A. J. (2007). Grades and data driven decision making: Issues of variance and student patterns. Unpublished PhD thesis, Michigan State University, East Lansing, MI.

Bowers, A. J. (2010). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment Research and Evaluation*, *15*(7), 1–18.

Bowers, A.J. (2015-2017) Using Big Data to Investigate Longitudinal Education Outcomes through Visual Analytics. National Science Foundation, Directorate for Computer & Information Science & Engineering (CISE), Division of Information & Intelligent Systems (NSF IIS-1546653).

Bowers, A.J. (2017) Quantitative Research Methods Training in Education Leadership and Administration Preparation Programs as Disciplined Inquiry for Building School Improvement Capacity. *Journal of Research on Leadership Education,* 12(1), p. 72-96.

Bowers, A.J., Blitz, M., Modeste, M., Salisbury, J., Halverson, R. (2017) How Leaders Agree with Teachers in Schools on Measures of Leadership Practice: A Two-Level Latent Class Analysis of the Comprehensive Assessment of Leadership for Learning. *Teachers College Record,* 119(4). http://www.tcrecord.org/Content.asp?ContentId=21677

Bowers, A. J., Krumm, A., Feng, M., and Podkul, T. (2016). Building a Data Analytics Partnership to Inform School Leadership Evidence-Based Improvement Cycles. *AERA Roundtable*.

Bowers, A. J., Shoho, A. R., and Barnett, B. G. (2014). Considering the Use of Data by School Leaders for Decision Making: An Introduction. *Considering the Use of Data by School Leaders for Decision Making*, 1–16.

Bowers, A. J., Sprott, R., and Taff, S. A. (2013). Do we know who will drop out?: A review of the predictors of dropping out of high school: Precision, sensitivity, and specificity. *The High School Journal*, *96*(2), 77–100.

Bowers and Zhao (2018) Personal Communication.

Bryk, A. S., Gomez, L. M., Grunow, A., and LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Bulkley, K. E., Christman, J. B., Goertz, M. E., and Lawrence, N. R. (2010). Building with Benchmarks: The Role of the District in Philadelphia's Benchmark Assessment System. *Peabody Journal of Education*, *85*(2), 186–204.

Charrad, M., Ghazzali, N., Boiteau, V. and Niknafs, A. (2014). NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, *61(6)*, 1-36. URL http://www.jstatsoft.org/v61/i06/.

Chaykowski, K. (2017) How ClassDojo built one of the most popular classroom apps by listening to teachers. *Forbes*, June 13.

Clow, D. (2012). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134–138). ACM.

Clow, D. (2014). Data wranglers: human interpreters to help close the feedback loop. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 49–53). ACM.

Coburn, C. E., Honig, M. I., and Stein, M. K. (2009). What's the evidence on districts' use of evidence? *The role of research in educational improvement*, 67-87.

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

Coburn, C. E., and Talbert, J. E. (2006). Conceptions of evidence use in school districts: Mapping the terrain. *American journal of Education*, *112*(4), 469-495.

Coburn, C. E., and Turner, E. O. (2012). The Practice of Data Use: An Introduction. *American Journal of Education, 118*(2), 99–111.

Copland, M. A. (2003). Leadership of Inquiry: Building and Sustaining Capacity for School Improvement. *Educational Evaluation and Policy Analysis*, *25*(4), 375–395.

Datnow, A., and Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record*, *117*(4), 1–26.

Datnow, A., Park, V., and Kennedy-Lewis, B. (2012). High School Teachers' Use of Data to Inform Instruction. *Journal of Education for Students Placed at Risk (JESPAR)*, *17*(4), 247–265.

Datnow, A., Park, V., and Kennedy-Lewis, B. (2013). Affordances and constraints in the context of teacher collaboration for the purpose of data use. *Journal of Educational Administration and History*, *51*(3), 341–362.

Davidson, K. L., and Frohbieter, G. (2011). District Adoption and Implementation of Interim and Benchmark Assessments. CRESST Report 806. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST).*

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340.

Denton, C. A., Swanson, E. A., and Mathes, P. G. (2007). Assessment-based instructional coaching provided to reading intervention teachers. *Reading and Writing*, *20*(6), 569–590.

Drake, T. A. (2015). *How technology, strategic decision making, and school context influence principals' use of a data warehouse: A latent class growth analysis*. (E. Goldring, Ed.). Vanderbilt University, Ann Arbor, United States.

Dumais, S., Jeffries, R., Russell, D. M., Tang, D., and Teevan, J. (2014). Understanding User Behavior Through Log Data and Analysis. In *Ways of Knowing in HCI* (pp. 349–372). Springer, New York, NY.

Dunn, K. E., Airola, D. T., and Garrison, M. (2013). Concerns, knowledge, and efficacy: An application of the teacher change model to data driven decision-making professional development. *Creative Education*, *4*(10), 673.

Dunn, K. E., Airola, D. T., Lo, W.-J., and Garrison, M. (2013a). Becoming data driven: The influence of teachers' sense of efficacy on concerns related to data-driven decision making. *Journal of Experimental Education*, *81*(2), 222–241.

Dunn, K. E., Airola, D. T., Lo, W.-J., and Garrison, M. (2013b). What teachers think about what they can do with data: Development and validation of the data driven decision-making efficacy and anxiety inventory. *Contemporary Educational Psychology*, *38*(1), 87–98.

Efron, B., and Gong, G. (1983). A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician*, *37*(1), 36–48.

Fishbein, M., and Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Psychology Press.

Fishman, B. J., Penuel, W. R., Allen, A. R., Cheng, B. H., and Sabelli, N. O. R. A. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *National society for the study of education*, *112*(2), 136-156.

Fuchs, L. S., Fuchs, D., and Bishop, N. (1992). Instructional Adaptation for Students at Risk. *The Journal of Educational Research*, *86*(2), 70–84.

Gallimore, R., Ermeling, B. A., and Saunders, W. M. (2009). Moving the learning of teaching closer to practice: Teacher education implications of school-based inquiry teams. *The Elementary School Journal*.

Gašević, D., Kovanović, V., and Joksimović, S. (2017). Piecing the learning analytics puzzle: a consolidated model of a field of research and practice. *Learning: Research and Practice*, *3*(1), 63-78.

Geertz, C. (2017). The interpretation of cultures. Basic books.

George, A. A., Hall, G. E., and Stiegelbauer, S. M. (2006). *Measuring implementation in schools: The Stages of Concern Questionnaire*. Austin, TX: Southwest Educational Development Laboratory.

Gibson, S., and Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, *76*(4), 569.

Gold, Thomas; Lent, Jessica; Cole, Rachel; Kemple, James; Nathanson, Lori; and Brand, Janet. (2012). Usage Patterns and Perceptions of the Achievement Reporting and Innovation System (ARIS). *Research Alliance for New York City Schools*.

Gold, T., Lent, J., Cole, R., Kemple, J., Nathanson, L., and Brand, J. (2012a). Usage Patterns and Perceptions of the Achievement Reporting and Innovation System (ARIS). Technical Appendices. *Research Alliance for New York City Schools*.

Graves, K., and Bowers, A. J., 2018, Toward a Typology of Technology-Using Teachers in the "New Digital Divide": A Latent Class Analysis of the NCES Fast Response Survey System Teachers' Use of Educational Technology in U.S. Public Schools, 2009 (FRSS 95)

Greenwood, G. E., Olejnik, S. F., and Parkay, F. W. (1990). Relationships between four teacher efficacy belief patterns and selected teacher characteristics. *Journal of Research and Development in Education*.

Guskey, T. R. (1981). Measurement of the responsibility teachers assume for academic successes and failures in the classroom. *Journal of Teacher Education*, *32*(3), 44–51.

Guskey, T. R. (1984). The influence of change in instructional effectiveness upon the affective characteristics of teachers. *American Educational Research Journal*, *21*(2), 245–259.

Halverson, R. (2014). Data-Driven Leadership for Learning in the Age of Accountability. In A. J. Bowers, A. R. Shoho, and B. G. Barnett (Eds.), *Using Data in Schools to Inform Leadership and Decision Making* (pp. 255–266). INFORMATION AGE PUB Incorporated.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E. B., Supovitz, J. A., Wayman, J. C., … Steele, J. L. (2009). *Using Student Achievement Data to Support Instructional Decision Making* (GSE Publications No. NCEE 2009-4067). I.E.S. National Center for Educational Evaluation and Regional Assistance.

Hands, S., and Everitt, B. (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivariate Behavioral Research*, *22*(2), 235-243.

Haney, J. J., Wang, J., Keil, C., and Zoffel, J. (2007). Enhancing Teachers' Beliefs and Practices Through Problem-Based Learning Focused on Pertinent Issues of Environmental Health Science. *The Journal of Environmental Education*, *38*(4), 25–33.

Hastie, T., Tibshirani, R. and Freidman, J. (2017). *The elements of statistical learning*. New York, NY: Springer.

Hawn, A. (2015). The bridge report: bringing learning analytics to low-income, urban schools. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge* (pp. 410–411). ACM.

He, C., Parra, D., and Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, *56*, 9-27.

Hershkovitz, A., and Nachmias, R. (2009). Learning about online learning processes and students' motivation through web usage mining. *Interdisciplinary Journal of E-Learning and Learning Objects*, *5*(1), 197–214.

Home | International Educational Data Mining Society. (n.d.). Retrieved April 19, 2016, from http://www.educationaldatamining.org/

Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., and Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, *60*, 377–386.

Hoover, N. R., and Abrams, L. M. (2013). Teachers' Instructional Use of Summative Student Assessment Data. *Applied Measurement in Education*, *26*(3), 219–231.

Horn, I. S., Kane, B. D., and Wilson, J. (2015). Making sense of student performance data: Data use logics and mathematics teachers' learning opportunities. *American Educational Research Journal*, *52*(2), 208-242.

Hubbard, L., Datnow, A., and Pruyn, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation*, *42*, 54–62.

Huffman, D., and Kalnin, J. (2003). Collaborative inquiry to make data-based decisions in schools. *Teaching and Teacher Education*.

Il-Hyun, J., Kim, D., and Yoon, M. (2015). Constructing Proxy Variables to Measure Adult Learners' Time Management Strategies in LMS. *Journal of Educational Technology & Society*, *18*(3), 214–225.

James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York: springer.

Jimerson, J. B. (2014). Thinking about data: Exploring the development of mental models for "data use" among teachers and school leaders. *Studies in Educational Evaluation*, *42*, 5–14.

Jorion, N., Roberts, J., Bowers, A.J., Tissenbaum, M., Lyons, L., Kuma, V., Berland, M. (2018) Uncovering Patterns in Constructionist Collaborative Learning Activities via Cluster Analysis of Museum Exhibit Log Files. A paper submitted to the annual meeting of the American Educational Research Association (AERA), New York, NY: April 2018.

Kalbach, J. (2016). *Mapping experiences: A complete guide to creating value through journeys, blueprints, and diagrams*. " O'Reilly Media, Inc.".

Kaushik, A. (2010). Web Analytics 2.0: The Art of Online Accountability & Science of Customer Centricity. Wiley Publishing, Inc. Indiana

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., and Barney, H. (2006). Strategies to Promote Data Use for Instructional Improvement: Actions, Outcomes, and Lessons from Three Urban Districts. *American Journal of Education*, *112*(4), 496–520.

Koedinger, K. R., and Mathan, S. (2004). Distinguishing qualitatively different kinds of learning using log files and learning curves. In *ITS 2004 Log Analysis Workshop* (pp. 39–46).

Kovanovic, V; Gasevic, D.; Dawson, S.; Joksimovic, S.; Baker, Ryan S. and Hatala, M. (2015). Penetrating the Black Box of Time-on-task Estimation. LAK

Krumm, A., Means, B., and Bienkowski, M. (2018). *Learning Analytics Goes to School: A Collaborative Approach to Improving Education*. Routledge.

Lai, M. K., and McNaughton, S. (2016). The impact of data use professional development on student achievement. *Teaching and Teacher Education*, *60*, 434–443.

Lai, M. K., McNaughton, S., Amituanai-Toloa, M., Turner, R., and Hsiao, S. (2009). Sustained acceleration of achievement in reading comprehension: The New Zealand experience. *Reading Research Quarterly*, *44*(1), 30-56.

Lai, M. K., McNaughton, S., Timperley, H., and Hsiao, S. (2009). Sustaining continued acceleration in reading comprehension achievement following an intervention. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 81-100.

Lai, Mcnaightonm Timperly, Hsiao (2009)

Lam, S. F., Wong, B. P., Yang, H., and Liu, Y. (2012). Understanding student engagement with a contextual model. In *Handbook of research on student engagement* (pp. 403-419). Springer US.

Lee, J. E., Recker, M., Bowers, A. J., and Yuan, M. (2016). Hierarchical Cluster Analysis Heatmaps and Pattern Analysis: An Approach for Visualizing Learning Management System Interaction Data.

Little, J. W. (2012). Understanding Data Use Practice among Teachers: The Contribution of Micro-Process Studies. *American Journal of Education*, *118*(2), 143–166.

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., and Reber, S. (2017). Descriptive Analysis in Education: A Guide for Researchers. NCEE 2017-4023. *National Center for Education Evaluation and Regional Assistance*.

Lust, G., Elen, J., and Clarebout, G. (2013). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education*, *60*(1), 385–395.

Mandinach, E. B., and Gummer, E. (2015). Data-Driven Decision Making: Components of the Enculturation of Data Use in Education. *Teachers College Record*, *117*(4), 1.

Mandinach, E. B., and Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*.

Mandinach, E. B., Honey, M., Light, D., and Brunner, C. (2008). A conceptual framework for data-driven decision making. *Data-Driven School Improvement: Linking Data and Learning*, 13–31.

Mandinach, E. B., and Jackson, S. S. (2012). *Transforming teaching and learning through data-driven decision making*. Corwin Press.

Mandinach, E. B., and Jimerson, J. B. (2016). Teachers learning how to use data: A synthesis of the issues and what is known. *Teaching and Teacher Education*.

March, J. G., Sproull, L. S., and Tamuz, M. (1991). Learning from samples of one or fewer. *Organization science*, *2*(1), 1-13.

Marsh, J. A. (2007). Democratic dilemmas: Joint work, education politics, and community.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record*, *114*(11), 1–48.

Marsh, J. A., Hamilton, L., and Gill, B. (2008). Assistance and Accountability in Externally Managed Schools: The Case of Edison Schools, Inc. *Peabody Journal of Education*, *83*(3), 423–458.

Marsh, J. A., Pane, J. F., and Hamilton, L. S. (2006). Making Sense of Data-Driven Decision Making in Education. *Rand Education Occasional Papers*.

Marsh, J. A., Sloan McCombs, J., and Martorell, F. (2010). How Instructional Coaches Support Data-Driven Decision Making: Policy Implementation and Effects in Florida Middle Schools. *Educational Policy*, *24*(6), 872–907.

Maull, K. E. (2013). Computational Methods for Analyzing and Understanding Online Curriculum Planning Behavior of Teachers.

Maull, K. E., Saldivar, M. G., and Sumner, T. (2010a). Observing the online behavior of teachers: From Internet usage to personalization for pedagogical practice. In *ACM SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA*. academia.edu.

Maull, K. E., Saldivar, M. G., and Sumner, T. (2010b). Online curriculum planning behavior of teachers. In *Educational Data Mining 2010*. educationaldatamining.org.

McDougall, D., Saunders, W. M., and Goldenberg, C. (2007). Inside the Black Box of School Reform: Explaining the how and why of change at Getting Results schools. *International Journal of Disability, Development and Education*, *54*(1), 51–89.

McFarland, D. J., and Hamilton, D. (2006). Adding contextual specificity to the technology acceptance model. *Computers in human behavior*, *22*(3), 427-447.

McNamara, D. S., Graesser, A. C., McCarthy, P. M., and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.

Means, B., Padilla, C., and Gallagher, L. (2010). Use of Education Data at the Local Level: From Accountability to Instructional Improvement. *NIH Guide for Grants and Contracts / U.S. Department of Health, Education, and Welfare*.

Meijer, C., and Foster, S. F. (1988). The effect of teacher self-efficacy on referral chance. *The Journal of Special Education*, *22*(3), 378–385.

Midgley, C., Feldlaufer, H., and Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology*, *81*(2), 247.

Moen, Ronald and Norman, Clifford L. (2010). Circling Back: Clearing up myths about the Deming Cycle and seeing how it keeps evolving. *Quality Progress*, 43(11), 22-28.

Moody, L., and Dede, C. (2008). Models of data-based decision making: A case study of the Milwaukee Public Schools. *Mandinach and M. Honey (Eds.), Data-Driven School Improvement: Linking Data and Learning*, 233–254.

Munzner, T. (2014). *Visualization analysis and design*. AK Peters/CRC Press.

Murnane, R. J., Sharkey, N. S., and Boudett, K. P. (2005). Using Student-Assessment Results to Improve Instruction: Lessons From a Workshop. *Journal of Education for Students Placed at Risk (JESPAR)*, *10*(3), 269–280.

Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification*, *31*(3), 274-295.

Natriello, G. (03/1987). The Impact of Evaluation Processes on Students. *Educational Psychologist*, *22*(2), 155–175.

Nelson, T. H., and Slavit, D. (2007). Collaborative inquiry among science and mathematics teachers in the USA: professional learning experiences through cross-grade, cross-discipline dialogue. *Journal of In-service Education*.

Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, *10*(1), 53.

Nitkin, D. (2018). *Technology-Based Personalization: Instructional Reform in Five Public Schools* (Doctoral dissertation, Columbia University).

Nunamaker, J. F., Romano, N. C., and Briggs, R. O. (2001). A framework for collaboration and knowledge management. In *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on* (p. 12 pp.–).

Nye, B. D., Graesser, A. C., and Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427–469.

Ocumpaugh, J., Baker, R. S., San Pedro, M. O., Hawn, M. A., Heffernan, C., Heffernan, N., and Slater, S. A. (2010). Guidance counselor reports of the ASSISTments college prediction model (ACPM).

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543–578.

Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., and Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 117–124). ACM.

Pardos, Z. A., Gowda, S. M., Baker, R. S., and Heffernan, N. T. (2011). Ensembling Predictions of Student Post-Test Scores for an Intelligent Tutoring System. In *EDM* (pp. 189–198).

Park, V., and Datnow, A. (2017). Ability Grouping and Differentiated Instruction in an Era of Data-Driven Decision Making. *American Journal of Education*, *123*(2).

Part 100 Regulations of the Commissioner of Education of New York State. (2018) 100.5 Diploma Requirements. Retrieved from http://www.p12.nysed.gov/part100/pages/1005.html#a.

Pea, R. (2014). The learning analytics workgroup: A report on building the field of learning analytics for personalized learning at scale. Stanford University Stanford, CA.

Piety, P. J. (2013). *Assessing the educational data movement*. Teachers College Press.

Piety, P. J., Behrens, J., and Pea, R. (2013). Educational data sciences and the need for interpretive skills. *American Educational Research Association*, *27*.

Piety, P. J., Hickey, D. T., and Bishop, M. J. (2014, March). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 193-202). ACM.

Podell, D. M., and Soodak, L. C. (1993). Teacher Efficacy and Bias in Special Education Referrals. *The Journal of Educational Research*, *86*(4), 247–253.

Poortman, C. L., and Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. *Teaching and Teacher Education*, *60*, 425–433.

Porter, K. E., and Snipes, J. C. (2006). The Challenge of Supporting Change: Elementary Student Achievement and the Bay Area School Reform Collaborative's Focal Strategy.

Preece, J., Rogers, Y., and Sharp, H. (2015). *Interaction design: beyond human-computer interaction*. John Wiley and Sons.

Prenger, R., and Schildkamp, K. (2018). Data-based decision making for teacher and student learning: a psychological perspective on the role of the teacher. *Educational psychology*, 1-19.

Quint, J. C., Sepanik, S., and Smith, J. K. (2008). Using Student Data to Improve Teaching and Learning: Findings from an Evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) Program …. *MDRC*.

Recker, M., and Lee, J. E. (2016). Analyzing Learner and Instructor Interactions within Learning Management Systems: Approaches and Examples. In M. J. Spector, B. B. Lockee, and M. D. Childress (Eds.), *Learning, Design, and Technology* (pp. 1–23). Springer International Publishing.

Reeves, T. D., Summers, K. H., and Grove, E. (2016). Examining the landscape of teacher learning for data use: The case of Illinois. *Cogent Education*, *3*(1), 1211476.

Research Tools - Megan Tschannen-Moran's Web Site. (2018). Retrieved April 7, 2017, from http://wmpeople.wm.edu/site/page/mxtsch/researchtools.

Reynolds, D., Stringfield, S., and Schaffer, E. C. (2006). The High Reliability Schools Project. *Improving schools and educational systems*, 56.

Riehl, C. J., Earle, H., Nagarajan, P., Schwitzman, T. E., & Vernikoff, L. (2018). Following the Path of Greatest Persistence: Sensemaking, Data Use, and the Everyday Practice of Teaching. In *Cases of Teachers' Data Use* (pp. 46-59). Routledge.

Riehl, C., Pallas, A. M., and Natriello, G. (1991). *More Responsive High Schools, Student Information, and Problem Solving*. Center for Research on Effective Schooling for Disadvantaged Students, the Johns Hopkins University.

Rodrigo, M., Mercedes, T., Baker, R. S., McLaren, B. M., Jayme, A., and Dy, T. T. (2012). Development of a Workbench to Address the Educational Data Mining Bottleneck. *International Educational Data Mining Society*.

Rogosa, D. (2005). Statistical Misunderstandings of the Properties of School Scores and School Accountability. *Yearbook of the National Society for the Study of Education*, *104*(2), 147–174.

Romero, C., Ventura, S., and García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384.

Rose, J. S., and Medway, F. J. (1981). Measurement of teachers' beliefs in their control over student outcome. *The Journal of Educational Research*, *74*(3), 185–190.

Ross, J. A. (1992). Teacher Efficacy and the Effects of Coaching on Student Achievement. *Canadian Journal of Education / Revue Canadienne de L'éducation*, *17*(1), 51–65.

Rotter, Julian B. (1966). Generalized Expectancies for Internal Versus External Control of Reinforcement. Psychological Monographs: General and Applied, 80(1), 1-28.

San Pedro, M. O. C. Z., Baker, R. S., Bowers, A., and Heffernan, N. T. (2013). Predicting College Enrollment from Student Interaction with an Intelligent Tutoring System in Middle School. In *EDM* (pp. 177–184).

Schildkamp, K., and Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, *26*(3), 482–496.

Schildkamp, K., Karbautzki, L., and Vanhoof, J. (2014). Exploring data use practices around Europe: Identifying enablers and barriers. *Studies in educational evaluation*, *42*, 15-24.

Schildkamp, K., Poortman, C., Luyten, H., and Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. *School effectiveness and school improvement*, *28*(2), 242-258.

Schunk, D. H., and Pajares, F. (2005). Competence perceptions and academic functioning. *Handbook of Competence and Motivation*, *85*, 104.

Schunk, D. H., and Pajares, F. (2009). Self-efficacy theory. *Handbook of Motivation at School*, 35–53.

Senge, P. M. (2006). *The Fifth Discipline: The Art and Practice of the Learning Organization*. Doubleday/Currency.

Shahid, J., and Thompson, D. (2001). Teacher Efficacy: A Research Synthesis. Retrieved April 5, 2017, from http://eric.ed.gov/?id=ED453170

Shaw, S. (2010). Utilizing multilevel modeling to examine teachers' sense of efficacy in relation to their use of data and student achievement. Dissertation

Shaw, S., and Wayman, J. C. (2012). Third-Year Results From an Efficacy Study of the Acuity Data System.

Shih, C.F. and Venkatesh, A. (2004). Beyond adoption: Development and application of a use-diffusion model. Journal of Marketing,

Shneiderman, Ben. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." In Proceedings of the IEEE Conference on Visual Languages, pp. 336–343. IEEE Computer Society, 1996.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4-14.

Skaalvik, E. M., and Skaalvik, S. (2007). Dimensions of Teacher Self-Efficacy and Relations With Strain Factors, Perceived Collective Teacher Efficacy, and Teacher Burnout. *Journal of Educational Psychology*.

Skaalvik, E. M., and Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education*, *26*(4), 1059–1069.

Skinner, E. A., Wellborn, J. G., and Connell, J. P. (1990). What it takes to do well in school and whether I've got it: A process model of perceived control and children's engagement and achievement in school. *Journal of Educational Psychology*, *82*(1), 22.

Snodgrass Rangel, V., Bell, E. R., Monroy, C., and Whitaker, J. R. (2015). Toward a New Approach to the Evaluation of a Digital Curriculum Using Learning Analytics. *Journal of Research on Technology in Education*, *47*(2), 89–104.

Soodak, L. C., and Podell, D. M. (1993). Teacher Efficacy and Student Problem as Factors in Special Education Referral. *The Journal of Special Education*, *27*(1), 66–81.

Stringfield, S., Reynolds, D., and Schaffer, E. C. (2008). Improving secondary students' academic achievement through a focus on reform reliability: 4- and 9-year findings from the High Reliability Schools project. *School Effectiveness and School Improvement*, *19*(4), 409–428.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, *10*(2-3), 211–227.

Supovitz, J. A. (2006). *The case for district-based reform: Leading, building and sustaining school improvement*. Harvard Education Pr.

Sutherland, S. (2004). Creating a Culture of Data Use for Continuous Improvement: A Case Study of an Edison Project School. *American Journal of Evaluation*, *25*(3), 277–293.

Teddlie, C., and Stringfield, S. (2007). A history of school effectiveness and improvement research in the USA focusing on the past quarter century. In *International handbook of school effectiveness and improvement* (pp. 131-166). Springer, Dordrecht.

Timperley, H. S., and Phillips, G. (2003). Changing and sustaining teachers' expectations through professional development in literacy. *Teaching and Teacher Education*, *19*(6), 627–641.

Tschannen-Moran, M., and Hoy, A. W. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.

Tschannen-Moran, M., Hoy, A. W., and Hoy, W. K. (1998). Teacher Efficacy: Its Meaning and Measure. *Review of Educational Research*, *68*(2), 202–248.

Tyler, J. H. (2013). If you build it will they come? Teachers' online use of student performance data. *Education Finance and Policy*, *8*(2), 168–207.

Venkatesh, V., and Davis, F. D. (1996). A model of the antecedents of perceived ease of use: Development and test. *Decision sciences*, *27*(3), 451-481.

Venkatesh, V.; Thong, J.; and Xu, X (2016) "Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead," *Journal of the Association for Information Systems*, 17(5), 328-376.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.

Venkatesh, V. (2000). Determinants of perceived ease of use: Integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information systems research*, *11*(4), 342-365.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., and Santos, J. L. (2013). Learning analytics dashboard applications. *The American Behavioral Scientist*, *57*(10), 1500–1509.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., and Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 1–16.

Walker, D. A., Reeves, T. D., and Smith, T. J. (2016). Confirmation of the Data-Driven Decision-Making Efficacy and Anxiety Inventory's Score Factor Structure Among Teachers. *Journal of Psychoeducational Assessment*, 0734282916682905.

Ware, C. (2008). Toward a perceptual theory of flow visualization. *IEEE Computer Graphics and Applications*, *28*(2).

Ware, C. (2012). *Information visualization: perception for design*. Elsevier.

Wayman, J. C., Cho, V., Jimerson, J. B., and Spikes, D. D. (2012). District-wide effects on data use in the classroom.

Wayman, J. C., Cho, V., and Shaw, S. (2009a). First-Year Results From an Efficacy Study of the Acuity Data System.

Wayman, J. C., Cho, V., and Shaw, S. (2009b). Survey of educator data use. *Unpublished Document*.

Wayman, J. C., Shaw, S., and Cho, V. (2011). Second-Year Results From an Efficacy Study of the Acuity Data System.

Wayman, J. C., Shaw, S., and Cho, V. (2017). Longitudinal Effects of Teacher Use of a Computer Data System on Student Achievement. *AERA Open*, *3*(1), 2332858416685534.

Wayman, J. C., Snodgrass Rangel, V. W., Jimerson, J. B., and Cho, V. (2010). Improving data use in NISD: Becoming a data-informed district. *Austin: The University of Texas. Retrieved October*, *13*, 2010.

Wayman, J. C., Wilkerson, S. B., Cho, V., Mandinach, E. B., and Supovitz, J. A. (2016). Guide to using the Teacher Data Use Survey.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 203–211). ACM.

Xhakaj, F., Aleven, V., and McLaren, B. M. (2016, September). How teachers use data to help students learn: Contextual inquiry for the design of a dashboard. In *European Conference on Technology Enhanced Learning* (pp. 340-354). Springer, Cham.

Xhakaj, F., Aleven, V., and McLaren, B. M. (2017, September). Effects of a teacher dashboard for an intelligent tutoring system on teacher knowledge, lesson planning, lessons and student learning. In *European Conference on Technology Enhanced Learning* (pp. 315-329). Springer, Cham.

Xu, B., and Recker, M. (2012). Teaching Analytics: A Clustering and Triangulation Study of Digital Library User Data. *Journal of Educational Technology & Society*, *15*(3), 103–115.

Yanchar, S. C. (2006). On the possibility of contextual–quantitative inquiry. *New Ideas in Psychology*, *24*(3), 212-228.

Yoo, Y., Lee, H., Jo, I. H., and Park, Y. (2015). Educational dashboards for smart learning: Review of case studies. In *Emerging issues in smart learning* (pp. 145-155). Springer, Berlin, Heidelberg.

Yildiz, O., Bal, A., and Gulsecen, S. (2015). Statistical and Clustering Based Rules Extraction Approaches for Fuzzy Model to Estimate Academic Performance in Distance Education. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(2).

Yu, T., and Jo, I.-H. (2014). Educational Technology Approach Toward Learning Analytics: Relationship Between Student Online Behavior and Learning Performance in Higher

Education. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 269–270). New York, NY, USA: ACM.

Appendix A

Survey Instruments and Subscales

The DBDM Efficacy and Anxiety Survey (3D-MEA) (Dunn, Airola, Lo, et al., 2013b)

Self-efficacy for the application of data to instruction (N of items = 6, Cronbach's Alpha = .92)					
Please indicate how much you agree or disagree with the following statements:	**Strongly Disagree**	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**
I am confident that I can use data to identify students with special learning needs					
I am confident that I can use data to identify gaps in student understanding of curricular concepts					
I am confident that I can use assessment data to provide targeted feedback to students about their performance or progress					
I am confident I can use assessment data to identify gaps in my instructional curriculum					
I am confident that I can use data to group students with similar learning needs for instruction					
I am confident in my ability to use data to guide my selection of targeted interventions for gaps in student understanding					

Self-efficacy for data identification and access (N of items = 3, Cronbach's Alpha = .84)					
Please indicate how much you agree or disagree with the following statements:	**Strongly Disagree**	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**
I am confident in my ability to access state assessment results for my students					
I am confident that I know what types of data or reports I need to assess group performance					
I am confident that I know what types of data or reports I need to assess student performance					

Self-efficacy for data technology use (N of items = 3, Cronbach's Alpha = .91)					
Please indicate how much you agree or disagree with the following statements:	**Strongly Disagree**	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**
I am confident I can use the tools provided by my district's data technology system to retrieve charts, tables or graphs for analysis					
I am confident I can use the tools provided by my district's data technology					

system to filter students into different groups for analysis					
I am confident that I can use my district's data analysis technology to access standard reports					

Self-efficacy for data analysis and interpretation (N of items = 3, Cronbach's Alpha = .81)					
Please indicate how much you agree or disagree with the following statements:	**Strongly Disagree**	**Disagree**	**Neither Agree nor Disagree**	**Agree**	**Strongly Agree**
I am confident in my ability to understand assessment reports					
I am confident in my ability to interpret student performance from a scaled score					
I am confident in my ability to interpret subtest or standard scores to determine student strengths and weaknesses in a content area					

Teacher Self-Efficacy Scale - Short Form (TSES - Short)

(Tschannen-Moran & Woolfolk Hoy, 2001; “Research Tools - Megan Tschannen-Moran’s Web Site,” 2018)

Directions:

Please indicate your opinion about each of the questions below by marking any one of the nine responses in the columns on the right side, ranging from (1) None at all to (9) A Great Deal as each represents a degree on the continuum.

Please respond to each of the questions by considering the combination of your current ability, resources, and opportunity to do each of the following in your present position.

Efficacy in Instructional Strategies (N of Items = 4, Cronbach’s Alpha = .86)									
	None at All 1	**2**	**Very Little 3**	**4**	**Some Degree 5**	**6**	**Quite a Bit 7**	**8**	**A Great Deal 9**
To what extent can you use a variety of assessment strategies?									
How well can you implement alternative teaching strategies in your classroom?									
To what extent can you provide an alternative explanation or example when students are confused?									
To what extent can you craft good questions for your students?									

Efficacy in Student Engagement

(N of Items = 4, Cronbach's Alpha = .81)

	None at All 1	2	Very Little 3	4	Some Degree 5	6	Quite a Bit 7	8	A Great Deal 9
How much can you do to motivate students who show low interest in school work?									
How much can you do to help your students value learning?									
How much can you do to get students to believe they can do well in school work?									
How much can you assist families in helping their children do well in school?									

Norwegian Teacher Self-Efficacy Scale (NTES) (Skaalvik & Skaalvik, 2007, 2010)

Adapt Instruction to Individual Needs (N of Items = 4, Cronbach's Alpha = .87)							
How certain are you that you can:	**Not certain at all 1**	**2**	**Quite uncertain 3**	**4**	**Quite certain 5**	**6**	**Absolutely certain 7**
organize schoolwork to adapt instruction and assignments to individual needs.							
provide realistic challenge for all students even in mixed ability classes.							
adapt instruction to the needs of low-ability students while you also attend to the needs of other students in class.							
organize classroom work so that both low- and high-ability students work with tasks that are adapted to their abilities.							

External Control Scale (N of Items = 5, Cronbach's Alpha = .79)						
Please indicate how true you feel each statement to be.	**False**	**Mostly false**	**More false than true**	**More true than false**	**Mostly True**	**True**
How much students can learn in school is primarily determined by their abilities.						
If students have not learned discipline at home, there is not much the school can do.						
A teacher cannot do much to improve students' achievements if those students have limited abilities for schoolwork.						
*Good teaching is more important to students' engagement in schoolwork than students' home environment						
It is practically impossible for a teacher to motivate a student for academic work if that student lacks support and stimulation at home.						
*Reverse Coded						

Appendix B: Profile Plots by Content Area

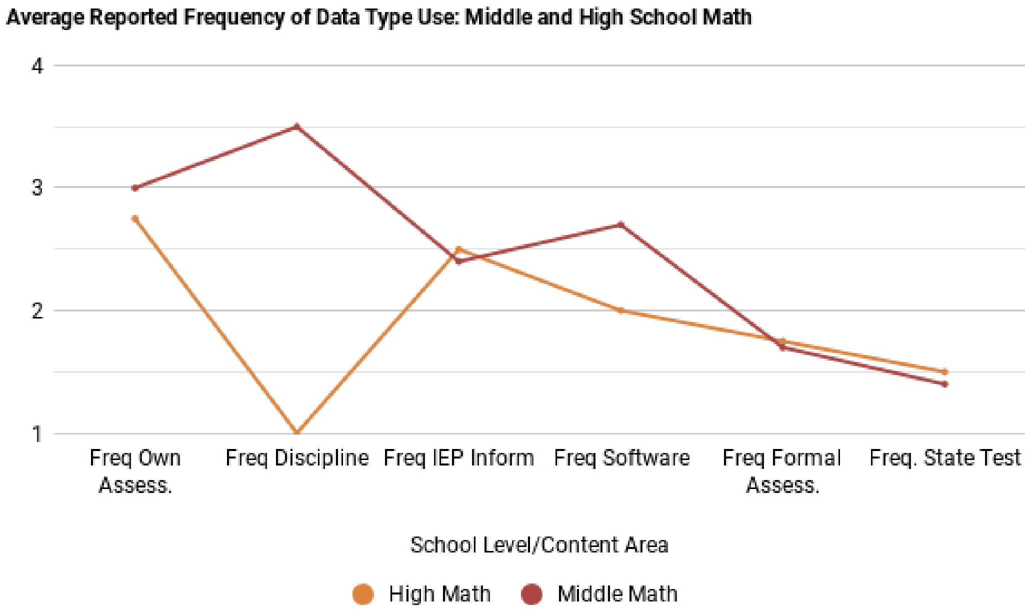


Figure 51. Average reported frequency of data type use: middle and high school Math

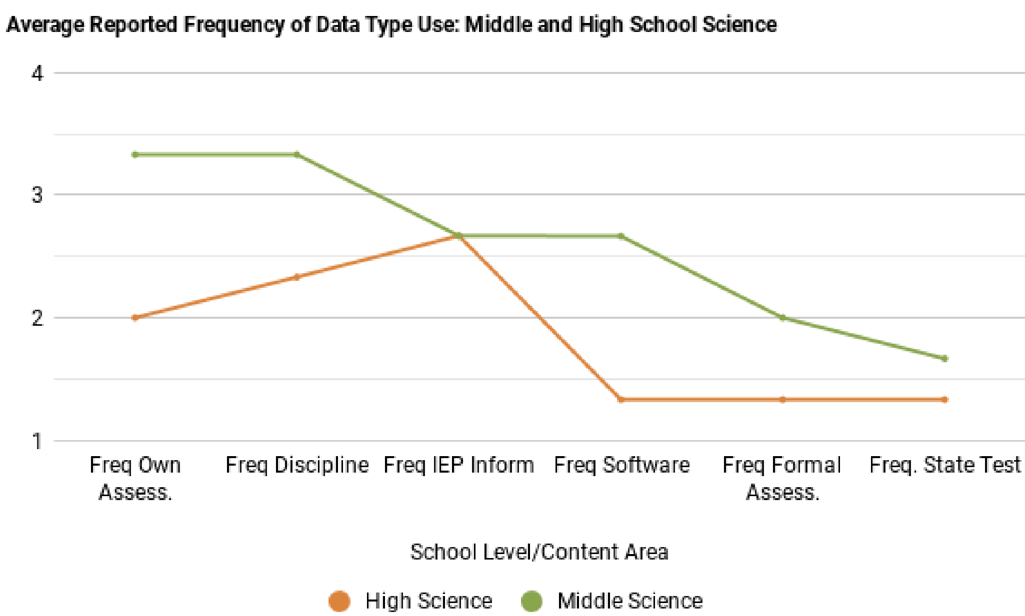


Figure 52. Average reported frequency of data type use: middle and high school Science

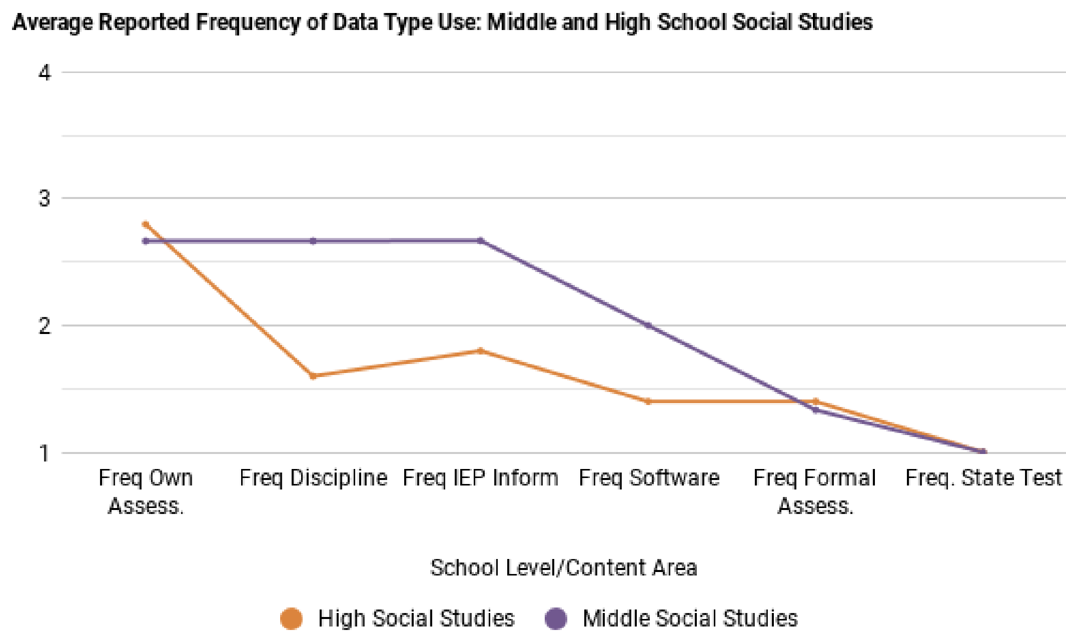


Figure 53: Average reported frequency of data type use: middle and high school Social Studies

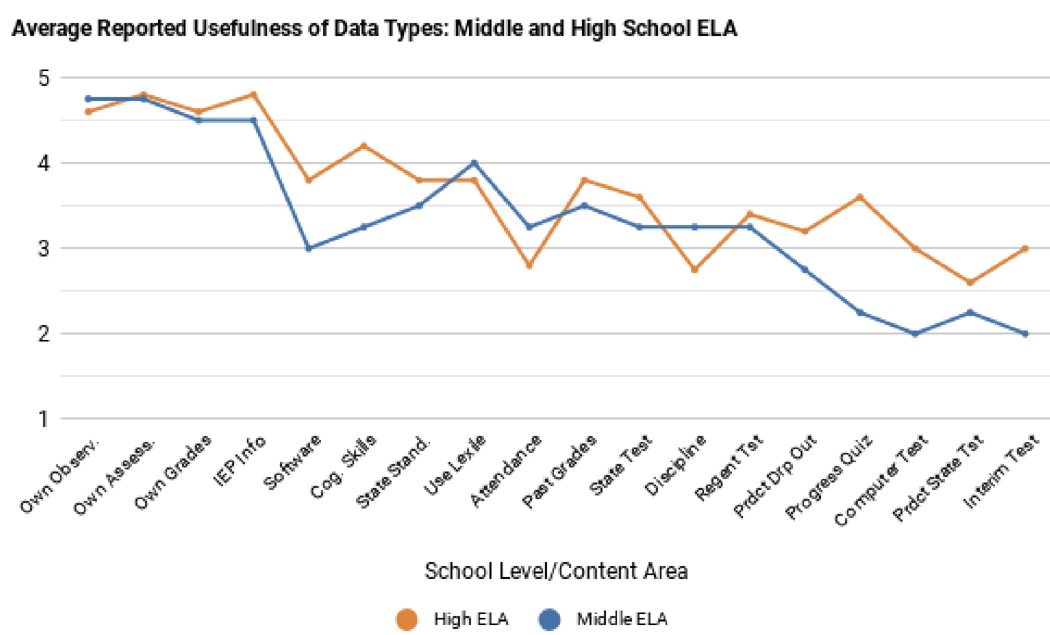


Figure 54. Average reported usefulness of data types: middle and high school ELA

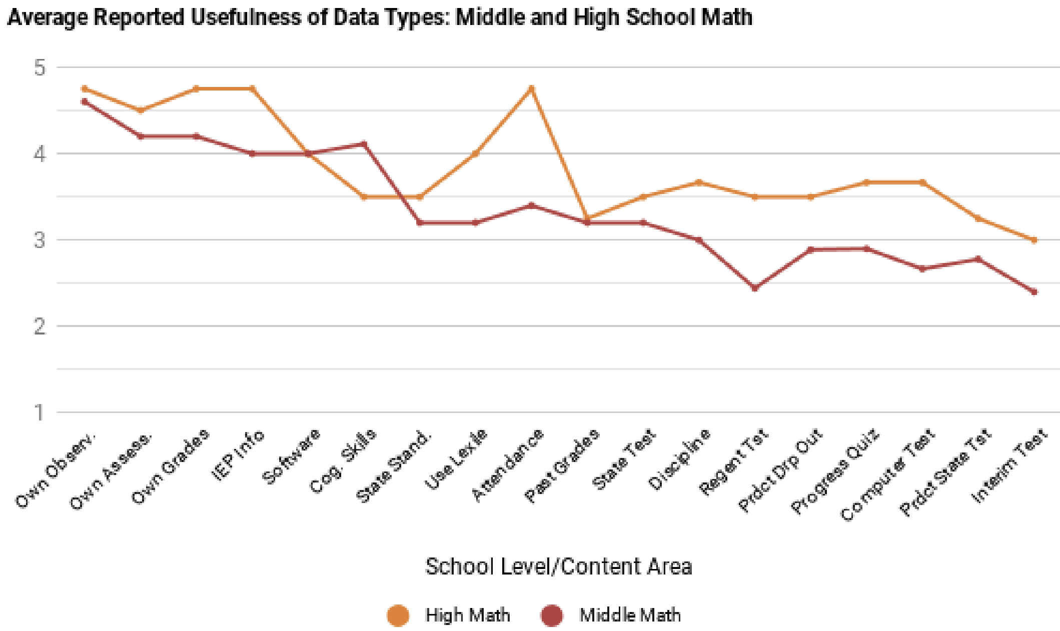


Figure 55. Average reported usefulness of data types: middle and high school Math

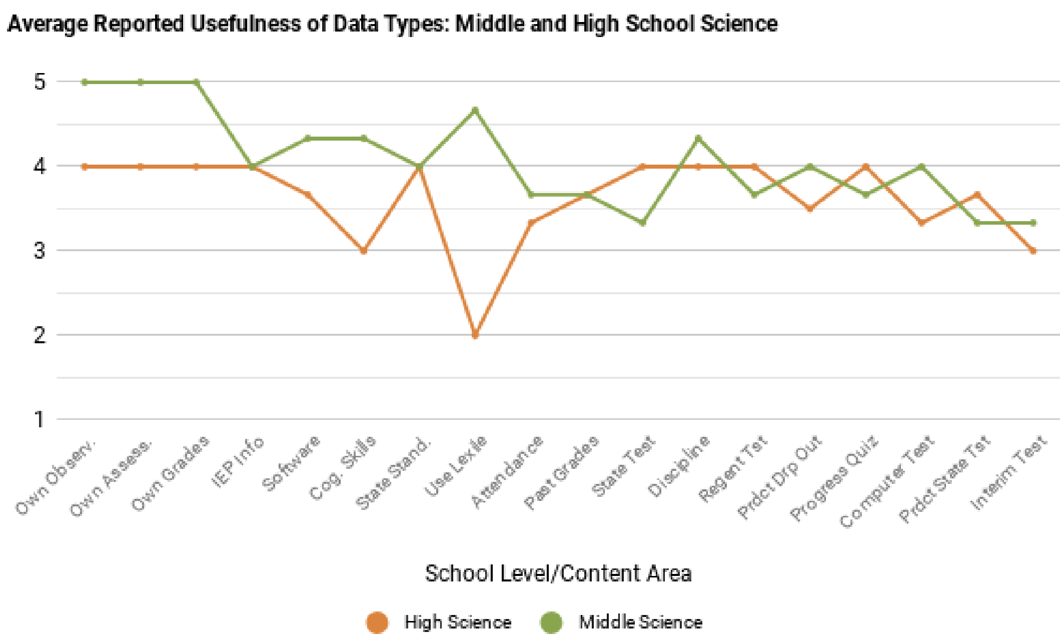


Figure 56. Average reported usefulness of data types: middle and high school Science

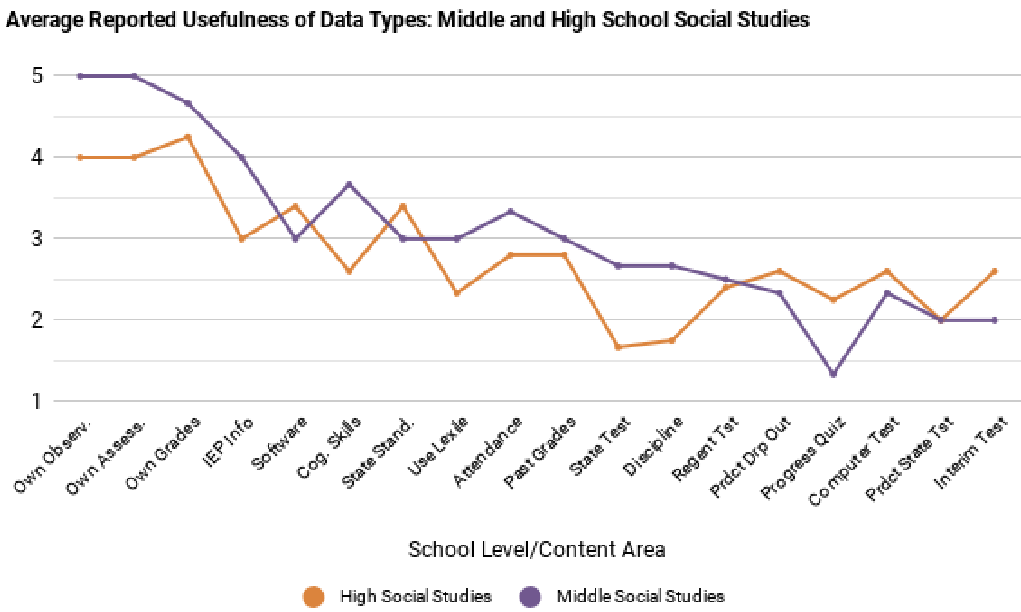


Figure 57. Average reported usefulness of data types: middle and high school Social Studies

Appendix C: Dendrograms with Distance Scale and Cluster Divisions

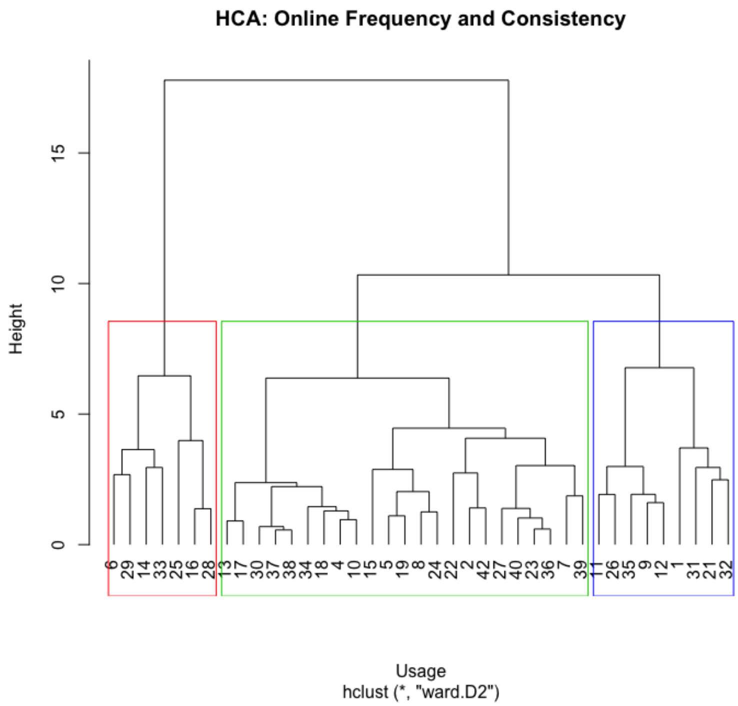


Figure 58. HCA dendrogram: online frequency and consistency

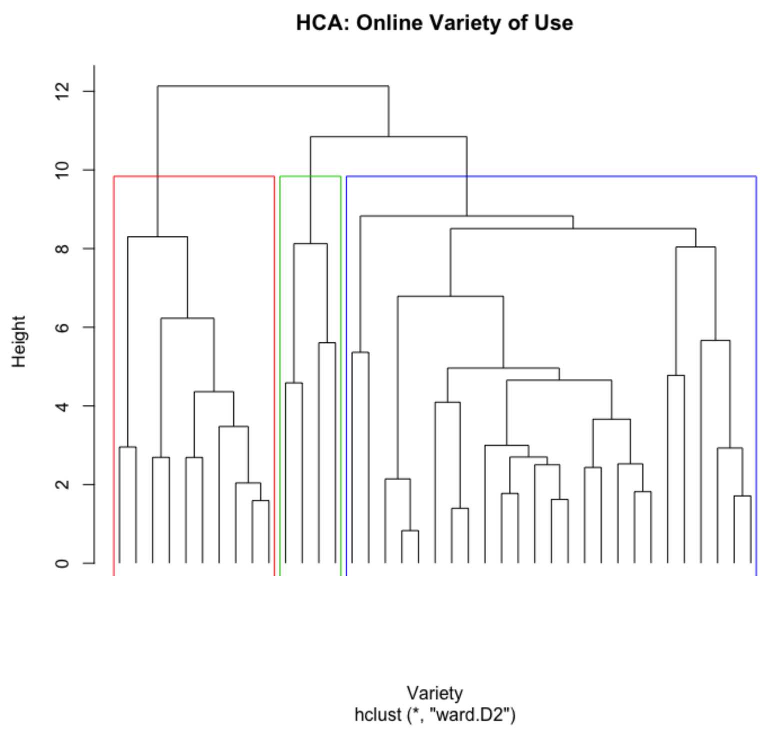


Figure 59. HCA dendrogram: online variety of use

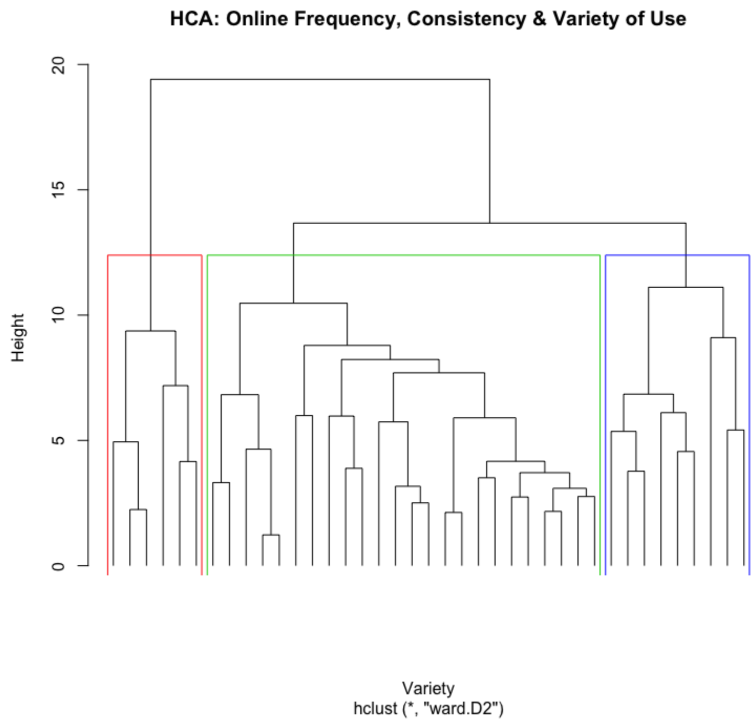


Figure 60. HCA dendrogram: online frequency, consistency, and variety of use

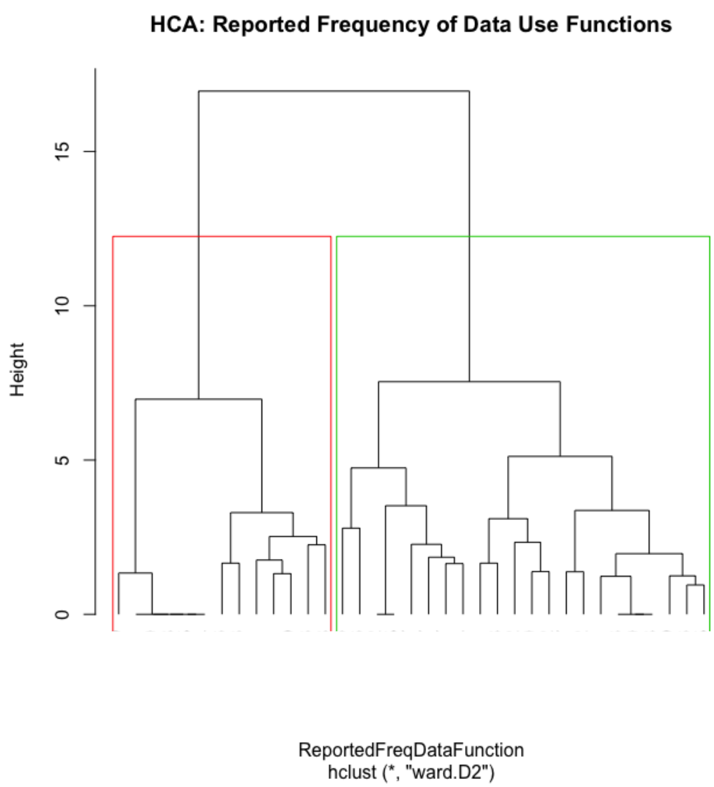


Figure 61. HCA dendrogram: reported frequency

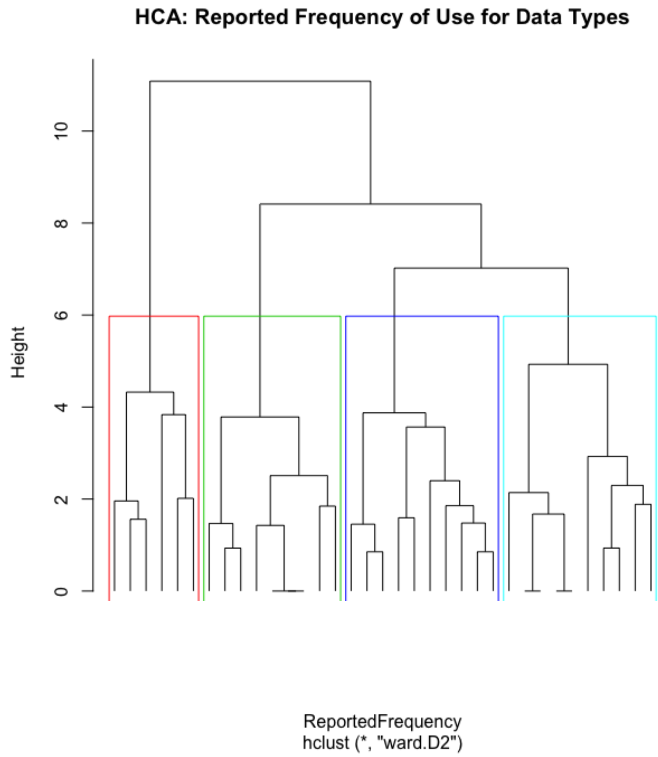


Figure 62. HCA dendrogram: reported frequency of use for data types

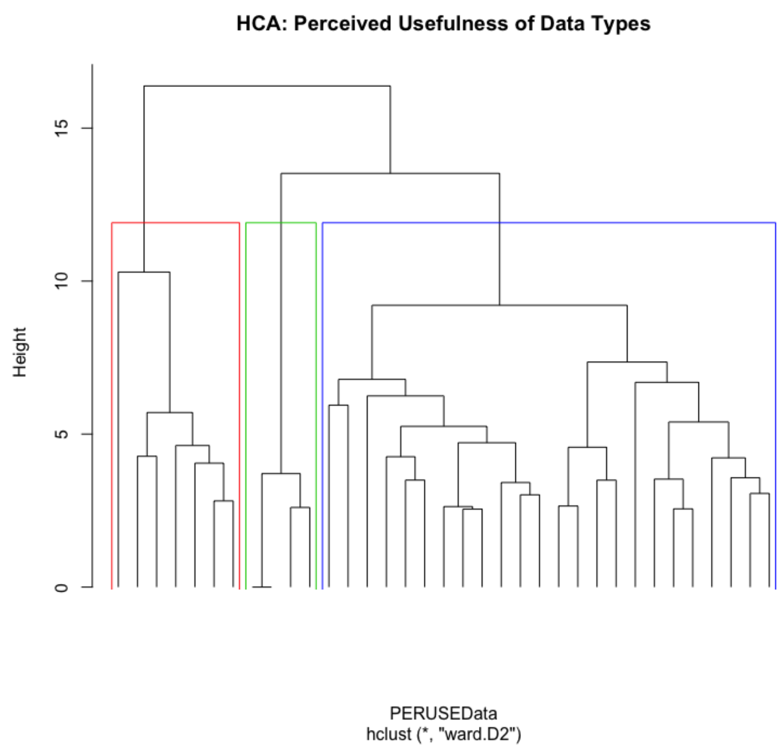


Figure 63. HCA dendrogram: Perceived usefulness of data types

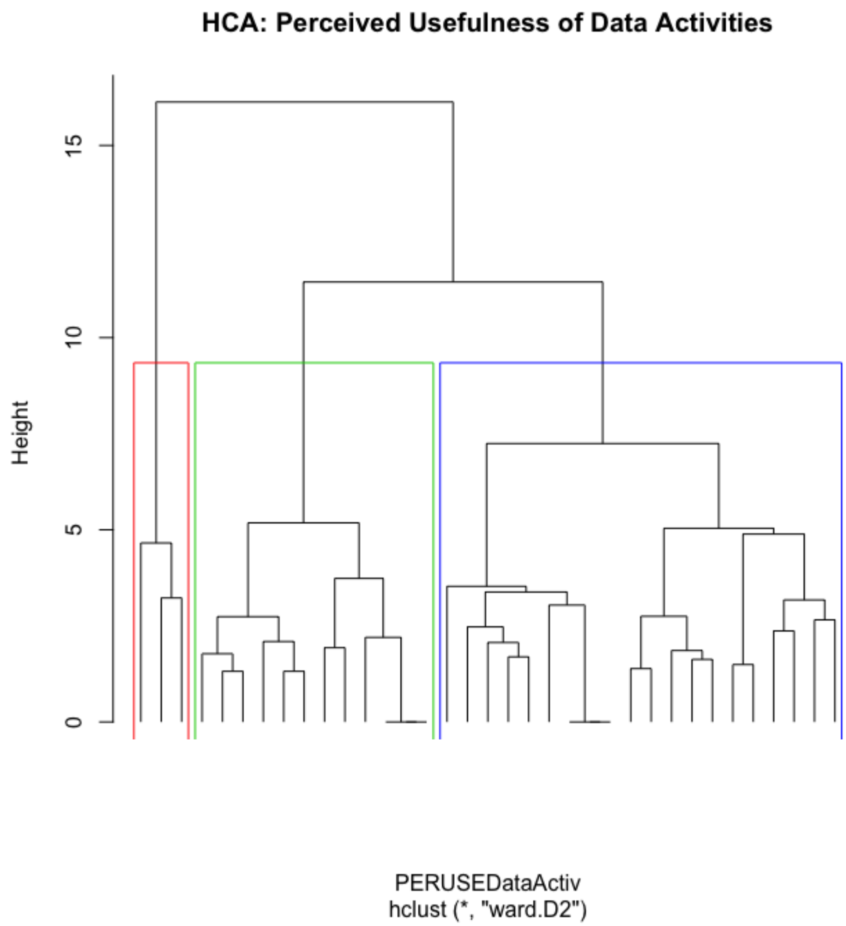


Figure 64. HCA dendrogram: perceived usefulness of data activities

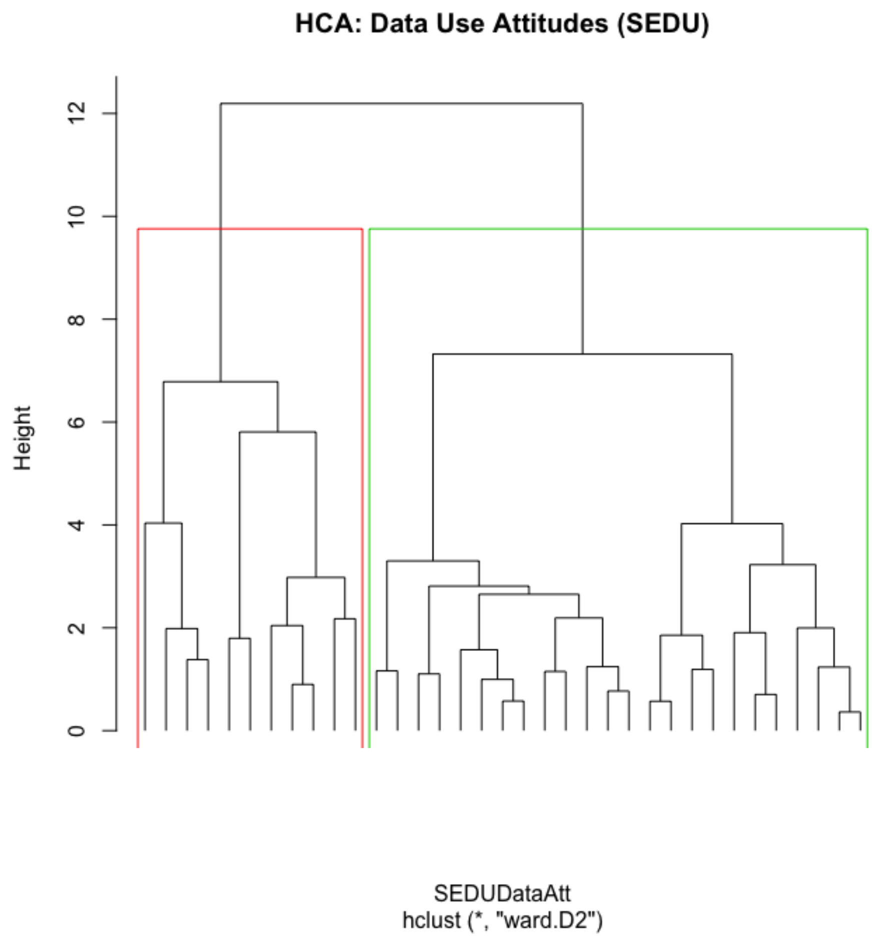


Figure 65. HCA dendrogram: survey of educator data use (SEDU). NbClust suggests 2 clusters (10 Methods) and 5 clusters (7 Methods).

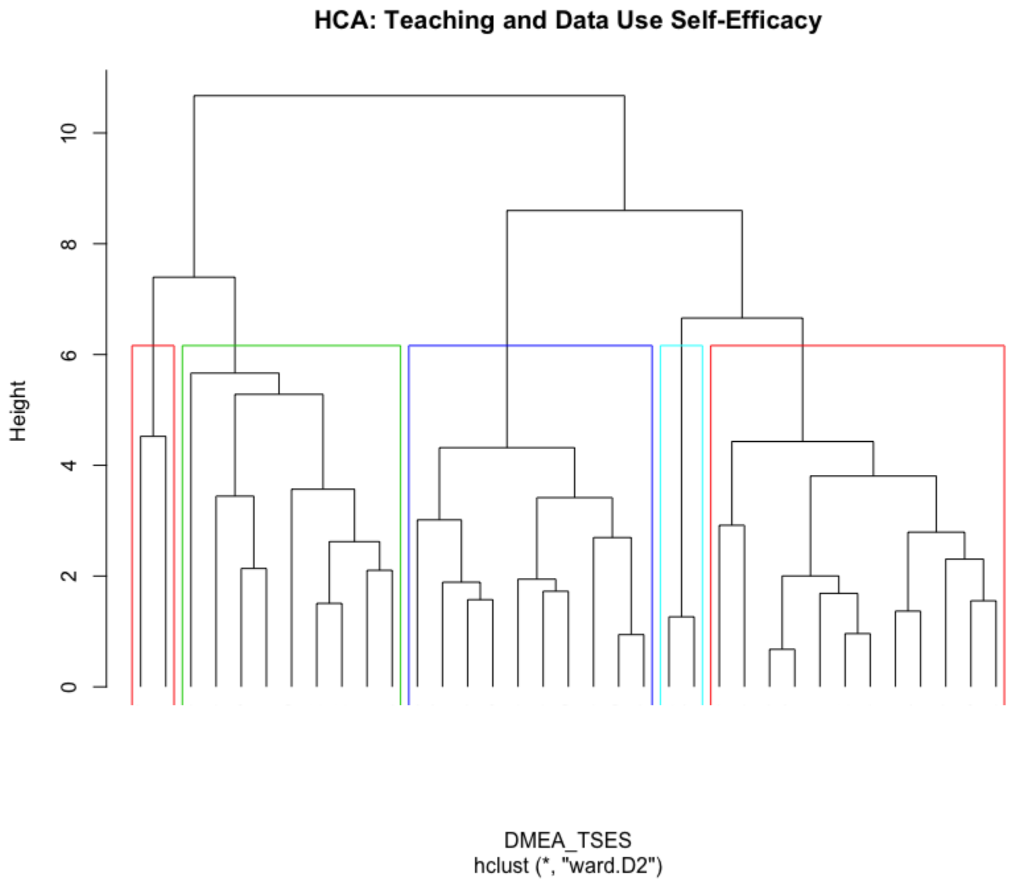


Figure 66. HCA dendrogram: data-driven decision making and teaching self-efficacy. NbClust suggests 5 clusters (12 methods)

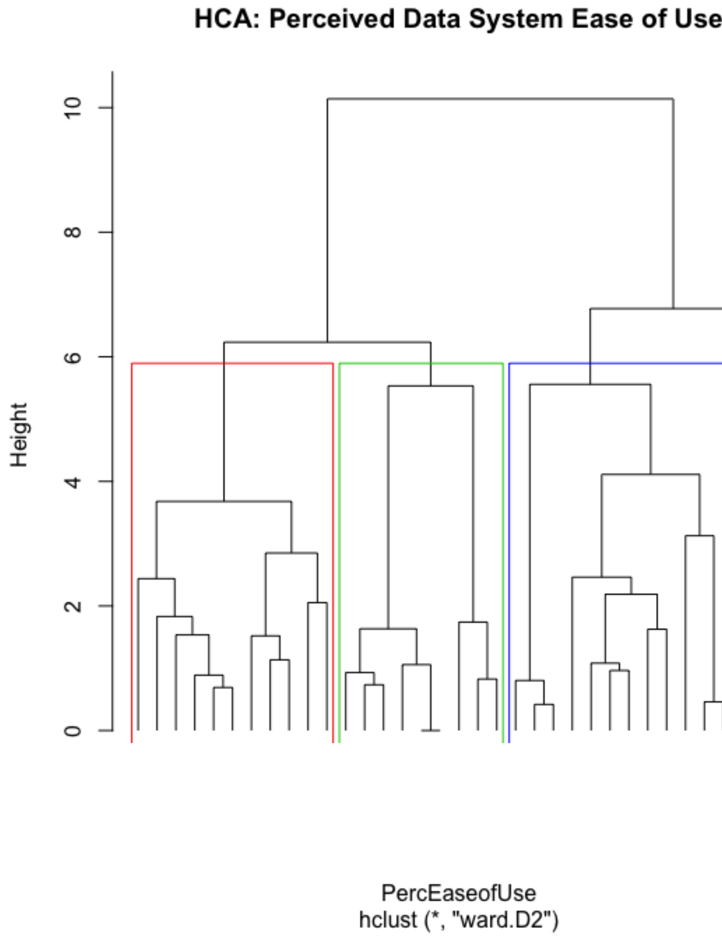


Figure 67. HCA dendrogram: perceived ease of use for Benchmark Data system. Nine NbClust methods proposed 5 clusters; 8 methods proposed 2 clusters.

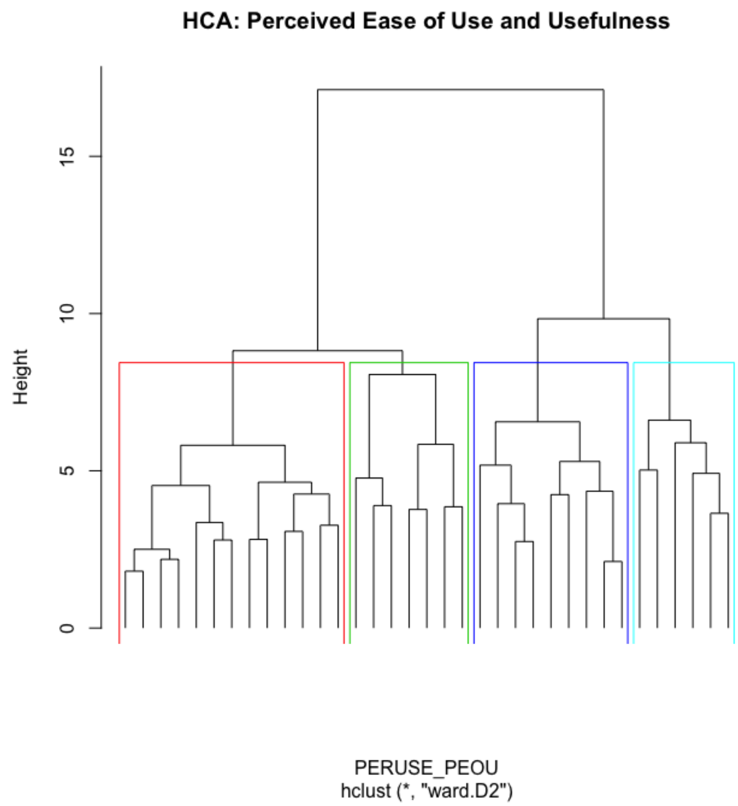


Figure 68. HCA dendrogram: perceived ease of use and usefulness of data types. 11 methods propose 2 clusters; 5 propose four clusters

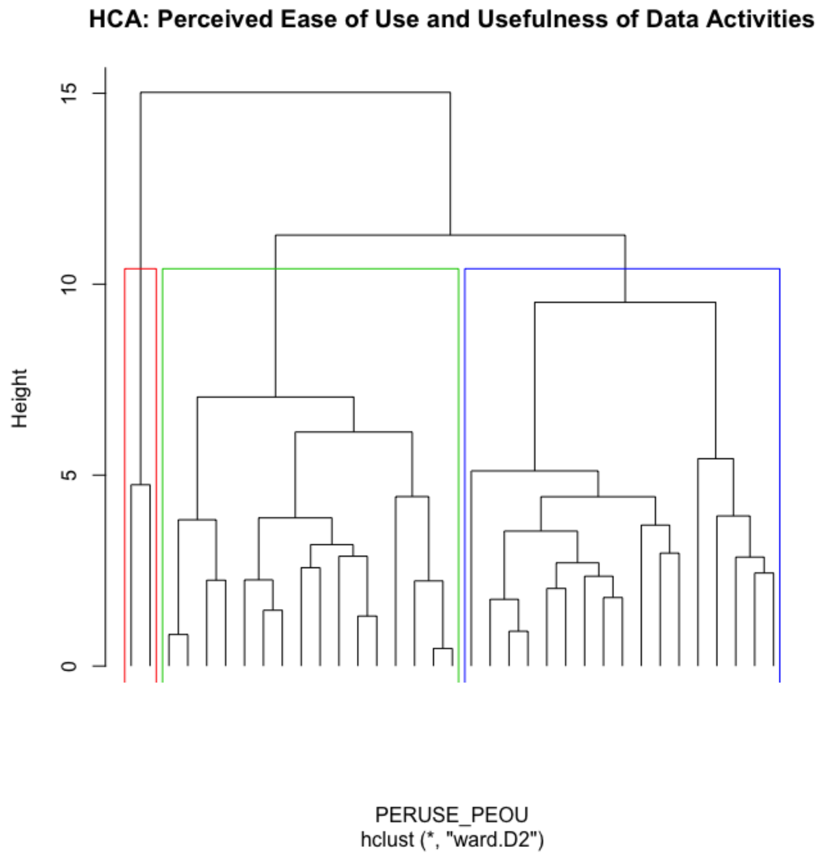


Figure 69. HCA dendrogram: perceived ease of use and usefulness of data activities. 15 methods from NbClust propose two clusters. Visual inspection suggests 3 clusters for interpretation.

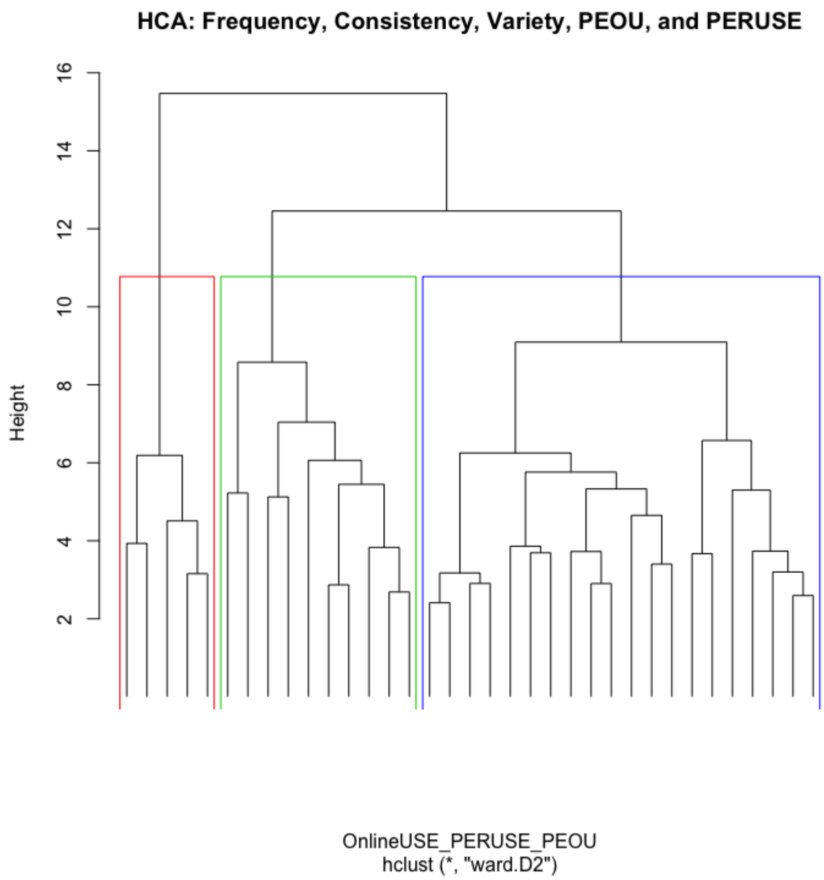


Figure 70. HCA dendrogram: frequency, consistency, variety, PEOU, and PERUSE. 11 NbClust Methods propose 2 clusters; 7 methods suggest 3 clusters. Visual inspection suggests 4 interesting clusters for interpretation.