Boston College
Lynch School of Education and Human Development

Department of
Measurement, Evaluation, Statistics, and Assessment

DEVELOPMENT OF A RASCH/GUTTMAN SCENARIO INSTRUMENT TO MEASURE TEACHERS' USE OF DATA TO INFORM CLASSROOM INSTRUCTION

Dissertation

by

CAITLIN DIANE HOGUE

submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

May 2022

Abstract

DEVELOPMENT OF A RASCH/GUTTMAN SCENARIO INSTRUMENT TO MEASURE TEACHERS' USE OF DATA TO INFORM CLASSROOM INSTRUCTION

Caitlin Diane Hogue

Larry H. Ludlow, Chair

Teachers in the United States are increasingly tasked with using data to inform their classroom instruction both through federal policies, such as the Every Student Succeeds Act (ESSA, 2016), and state policies requiring the use of teacher-determined data-driven goals for performance evaluations (MA 603 CMR 35.07). Many teachers, however, report that they feel underprepared to engage in this type of work (Dunn et al., 2013), also called Data-Driven Decision Making (DDDM). In addition, there is currently a limited set of instruments to measure the construct of using data to inform classroom instruction and the instruments that currently exist measure this construct using a typical Classical Test Theory design.

This work developed an instrument called the Using Data to Inform classroom Instruction (UDII) scale to measure teachers' use of data to inform classroom instruction. It used the Rasch/Guttman Scenario (RGS) methodology, an approach that develops scenarios that reflect the rich lived experiences of individuals (Antipkina & Ludlow, 2020; Ludlow et al., 2014). The RGS approach utilizes the Rasch model, part of the family of Item Response Theory models, which conceptualizes a construct as a hierarchical continuum. Scenario items and people are plotted on the same variable map, which allows for the development of rich descriptions of individuals at particular raw

score locations on the continuum. An interpretative variable map is included to help schools and districts use the results of the survey.

This work adds to the growing body of literature utilizing the RGS approach, as well as the literature focused on the use of data to inform classroom instruction (or DDDM). The UDII scale can be utilized by schools and districts who are engaged in the work of using data to inform classroom instruction to identify the current skillsets of teachers and/or teams of teachers to provide differentiated support, or it can be used before and after an intervention focused on using data to inform classroom instruction to measure change.

Acknowledgements

This has been a long time coming and there are many people to whom I am indebted for their help in making this dream of mine a reality. All of these individuals have supported my dissertation work through a full-time job, the birth of both of my children, and a global pandemic that drastically changed my work requirements, requiring a pause in my dissertation work. Without the support of these people, this work would never have been completed and I am so utterly grateful to each and every one of them.

I want to thank my dissertation committee for their (enduring) patience and strong support. There are not enough words to thank Dr. Larry Ludlow, my advisor, for his incredibly fast turnaround with very detailed feedback and his continued support of me as the pace of my dissertation work ebbed and flowed. Dr. Laura O'Dwyer and Dr. Vincent Cho, your feedback made this work much stronger, and I appreciate your patience with me as I popped in and out of your lives.

I owe much gratitude to my parents, Terry and Bob Green, who have been my ultimate cheerleaders throughout my life and have always believed that I can do anything I set out to do, even when I doubt myself. That support, and their tireless help taking care of my children, has made my work successful.

Finally, I thank my spouse, Adam, and my children, CC and Ryan, for their support and love. There were many weekends that I was holed up in the library or my office missing our family life, and I thank Adam and CC for supporting me in this work (and CC, the drawings you made for me helped me when I was missing you!). CC and

Ryan, you two are my inspiration and I hope that one day you will read this work and remember that you can always accomplish what you set out to do, even when it seems impossible, especially if you rely on your strong network of support. I will always be that support for you. PS – Ryan, thank you for being such a great napper as a newborn so that I could finish this work.

Table of Contents

List of Tables

List of Figures

Chapter 1 : Introduction

The Problem

Teachers in the United States are increasingly being tasked with using data to inform their classroom instruction both through federal policies, such as No Child Left Behind (NCLB, 2001) and its reauthorization in the form of the Every Student Succeeds Act (ESSA, 2016), and state policies that require the use of teacher-determined data-driven goals for teacher performance evaluations (such as MA 603 CMR 35.07). Although NCLB promoted the use of data to inform instruction, ESSA expands on this directive with the expectation of data use at all levels in the education system (Mandinach & Gummer, 2016b) and includes the development of skills to use data to inform classroom instruction using a variety of data sources as part of the description of professional development (ESSA, 2016).

State policies that incorporate the use of data to inform classroom instruction require teachers to identify a goal based on an analysis of student data and track progress towards the achievement of that goal. For example, in Massachusetts, as part of the teacher evaluation process, teachers are required to set student learning goals for the year and track progress towards achieving those goals by collecting and analyzing student data throughout the year (MA 603 CMR 35.07). This use of data to inform classroom instruction has often been called data-driven decision making, or DDDM, in both research and practice, and its underlying logic assumes that if data are used by teachers in a timely manner to inform instructional decisions, those decisions (which are driven by data) will result in improvements in student achievement (Herman & Haertel, 2005).

The increased expectation for teachers to use data to inform their instruction, however, has not necessarily been accompanied by training or targeted support for teachers to engage in this activity in a meaningful way (Bocala et al., 2014; Champion, 2017; Dunn et al, 2013; Hamilton et al., 2009; National Council of Teachers of Mathematics, 2010). For many teachers, the requirement to use data to inform classroom instruction comes as an additional undertaking without any training or support, and many teachers feel underprepared to use data to inform their classroom instruction (Dunn et al., 2013). When teachers' performance evaluations include a measure of teachers' use of data to inform classroom instruction, such as in Massachusetts, this lack of training or support becomes very problematic: how can teachers be evaluated on a set of tasks that they have not been prepared for or supported in doing?

Along with a lack of training or support for teachers to engage in using data to inform instruction is a limited set of instruments to measure teachers' capabilities for using data to inform instruction. This is problematic, as this limited set of instruments makes it difficult to measure either current status for teachers, which could be used to provide targeted support or differentiated professional development to teachers based on their own needs, or changes in teacher status after specific training or support.

Purpose

This dissertation focuses on the following research objective: the development of a measurement instrument to support teachers in using data to inform their classroom instruction. To meet this objective, I used the Rasch/Guttman Scenario (RGS) approach (Antipkina & Ludlow, 2020; Ludlow et al., 2014) to develop a measurement instrument to measure teachers' own perceived level of a set of skills hypothesized to be utilized when using data to inform instruction. The hypothesized set of skills that are utilized when using data to inform instruction were developed as part of this dissertation. The measurement instrument was developed using the RGS measurement approach, which is a scenario-based scale methodology that reflects "lived experiences" for those responding to the instrument. The RGS measurement approach for a scenario-based scale is a novel approach that utilizes Guttman's facet theory design component (Borg & Shye, 1995; Guttman & Greenbaum, 1998) and sentence mapping (Hackett, 2014), along with Rasch measurement principles (Rasch, 1966) to develop short scenarios, or vignettes, that describe the lived experiences of an individual who possesses specific levels of individual facets (or components of the construct being measured). The scenarios as a whole encompass the entire continuum of the construct being measured. Those responding to the measurement instrument are asked to evaluate how similar or different their own lived experiences are to each individual scenario. The RGS approach provides scores that are linked to a detailed description of an individual at that score along the construct's continuum (Ludlow et al., 2020). Further detail on the RGS approach and scenario-based scales, Rasch measurement principles, Guttman's facet theory, and sentence mapping are described in the Literature Review section.

This dissertation stems from previous instrument development work that I engaged in during my doctoral coursework. As part of the Boston College ERME8864 Survey Methods course in Fall 2015, I piloted an instrument developed using Classical Test Theory (CTT) methods and based on the theoretical framework of Data-Driven Decision Making (DDDM) posited by Mandinach (2012). This pre-dissertation survey was not developed utilizing the RGS approach, but is included here as the pre-dissertation form of the survey that was developed for this dissertation. Mandinach's (2012) theoretical framework describes a hypothesized series of stages that teachers progress through when using data to inform instruction and is described in detail in the Literature Review section. It is very similar to many of the data inquiry cycle frameworks currently described in the literature and in educational practice.

An important consideration for my dissertation, however, is that I prefer to use the terminology “using data to inform instruction” rather than Data-Driven Decision Making (or DDDM) as it is often called in the literature, because the phrase “using data to inform instruction” places the emphasis on instruction, while DDDM places the emphasis on data. I believe the emphasis should be on instruction, as this is the purpose of engaging in these activities for teachers, and maintaining that focus on instruction in the terminology can both reinforce that purpose and help reassure teachers who may be adverse to “data” about the main purpose of this activity. This emphasis on instruction rather than data was also identified during the expert review of the pre-dissertation pilot survey items (described later in this dissertation) as more palatable than “data-driven decision making”, given the focus on instruction and the implication that it conveys a somewhat more holistic approach to influencing teaching and learning. This shift in nomenclature is beginning to be represented in the literature, with a focus on the use of the term “data-informed decision making”, indicating that data inform decisions which is combined with professional knowledge to contribute to both achievement and learning (Brown et al., 2017; Schildkamp et al., 2019). It has also been described as evidence-based practice in the literature (Horn et al., 2015) and data-based decision making (Brown et al., 2017).

This dissertation expands upon my pre-dissertation instrument which was based on Mandinach's (2012) DDDM framework (see Figure 2.2), which describes a set of hypothesized cognitive skills involved in DDDM. This dissertation further develops the hypothesized set of skills involved in using data to inform classroom instruction by synthesizing existing frameworks and cycles of inquiry currently described in the literature. Once this hypothesized set of skills was synthesized, I utilized the RGS approach to write scenarios to measure these skills. As part of the RGS approach, a Rasch model (Rasch, 1966) was utilized to test if the actual responses to these scenarios support the hypothesized continuum of skills.

Study Significance

Schools and districts will be able to use the data from this instrument to identify where teachers perceive themselves to be on the set of skills involved in using data to inform classroom instruction. The location of teachers along this continuum of skills is linked to detailed descriptions of a person at that location along this continuum (Ludlow et al., 2020). This information can then be used by school or district staff to differentiate professional development for teachers based on their perceived skill levels for this process; a teacher who is more advanced would receive different experiences or supports than a teacher who is lower on this set of skills. Alternatively, school or district staff can use this information when supporting teacher teams, as it can provide information on specific skills for which a particular teacher team may need additional support.

This dissertation also contributes to the growing body of literature on the RGS approach and its potential for use with a variety of different constructs. Additional studies utilizing the RGS approach are described in the Literature Review. Finally, this dissertation adds to the body of literature focused on how teachers are using data to inform their instruction and how data can be used in classrooms.

Chapter Summary

This chapter provides an introduction to the construct of using data to inform classroom instruction and elaborates on the context for this dissertation. It describes the need for an additional instrument to measure teachers' perceived levels of skill with using data to inform classroom instruction and suggests that this need can be filled with the use of the RGS approach to develop a scenario-based instrument. This chapter also describes the benefits of the use of this type of survey instrument and notes that this dissertation adds to the growing body of literature describing the use of the RGS approach in survey design. The next chapter provides an overview of the construct of using data to inform classroom instruction and describes the particular set of skills (also called facets in the RGS approach) in detail. It also describes existing instruments that have been developed thus far to measure teachers' use of data to inform classroom instruction and the different methodological approaches used in the pre-dissertation instrument and in the instrument developed for this dissertation.

Chapter 2 : Literature Review

This chapter focuses on a review of the literature related to my dissertation. The literature focused on teachers' use of data to inform instruction (also referred to as DDDM in the literature) is discussed to provide a broad overview of this domain, followed by details on the RGS approach and scenario-based scales. The literature review then describes nine peer-reviewed or published models of DDDM/data inquiry cycles in detail, which form the basis for the facets of the construct of using data to inform classroom instruction. These facets are explained with detailed descriptions. The self-efficacy literature is discussed briefly, as the pre-dissertation instrument was designed to measure teachers' self-efficacy with using data to inform classroom instruction. The literature review closes with an overview of existing instruments to measure the use of data to inform instruction, as well as a discussion of Classical Test Theory (CTT) and Item Response Theory (IRT).

Overview of DDDM

The formalized use of data in the education field was initially focused on accountability and holding teachers, principals, schools, and districts accountable for student achievement (Brown et al., 2017; Mandinach, 2012), which may have prompted a negative connotation with data in the minds of some educators. The shift in education to using data not solely for accountability purposes, but also to empower teachers, principals, schools, and districts to utilize data to make better informed decisions signals a paradigm change in the way data use is conceptualized (Brown et al., 2017; Coburn & Turner, 2012; Jimerson, 2013; Little, 2012; Mandinach, 2012). Understandably, many teachers have not yet internalized this paradigm shift and may still hold a negative view towards data use, prompting my preference to utilize the terminology of using data to inform instruction rather than data-driven decision making (DDDM). As mentioned in Chapter 1, this shift in nomenclature is occasionally reflected in the literature, specifically in some shifts to labeling this process as data-informed decision making (Brown et al., 2017; Schildkamp et al., 2019), evidence-based practice (Horn et al., 2015) and data-based decision making (Brown et al., 2017). These slight changes in nomenclature still emphasize data, although in different ways; my preference is to emphasize instruction. Given, however, that the literature utilizes DDDM as the primary terminology, this section will focus on the literature describing DDDM.

Although there are many models of DDDM that currently exist, the practice essentially focuses on the following activities: teachers identify areas of interest in their classrooms and ask questions based on those areas of interest, they gather data to help refine their questions, they analyze that data to help answer their questions, they use the results of their analysis, along with their pedagogical expertise, to identify an instructional intervention, they implement that intervention and determine if it was successful using data, and repeat the process if their intervention was not successful (Mandinach & Gummer, 2016b; Mandinach, 2012). This process is often referred to as an inquiry cycle. In this process, "administrators and teachers collect and analyze data to help *inform* educational decisions" (Datnow, 2011, p. 148; italics in original). Datnow (2011) points out that data alone do not identify an action plan or decision, but rather provide information that can be used to inform decision making. This requires that teachers (or administrators) interact with the data to utilize it to inform decision making

(Coburn & Turner, 2011; Schildkamp, 2019; Spillane, 2012) and, specifically, requires a set of skills that are utilized during this interaction.

The underlying logic of DDDM for teachers posits that if teachers use data to inform their classroom instruction in a timely manner, the academic achievement of students in those classrooms will improve (Hamilton et al., 2009; Herman & Haertel, 2005; Marsh, Pane, & Hamilton, 2006). In addition, the use of data to inform classroom instruction can allow teachers to differentiate their instruction based on individual student needs (Hamilton et al., 2009). Prior research has shown that the use of data to inform decisions in schools can lead to increases in student academic achievement (Anfara & Donhost, 2010; Doyle, 2003; Mason, 2002).

Despite research indicating increases in student academic achievement as a result of teacher engagement in DDDM, other research indicates that engaging in the process of using data to inform classroom instruction may not lead to increases in student achievement (Neuman, 2016). A likely reason that DDDM may fail to improve student outcomes is the need to identify an instructional intervention that can be successful for student learning. If the teacher or team engaging in this process is unable to identify an intervention that may have a positive impact on student achievement, this may result in no improvement, or even a negative effect, on student outcomes (Brown et al., 2017; Neuman, 2016). In fact, Neuman (2016) suggests that selecting the wrong type of intervention (for example, increasing the number of worksheets that students complete instead of improving content-rich instruction) can have a negative impact on student achievement. This research highlights the importance of high quality interventions focused on content and learning during this process.

The use of data to inform classroom instruction is not something that is new to teachers, as many teachers have been implementing this practice informally for years (Datnow, 2011; Kekahio & Baker, 2013; Mandinach, 2012; National Forum on Education Statistics, 2012; Pella, 2012). What is new is the formalized and more automated processes that are in place to help teachers engage with data to inform their classroom instruction in a systematized manner; this systemization of the process is at the core of DDDM (Brown et al., 2017; Dunn et al., 2013; Hamilton et al., 2007; Marsh et al., 2006; Mandinach, 2012). The DDDM framework described by Mandinach (2012) in Figure 2.2 is an example of this systematization. The formalized description of this process, which can also be described as a cycle of inquiry, is one example of the way that this practice has become more systematic. Additionally, federal policies such as NCLB, and more explicitly, its successor ESSA, require states and districts to provide teachers with support in using data and assessments to inform instruction (Mandinach & Gummer, 2016b). ESSA specifically requires all levels of the education system to use data to make educational decisions and requires that administrators and policy makers help to facilitate the use of data to make instructional decisions (Mandinach & Gummer, 2016b).

Under ESSA, important for the use of data to inform classroom instruction, data are not restricted solely to assessment data, but also include other sources of data such as behavioral data, demographic data, motivational data, attendance data, school climate data, financial operation data, and observational data, among others (Bernhardt, 2004; Brown et al., 2017; Coburn et al., 2020; Marsh et al, 2006; Mandinach & Gummer, 2016b; Schildkamp, 2019; Schildkamp et al., 2019). This broad definition of data is important in this process, as restricting data to data from assessments only becomes problematic when the goal is to inform instruction. Assessments, especially standardized assessments, may provide only general information due to the nature of their design, making an analysis of instructional interventions difficult (Pella, 2012). Additionally, assessments are often administered infrequently, with large amounts of time in between the administration and the availability of data, making instructional interventions even more difficult (Pella, 2012). Given these constraints, Pella (2012) states, “A variety of forms of classroom data are necessary to support instructionally embedded formative assessment; data provide a focus for teachers to reflect on the pedagogical reasoning that occurred during instruction” (p. 60), and specifically notes that qualitative data should be incorporated into inquiry cycles. Qualitative data can take the form of observations (of student participation and engagement), work samples, and teachers’ reflections (Pella, 2012). The National Forum on Education Statistics (2012) also notes that a variety of data sources should be utilized when using data to inform classroom instruction, noting similar data sources as those described under ESSA, as well as additional sources such as heath data, program data, transportation data, and workforce data (p. 4).

Despite this requirement for all levels of the educational system to use data to inform decision making, there has been a lack of training both for current teachers (Dunn, Airola, Lo, & Garrison, 2013; National Council of Teachers of Mathematics, 2010) and for pre-service teachers in teacher education programs (Mandinach, 2012), leading many teachers to report that they do not feel prepared or feel underprepared to use data to inform instructional decision making (Dunn et al., 2013). This lack of preparation offers an opportunity to provide teachers with training in using data to inform their instruction based on their specific needs; however, there is currently a limited set of instruments or measurement tools to identify the specific needs of individual teachers.

Additionally, the use of data to inform instruction is most frequently a group activity, where a teacher team engages in this process together, although individual teachers do also engage in this process on their own. When individual teachers who are part of teacher teams feel underprepared or not prepared to engage in this work, it can lead to sub-optimal use of team meeting time and can potentially result in teams who do not believe that engaging in this process can be beneficial for students if the process is unsuccessful due to lack of support or preparation.

The next section describes the RGS approach and scenario-based scales in detail.

RGS Measurement Approach and Scenario-Based Scales

Scenario-based scales are relatively new to the field of survey design (Ludlow et al., 2014) and, at a broad level, present scenarios to respondents that describe particular lived experiences of an individual and ask the respondents to compare their own experiences to that of the individual in each scenario. The scenarios are carefully constructed based on the definition of the construct being measured. The definition of the construct includes identifying the specific facets, or components, that make up the construct as a whole. The construct is hypothesized to be hierarchical in nature and the facets are utilized to write each scenario by varying the level of the individual facets in each scenario to provide coverage of the entire hierarchical construct. In this way, each scenario reflects different levels of the construct along the hierarchical continuum and the set of scenarios as a whole provides coverage of the entire hierarchical continuum of the construct. Respondents respond to these scenarios using a rating scale format that has been crafted to measure the similarity of their own experiences to those described in the scenarios.

A key feature of scenario-based scales is that the scenarios reflect the lived experiences of individuals (Ludlow et al., 2014) and provide a more comprehensive and rich description of these lived experiences than typical Likert-based scales designed to measure particular presumably independent features of the construct. Scores from scenario-based scales are linked to a detailed description of an individual at that score level, enhancing the interpretation of an individual's status, or location, on the construct (Ludlow et al., 2020). Recent scenario-based scales that have been developed utilizing the RGS approach include those focused on living with meaning and purpose (Ludlow et al., 2020), parental involvement (Antipkina & Ludlow, 2020), teaching for equity (Chang, 2017; Chang et al., 2019), productive engagement (Ludlow et al., 2019), engagement in later life activities (Ludlow et al., 2014), faculty availability outside of class (Reynolds, 2020), and teachers' promotion of sociocultural integration (Báez Cruz, 2021); details focused on the methodology itself have also been recently published (Ludlow et al., 2021; Ludlow et al., 2020).

Because of the rich and comprehensive nature of scenario-based scales, the scenarios are, by nature, multi-barreled, meaning that each scenario measures multiple facets within a particular construct. In Classical Test Theory (CTT), this is something that is avoided during item development, as it can be difficult to identify what a participant is responding to in a double- or multi-barreled item. With scenario-based scales, however, scenarios are constructed utilizing sentence mapping (described in the next section) to reflect multiple parts of the construct. This results in scenarios that are multi-barreled but which focus on a single construct. To deal with the fact that scenario-based scales are, by definition, multi-barreled, careful construction of the response options for participants is necessary. It is also very useful to provide participants with a practice, or “start-up”, item to help participants understand how to engage with this new type of survey (Ludlow et al., 2014). Previous studies have shown that there is often a start-up effect for participants with scenario-based scales, as participants are learning how they should engage with the scenarios and the response scales (Ludlow et al., 2018).

The process for designing scenario-based scales consists of the following steps (Ludlow et al., 2014; Ludlow et al., 2019; Ludlow et al., 2020), as displayed in Figure 2.1:

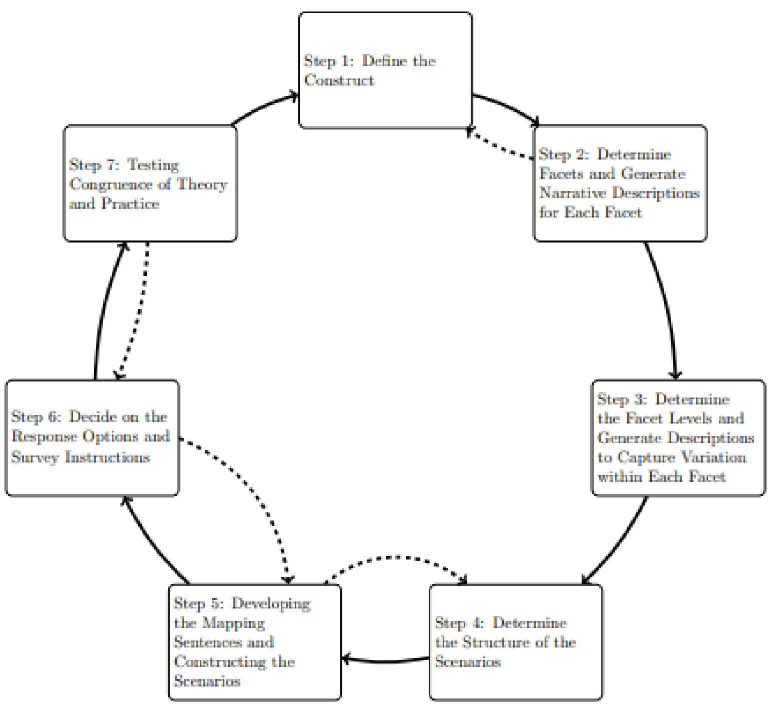
- define the construct
- determine the facets and generate narrative descriptions for each facet
- determine the facet levels and generate descriptions to capture variation within each facet
- determine the structure of the scenarios
- develop the mapping sentences and construct the scenarios
- decide on the response options and survey instructions
- test congruence of theory and practice

Ludlow et al. (2020) note that although these steps are presented sequentially, it is common to move backwards and forwards between certain steps as the instrument development proceeds and more detailed understandings of the construct emerge.

Specifically, it is possible that those utilizing the RGS methodology may move back and forth between defining the construct and determining the facets, as well as between determining the structure of the scenarios and developing the mapping sentences/constructing the scenarios, between developing the mapping sentences/constructing the scenarios and deciding on the response options/survey instructions, and between deciding on the response options/survey instructions and testing congruence of theory and practice (Ludlow et al., 2020). Each of these steps is described in general terms below. A detailed description of each design stage specific to my dissertation is described in the methodology section in Chapter 3.

Figure 2.1

RGS scenario development process (Ludlow et al., 2020)



The definition of the construct consists of a comprehensive literature review to fully define the area under investigation, as well as consultation with content experts (Ludlow et al., 2020). This literature review drives the definition of the construct under measurement, which provides the basis for all future steps in the design process. The definition of the construct that emerges from the literature review is then utilized to extract the facets of the construct and define the levels of each facet. A facet is a piece of the overarching construct that, when combined with the other facets identified in this process, makes up the construct as a whole.

The facets are extracted by taking the construct definition and identifying the individual pieces (i.e., facets) that make up the definition of the construct. Once the facets have been extracted, levels of each facet are identified and described in detail (typically low, medium, and high levels for each facet). Then the structure of the scenarios is determined by identifying how facets and levels will be combined to form individual scenarios. In some cases, there are too many facets to include a level of each facet in each scenario because the resulting scenario would be too long; in these cases, scenarios may include only a subset of the facets of a particular construct. Additionally, to ensure the length of the entire instrument is not too long, not all combinations of facets and levels are included in the final scenarios. For example, for a construct with five facets and three levels, there are 15 facet/level sentence mappings, which results in 3,003 possible scenario combinations (assuming order does not matter and all five facets are included in each scenario). The decision of which combinations of facets and levels to include in each scenario determines the structure of the instrument.

The descriptions of the facet levels and the structure of the scenarios are then utilized in the sentence mapping process. The sentence mapping process consists of the development of the particular sentences that will be used in each scenario. The sentence maps have multiple parts: the stem, which includes the description of the facet level, and the content, which includes the description of the facet. Once the sentence maps have been developed, they are used to create the scenarios by taking sentences from the sentence maps for each facet and level and creating a scenario. Once the scenarios are constructed, response options are developed, along with specific instructions for how participants should engage with the instrument. To test congruence of theory and practice, the scenarios are then shared with a small group of experts to review the face validity of the instrument, and then administered in a small pilot administration to gather preliminary item statistics and test the instructions and response options. The scenarios, instructions, and response options are revised as needed based on the preliminary item statistics and pilot feedback and then the scenarios are administered in a full administration.

The next section consists of the literature review of peer-reviewed models of DDDM/inquiry cycles that define the construct of using data to inform classroom instruction. This section focuses on step one of the RGS methodology, which is to define the construct.

Peer-Reviewed Models of DDDM/Data Inquiry Cycles

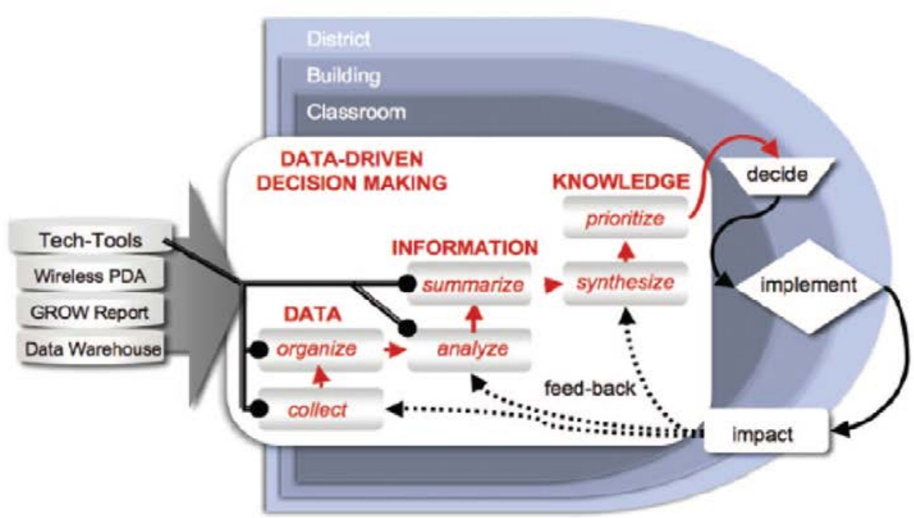
Many different models exist in the literature and in the field of teaching with the main purpose of helping teachers utilize data to inform classroom instruction. The majority of these models are self-described as inquiry cycles. Many organizations and schools have created their own models of inquiry cycles, although they all follow a similar structure. These models of inquiry cycles can be thought of as organizational routines that are utilized in schools, which can help engender efficient action among those in an organization and potentially reduce conflict in doing this work (Spillane, 2012). Rather than utilize a specific data inquiry model as the theoretical basis for this instrument, this section describes nine inquiry cycle models published in peer-reviewed journals, books, or government guides. Based on this review, I identified elements common to all of these models to use as the theoretical basis to develop this instrument. This method allows my instrument to be utilized with any data inquiry cycle model and does not constrain districts or schools who want to use this instrument to the use of a specific inquiry cycle model. Additionally, it helps frame this research in the context of a generalized organizational routine that may be used in schools, and prior research indicates that organizational routines can drive practice, especially in the context of using student achievement data (Spillane, 2012). These nine models in peer-reviewed journals, books, or government guides are each described next.

Mandinach (2012)

The DDDM framework in Figure 2.2 is described as “outlining the cognitive skills that are hypothesized to be involved in DDDM” (Mandinach, 2012, p. 77) and was developed utilizing a cognitive analysis of research on practitioners (Mandinach, 2012). Although the general process of DDDM described by Mandinach (2012) is cyclical in nature as teachers revisit different stages of the process based on the impact they observe of the instructional interventions that they implement, at the beginning of the process these stages are hierarchical in nature. When beginning to use data to inform classroom instruction, teachers must first collect and organize data, which are the cognitive skills described in the "Data" stage in the framework. They then analyze and summarize the organized data in the "Information" stage of the framework. The results from the "Information" stage are used to synthesize and prioritize in the "Knowledge" stage, which results in a decision of a particular intervention to be implemented in the classroom based on the results of the cognitive processes utilized in the "Data", "Information", and "Knowledge" stages. When teachers first engage in this process, the cognitive processes described in the "Data", "Information", and "Knowledge" stages must be utilized in that order to come to an instructional intervention decision. After implementing an instructional intervention and observing the impact, teachers may revisit the "Data", the "Information", or the "Knowledge" stage(s) based on the impact they have observed and their analysis of next steps.

Figure 2.2

Data-Driven Decision Making framework (Mandinach, 2012)



Mandinach and Gummer (2016a)

The data-driven decision making framework presented by Mandinach (2012) resulted in continued research that expanded into the construct of data literacy for teachers, or DLFT (Mandinach & Gummer, 2016a). Mandinach and Gummer (2016a) describe DLFT as "the ability to transform information into actionable instructional knowledge and practices by collecting, analyzing, and interpreting all types of data (assessment, school climate, behavioral, snapshot, longitudinal, moment-to-moment, etc.) to help determine instructional steps" (p. 14). The overall construct of DLFT is more comprehensive than an inquiry cycle only and includes subject matter content knowledge and pedagogical content knowledge, along with elements of an inquiry cycle, in the framework (Mandinach & Gummer, 2016a). As displayed in Figure 2.3, the DLFT framework displays subject matter content knowledge, pedagogical content knowledge, and knowledge of learners and educational contexts and purposes as inputs into the funnel that filter down to data use for teaching, which encompasses an inquiry cycle (Mandinach & Gummer, 2016a). Mandinach and Gummer (2016a) also mention that an understanding of the curriculum content and the scope and sequence is part of step four of their inquiry cycle.

For the purposes of this dissertation, I include only the inquiry cycle that is encompassed by the Data Use for Teaching component of the DLFT framework in this section, while acknowledging that subject matter content knowledge, pedagogical content knowledge, and knowledge of learners and educational contexts and purposes are certainly important inputs into this process. I argue, however, that the inquiry process can be measured independently, given that the audience for this instrument is teachers who will have some level (although varied) of these three additional parts of this framework, and that each individual engaging in an inquiry cycle will bring their own varied experiences and knowledge bases into that inquiry cycle. This inquiry cycle can be engaged in by all, understanding that differing levels of these other inputs may alter the outcome of the inquiry cycle based on teachers' skillsets. Therefore, these additional inputs are not included in the measurement of the inquiry cycle itself, as they are not required (although they are helpful) to engage in this process. Given this, these additional inputs are not included in the discussion of the inquiry cycle included in DLFT which follows below.

The Data Use for Teaching domain of DLFT is a five step process that is cyclical, as displayed in Figure 2.4. The first step in this process is Identify Problems/Frame Questions, which consists of identifying and communicating the problem or question under discussion. A key point in this step is the need to involve other participants. Mandinach and Gummer (2016a) also point out that an understanding of student privacy and contextual issues related to the problem or question are also important in this step. During this step, the problem or question under discussion is developed into a question or multiple questions that can be analyzed empirically (Mandinach & Gummer, 2016a).

The second step in this process is Use Data, and requires participants to identify potential sources of data (Mandinach & Gummer, 2016a). Based on these potential data sources, participants need to understand how to generate or collect this data, as well as understand the properties of these data and any data quality issues. Key in this step is understanding what data are appropriate to use for which questions, while also understanding how to generate and collect data that does not yet exist. Teachers must also understand data quality, specifically for the data that they are using during the inquiry process, and understand how to access and analyze these data. Part of analyzing these data requires knowledge on merging datasets, manipulating data, and aggregating/disaggregating data as needed, as well as an understanding of statistics and psychometrics (Mandinach & Gummer, 2016a), although it is important to note that Mandinach and Gummer (2016a) state that what an understanding of statistics and psychometrics is for teachers is an open question.

The third step in this process is Transform Data into Information. In this step, teachers must understand how to interpret data to take the analyzed data from step two and test their assumptions, generate hypotheses connected to instruction, and consider the impact or consequences of their analysis. Key in this step is taking analyzed data and interpreting them within their context (Mandinach & Gummer, 2016a). Mandinach and Gummer (2016a) state, “Without that transformation, data remain something to point towards, but not act on” (p. 52), but also note that this step from analyzed data to interpretation is not well defined in terms of knowledge and skills. When interpreting data, teachers need to understand how data can be represented or displayed, look at trends and patterns, think about causality or correlation (note, Mandinach and Gummer (2016a) only include the term causality in their description, but I believe correlation should be included as it is often difficult to assess causality in an inquiry cycle), synthesize multiple data analyses, summarize and explain data, and then describe inferences and conclusions (Mandinach & Gummer, 2016a). Mandinach and Gummer (2016a) point out that when considering impact or consequences of the analysis, it is important to consider both intended and unintended consequences (with the caveat that it may not be possible to identify all unintended consequences at this step).

The fourth step in this process is Transform Information into Decisions. In this step, teachers decide on their next steps for instruction and act on those, while thinking about and understanding the context for these decisions (Mandinach & Gummer, 2016a). This context includes the scope and sequence of the curriculum, as well as the curriculum content itself. Teachers also engage their understanding of pedagogy during this step. Teachers monitor student performance during the next steps in instruction by not only using data, but also identifying what else students need and making arrangements for that additional instruction as necessary (Mandinach & Gummer, 2016a). For example, reteaching a lesson may require a shift in pedagogy; teachers may build on prior lessons in future lessons to reinforce particular concepts where they have identified that students need additional support.

The fifth and final step in this process is Evaluate Outcomes. In this step, teachers revisit the original question(s) and compare outcomes to the data collected at the beginning of the cycle (Mandinach & Gummer, 2016a). Teachers look at both changes in student performance and changes in classroom practices in comparison to the decision that was made in step four (Mandinach & Gummer, 2016a). The need for iterative cycles is considered in this final step after evaluating outcomes. At this point, teachers determine if the initial issue that prompted the inquiry cycle has been addressed; part of this determination includes assessing any unintended consequences of the decisions made during the inquiry cycle at the student level or classroom level (Mandinach & Gummer, 2016a).

Figure 2.3

Data Literacy For Teachers (DLFT) framework (Mandinach & Gummer, 2016a)

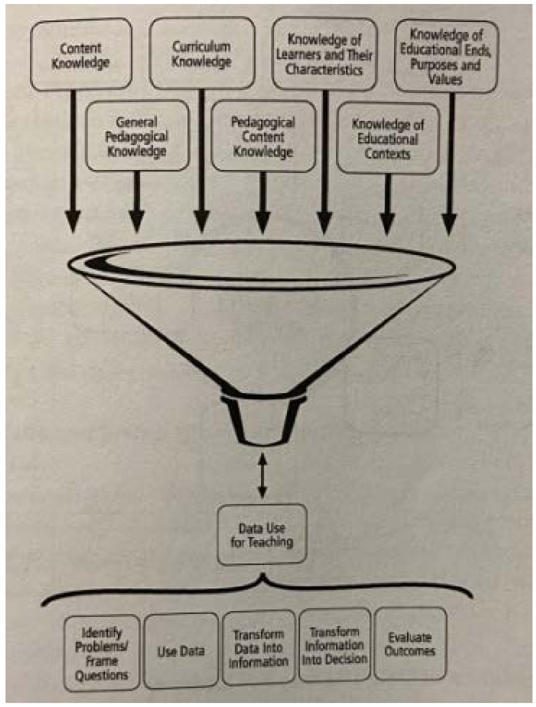
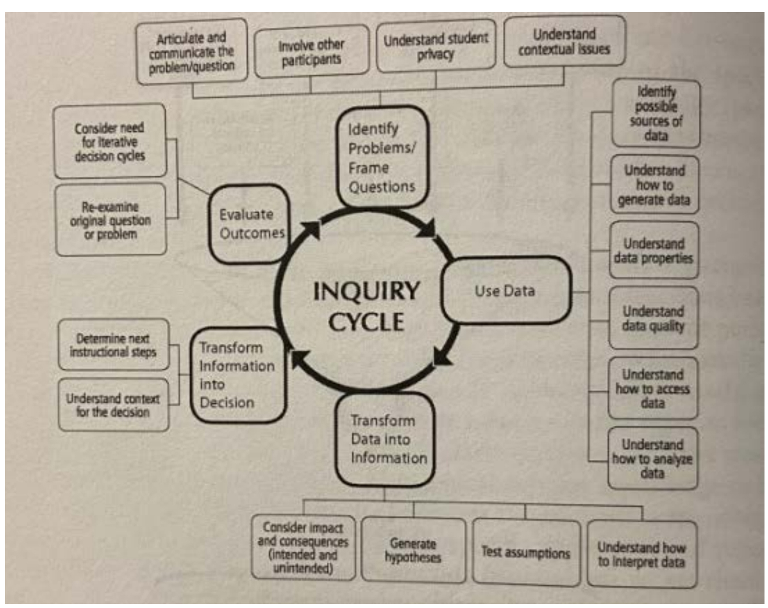


Figure 2.4

Data Use for Teaching component of the DLFT framework (Mandinach & Gummer, 2016a)



Data Wise

The Data Wise improvement process stems from a partnership between the Harvard Graduate School of Education (HGSE) and the Boston Public Schools (BPS) which focused on improving the skills of school leaders to use student assessment results to improve teaching and learning (Boudett et al., 2013). A key feature of the Data Wise process is that it is a collaborative process, meant to be engaged in by teams, and focuses on assessment data as the key data source for this process (although, as noted previously, multiple forms of data should be used in this process and are incorporated into the Data Wise process).

The Data Wise improvement process (Boudett et al., 2013) consists of three phases, with eight total steps. The first phase, Prepare, consists of two steps: Organize for Collaborative Work and Build Assessment Literacy. As displayed in Figure 2.5, the first phase is required at the beginning of this type of work, but once completed, is not necessarily revisited as the inquiry cycle for a particular team continues. During the Prepare phase, "educators lay a foundation for evidence-based decision making, developing the processes and skills they need to invite whole-faculty collaboration in the next two phases" (Boudett & Steele, 2007, p. 8). The first step, Organize for Collaborative Work, sets the stage and lays the groundwork for the team engaging in this process. This step consists of putting together a data team that will engage in the Data Wise cycle, organizing the collection of data that will be used during the cycle, and coordinating the schedules of the team (Boudett et al., 2013). The second step, Build Assessment Literacy, requires the team to learn basic principles about assessments, such as sampling, discrimination, measurement error, reliability, and score inflation, as well as to understand different ways that scores are reported, to ensure that they are responsible users of the student data they will analyze later in the process (Boudett et al., 2013).

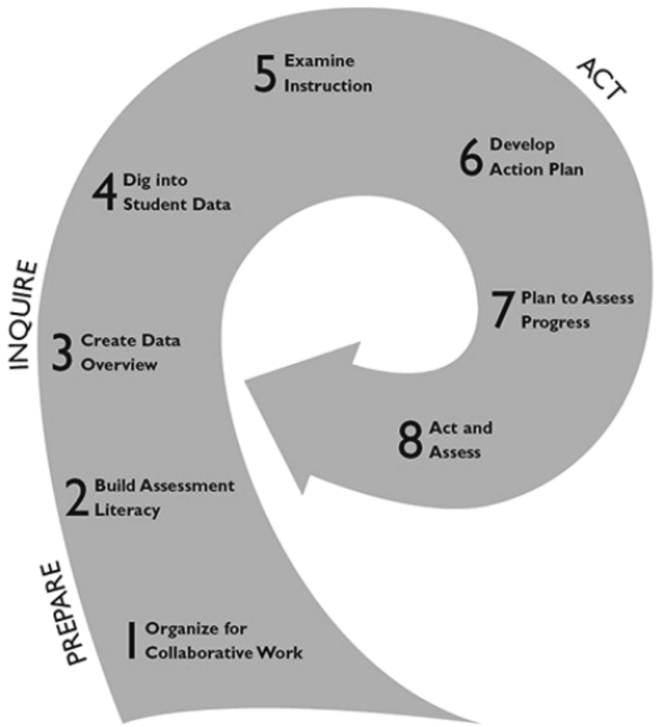
The second phase, Inquire, focuses on exploring “data from a range of sources in an effort to understand students’ learning and teachers’ practice” (Boudett & Steele, 2007, p. 8). This phase consists of three steps: Create Data Overview, Dig into Student Data, and Examine Instruction. During the Create Data Overview step, the team first identifies a focus area for this work, as well as a list of specific questions related to the focus area, and then analyzes and displays the data with the focus area and specific questions in mind (Boudett et al., 2013). A key point made by Boudett et al. (2013) specific to this stage is that assessment data is often presented in tables, which can make it difficult to visualize patterns. Boudett et al. (2013) suggest presenting assessment results related to the focus area in graphical formats. During the Dig into Student Data step, the team uses a variety of other data sources (focused on student academic performance) to gather more information on the learning gaps identified in the Create Data Overview step (Boudett et al., 2013). At the end of the Dig into Data step, the team identifies a learner-centered problem, which Boudett & Steele (2007) define as “a gap in skill or understanding common to many students that, if corrected, would have far-reaching implications for students’ continued academic growth” (p. 8). In the last step in this phase, Examine Instruction, the team looks at the instruction that students have already received in relation to the learner-centered problem, with the goal of reframing the learner-centered problem as a problem of practice, specific to teaching (Boudett et al., 2013). The problem of practice is defined as “an instructional challenge that teachers believe to be worth tackling collectively" (Boudett & Steele, 2007, p. 8) and is a statement rather than a question.

The third phase, Act, consists of three steps: Develop an Action Plan, Plan to Assess Progress, and Act and Assess. During this third phase, the members of the team "develop and carry out a plan for addressing the problem of practice and improving student learning" (Boudett & Steele, 2007, p. 8). In the Develop an Action Plan step, the team picks an instructional strategy that it believes can be a solution to the problem of practice, with the goal of improving student outcomes (Boudett & Steele, 2007). Important in this step, the team also develops a professional development plan for this instructional strategy (Boudett & Steele, 2007), with the understanding that teachers may need training and support in implementing this new instructional strategy. This step requires the action plan to be put in writing to document the roles and responsibilities of each team member, as well as to list the steps that teachers will take (Boudett et al., 2013). In the Plan to Assess Progress step, the team identifies goals for student learning in the short-, medium-, and long-term, along with indicators to track progress towards meeting those goals (Boudett & Steele, 2007). The team also identifies how short-term data, such as classwork, homework, classroom observations, and student conferences, will be used to measure progress toward short-term goals; how medium-term data, such as school-based assessments like benchmark or interim assessments, will be used to measure progress towards medium-term goals; and how long-term data, such as statewide assessment data, will be used to measure progress towards long-term goals (Boudett et al., 2013). The main purpose of this step is to be able to identify if the new instructional strategy identified in the previous step made a difference for the learner-centered problem. Finally, in the Act and Assess step, the team (or the school, depending on the scope of the learner-centered problem and the problem of practice) enacts the professional development and strategy developed in the Action Plan step and monitors progress towards the goals they set in the Plan to Assess Progress step (Boudett & Steele, 2007). Communication is key in this step, both prior to implementation and during implementation (Boudett et al., 2013). A key point for this step is that the team should carefully monitor student outcomes during this step, so that they are able to adjust teaching and instructional strategies as needed based on student outcome data (Boudett & Steele, 2007).

Once the Act phase is complete, the team continues the inquiry process by returning to the Inquire stage, as the Prepare stage is only necessary once with each individual team. Boudett et al. (2013) explain, "Three things you can do once you have made it around the steps of the improvement cycle are to celebrate success, revisit your criteria and raise the bar, and plan how to keep the work fresh and ongoing" (p. 185).

Figure 2.5

Data Wise inquiry cycle (Boudett et al, 2013)



Practitioner Data Use Workshop

The Practitioner Data Use Workshop (Bocala et al., 2014) provides a detailed toolkit "designed to help practitioners develop skills in collaborative, data-driven inquiry and instructional decisionmaking" (p. Introduction-1) and recommends that teams of workshop participants engage with the toolkit, rather than individuals. The toolkit was created by reviewing research, tools, and resources focused on data inquiry in education (Bocala et al., 2014). The main focus of the toolkit is on the data inquiry cycle and how it can be applied in educational settings (Bocala et al., 2014). The toolkit provides two models of inquiry cycles, which are described below.

National Forum on Education Statistics (2012)

The first model included in the toolkit, displayed in Figure 2.6, is a data inquiry cycle developed by the National Forum on Education Statistics (2012) as part of a guide designed to provide information on "the knowledge, skills, and abilities needed to identify, access, interpret, and use data to improve instruction in classrooms" (p. 1). There are five stages to this model. The first stage, Seek Information, focuses on identifying key questions and "refers to the process of finding the right data to address the specific information needs at hand" (National Forum on Education Statistics, 2012, p. 4). These key questions typically arise from the identification of gaps between what a person currently knows and what that person would like to know about a specific area (which could be related to any unit in an educational system, such as students, teachers, curricula, etc.) (National Forum on Education Statistics, 2012). Key steps in this stage include the recognition that information is needed to inform a decision, the definition of key questions (which are meaningful and achievable), the assessment of relevant available data as well as the identification of necessary data that is currently not available, and the identification of any possible barriers to finding the data that is not currently available and ways to remove those barriers (National Forum on Education Statistics, 2012). An important point for this stage is that the key questions should be carefully worded and well-defined, as opposed to broad questions, as concise and well-defined questions can help identify data sources to answer the question(s) (National Forum on Education Statistics, 2012).

The second stage in this model, Access and gather data, is focused on accessing and gathering the data identified in the first stage. In many cases, the data needed to answer the key questions already exist, although sometimes the data needs to be collected. The main steps in this stage include pulling the data relevant to the key questions, which may require accessing existing data or collecting new data, and learning about the data to understand what variables are included, the limitations of the data, the timestamp of the data (when it was collected) and any formatting that is included (National Forum on Education Statistics, 2012). The step focused on learning about the data in this stage is extremely important to ensure that the data are used in appropriate ways: for example, understanding which variables are included and what they mean is necessary before any analysis can be conducted, as the analysis will depend on the variables included in the dataset.

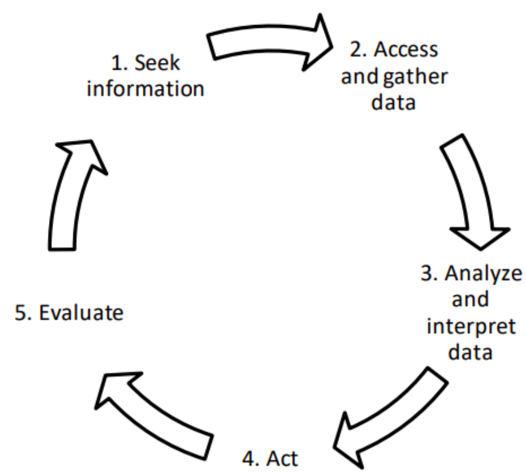
The third stage of this model, Analyze/Interpret Data, focuses on obtaining "sound evidence to inform decisionmaking and action" (National Forum on Education Statistics, 2012, p. 5), which comes from the data collected in the second stage. In this stage, specific steps include the formatting of data so that analysis and interpretation can occur and the determination of data constraints, which are derived from an understanding of the data obtained in the second stage (National Forum on Education Statistics, 2012). The data constraints described by the National Forum on Education Statistics (2012) include the following: the unit of analysis, the design of the data (how it was obtained, such as random sampling or comparison groups), the timestamp of the data, the purposes of the data (such as formative or summative assessments), the quality of the data, any potential bias in the data, and any possible misuse of the data. The understanding of these constraints for the data being utilized helps the user produce logically or statistically sound conclusions (National Forum on Education Statistics, 2012). A key point about this stage is that the specific analysis will depend on the key questions that are asked, the data and analytic tools that are available, and the analytic skills of the people engaging in this process, but everyone engaging in this model of inquiry will engage in some type of analysis during this stage (National Forum on Education Statistics, 2012). Although not explicitly stated as part of stage three by the National Forum on Education Statistics (2012), they note that the ability to understand the meaning of error in data and the ability to use tables and/or graphs to make meaning from data are part of the knowledge, skills, and abilities needed to effectively analyze and interpret data.

The fourth stage, Act, focuses on taking action based on the analysis of data that occurs in the third stage of this model (National Forum on Education Statistics, 2012). A key point in this stage is that ethics play an important role in the determination of appropriate actions based on the data and the analysis (National Forum on Education Statistics, 2012). Although the National Forum on Education Statistics (2012) provides some examples of inappropriate uses of data (such using a single test score to evaluate a student) and appropriate uses of data (such as using prior summative assessment scores to identify students who need additional instruction), this stage does not include a detailed description of how users might identify appropriate actions based on the analysis conducted in stage three. This is often cited as the possible reason that data inquiry cycles do not result in increased student achievement, as the determination of actions to take based on analysis can be difficult without pedagogical support (Neuman, 2016).

The fifth stage in the cycle, Evaluate, focuses on evaluating whether the action that occurred in the fourth stage has resulted in any changes (National Forum on Education Statistics, 2012). Important questions to be addressed during this stage include the extent to which the initial issue that arose in the first stage was addressed, whether any new concerns have surfaced, which parts of the initial issue are now understood and which parts require additional analysis, and whether any new gaps have been identified (National Forum on Education Statistics, 2012). Importantly, if any new concerns have arisen, new gaps have been identified, or especially if the initial issue has not been fully addressed, the data cycle begins again from step one.

Figure 2.6

The Cycle of Data Use framework (National Forum on Education Statistics, 2012)



Kekahio and Baker (2013)

The second model in the toolkit comes from a facilitation guide designed to showcase how teams of educators can apply data for "strategic actions" in data-informed conversations (Kekahio & Baker, 2013, p. i). This model consists of five steps that are involved in these data-informed conversations: setting the stage, examining the data, understanding the findings, developing an action plan, and monitoring progress and measuring success (Kekahio & Baker, 2013). Each of these steps is described in detail below and displayed in Figure 2.7. Kekahio and Baker (2013) frame this work as a team-based activity rather than an individual teacher activity.

The first step in Kekahio and Baker's (2013) model, setting the stage, focuses on identifying the question that will be the focus of the data-informed conversation, identifying the information that will be needed to answer the question, and determining if that information is available. Kekahio and Baker (2013) acknowledge that teams may start this process with broad, simply framed questions to identify a particular issue, but that the question needs to become specific in order to progress with this process. Once the question has been identified, the team identifies particular data sources to answer the question and determines if these data sources currently exist (Kekahio & Baker, 2013). Kekahio and Baker (2013) point out that in some cases, data may already exist in analyzed forms, but in other cases, teams must obtain raw data to answer their questions. Additionally, in many cases, multiple sources of data are necessary to answer the question, and sometimes, data necessary to answer the question do not currently exist and must be collected (Kekahio & Baker, 2013).

The second step in this model, examining the data, focuses on identifying the patterns in and/or making observations from the data, while also identifying limitations in the data (Kekahio & Baker, 2013). Kekahio and Baker (2013) note that identifying patterns and making what they refer to as "snapshot" observations are the first step in answering the question; these snapshot observations are observations that are made from an initial examination of the data. These snapshot observations and/or patterns can be classified as strengths, which indicate success, or challenges, which indicate "something is blocking improvement or higher achievement" (Kekahio & Baker, 2013, p. 6). These strengths and challenges must be specific, factual, and related to the question (Kekahio & Baker, 2013). Additionally, during this step, teams must identify limitations in the data to ensure that any analysis and conclusions are appropriate (Kekahio & Baker, 2013). Specifically, Kekahio and Baker (2013) suggest that teams must identify whether different data sources can be compared and discuss whether the data permits robust conclusions. This is particularly important for establishing causality (as stated in Kekahio & Baker, 2013), especially when data has not been collected or maintained in a format that allows group comparisons or rigorous analysis (Kekahio & Baker, 2013).

The third step in the model, understanding the findings, focuses on identifying possible causes, or driving factors, for the patterns or observations identified in step two (Kekahio & Baker, 2013). Kekahio and Baker (2013) suggest that a discussion among the team about one or two of the key challenges identified in step two is helpful to understand why the challenge(s) are occurring. The key challenge(s) that become part of the discussion should be actionable (meaning that they have driving factors that the team can address or influence) and be aligned with district priorities (Kekahio & Baker, 2013). Kekahio and Baker (2013) provide protocols to help teams with this step, and specifically suggest that teams can ask a series of "why" questions that are answered with "because" responses to help identify driving factors (p. 7). The answers to these "why" questions come from educated guesses (Kekahio & Baker, 2013), and these answers become the next "why" question until the team comes to a potential driving factor. The answers that lead to each potential driving factor are then investigated by determining if the data support the responses that led to the driving factors (Kekahio & Baker, 2013). If the data supports these answers for the potential driving factor, that driving factor can remain; if the data does not support any of these responses, the team should revisit their answers to identify another potential driving factor (Kekahio & Baker, 2013).

The fourth step in the model, developing an action plan, focuses on making an effective plan to address the issue identified (Kekahio & Baker, 2013). Kekahio and Baker (2013) point out that an effective plan should have both short-term objectives and a long-term goal focused on reducing the challenges identified in step three and increasing successes; they suggest that the SMART goal structure (specific, measurable, attainable, relevant, and timely) should be utilized to create these goals. Kekahio and Baker (2013) suggest that these SMART goals can be developed through brainstorming sessions focused on identifying possible strategies and actions to reach the goals. Any possible strategies could be considered with the following lenses: time (is the strategy possible within the given timeframe), resources (are there sufficient resources for the strategy), relevance (in relation to the goal), and data availability (in relation to monitoring the action plan) (Kekahio & Baker, 2013). Key stakeholders who will be involved in the plan should be identified and involved in the process to ensure success (Kekahio & Baker, 2013).

The last step in the model, monitoring progress and measuring success, focuses on keeping the action plan from step four on track and determining if progress is made on the identified issue (Kekahio & Baker, 2013). In this step, the team monitors their action plan to ensure it is being implemented as intended and that the goals are being attained (Kekahio & Baker, 2013). Kekahio and Baker (2013) state that in this stage, the team should collect data from the same source(s) utilized to identify the challenge in step two to avoid any issues with differences in the way data has been collected from different sources.

Although Kekahio and Baker (2013) do not describe their model as an inquiry cycle, it is described as such in Bocala et al. (2014) and presented in their toolkit as an option for a data inquiry cycle. Thus, although Kekahio and Baker (2013) do not describe their model as cyclical, most teams using their model would return to step one at the end of step five with new, refined questions based on the results of the evaluation conducted in step five or new questions that arise from the evaluation conducted in step five.

Figure 2.7

Five steps in data-informed conversations framework (Kekahio & Baker, 2013)

This guide describes five steps in data-informed conversations that lead to strategic decisionmaking and action:

1. *Setting the stage*. What question is to be addressed in this data-informed conversation? What information is needed to answer the question? Is the information available?
2. *Examining the data*. What patterns do the data reveal, or what "snapshot" observations can be made about the question?
3. *Understanding the findings*. What are the possible causes for the patterns?
4. *Developing an action plan*. How can a data team create an effective plan for addressing the issue?
5. *Monitoring progress and measuring success*. How can a data team know whether progress is being made on the issue?

Deming's PDSA Cycle (2018)

The PDSA Cycle (Deming, 2018) was originally developed as "a flow diagram for learning, and for improvement of a product or a process" (p. 91). This cycle was originally developed by Deming in 1950 in a booklet that is now out of print, as described by Deming (Deming, 2018, p. 91), and is included in his book focused on a new style of management for industry, government, and education. The PDSA cycle is presented in the chapter titled “Management of People” (Deming, 2018), where a main goal of the chapter is to “help people to optimize the system so that everybody will gain” (p. 86). Although this model was not specifically designed as an inquiry cycle for education, it has been used as such by others (Bernhardt, 2004; Tichnor-Wagner et al., 2017) and is described by Deming as a process to be used for process or product improvement, similar to other inquiry cycles; therefore it is included in this review. The PDSA Cycle is displayed in Figure 2.8.

The PDSA Cycle stands for Plan-Do-Study-Act. The first step, Plan, begins when a person thinks about improving a product or a process. Deming (2018) describes this as the “0-th stage, embedded in Step 1” (p. 91). This prompts a plan for a “test, comparison, experiment” (Deming, 2018, p. 91). This plan may require choosing among multiple options for testing, and Deming (2018) suggests that the decision of which option to test should be based on an analysis of the one whose probable outcome seems most likely to result in “new knowledge or profit” (p. 91). Another key point for this stage is that brainstorming is important to try to avoid the need to backtrack (as much as is possible) in a later step, indicating that this step should not be rushed (Deming, 2018). Although not explicitly stated by Deming (2018), this step requires that the plan be measurable (Tichnor-Wagner et al., 2017), and that those engaging in this process plan to measure those outcomes when they arrive at Step 3 (described below).

The second step, Do, consists of implementing the test, comparison, or experiment based on the plan identified in Step 1 (Deming, 2018). A key point in this step, according to Deming (2018), as that it is preferable to conduct the Do step on small scale. This is emphasized in Tichnor-Wagner et al. (2017), who describe the use of PDSA cycles in school settings as multiple tests of small changes. Again, although not explicitly stated by Deming (2018), information should be gathered on what happened during the test and as a result of the test (Tichnor-Wagner et al., 2017).

The third step, Study, focuses on studying the results from the second step, Do. During this step, the results should be examined and compared to the predictions and expectations from Step 1, Plan (Deming, 2018). If the results do not meet the predictions and expectations from Step 1, an analysis of what could have gone wrong should be undertaken, as well as an examination of whether the predictions from Step 1 are possibly incorrect (Deming, 2018).

The fourth and last step, Act, consists of making a decision on how to proceed. The change can be adopted, abandoned, or the cycle may begin again. If the third step, Study, ends with results that meet the predictions or expectations from Step 1, the change should be adopted on a larger scale (Deming, 2018). If the third step ends with results that do not meet the predictions or expectations, the fourth step should result in abandoned changes or another PDSA cycle with changes or modifications to the first one. If it seems likely that the predictions from Step 1 are incorrect, Deming recommends starting over (2018). If the PDSA cycle begins again, it is likely that different environmental conditions, materials, or people should be considered (Deming, 2018).

If the change is adopted on a larger scale, Tichnor-Wagner et al. (2017) explain that the PDSA cycles repeat as the change is adopted on a larger and larger scale. Each cycle provides new knowledge to those implementing it and may require some adjustment to continue to scale up this change. In this way, the PDSA cycles can be utilized as a cycle of inquiry (Tichnor-Wagner et al., 2017). Bernhardt (2004) utilizes a version of the PDSA cycle in her book focused on the use of data analysis for school improvement, pointing out that a key principle in the use of these cycles is the need for focused data analysis, when "schools are clear on their purpose and clear on what they expect students to know and be able to do, and when students and the community are aware of these expectations" (p. 14). The use of these cycles without this focus can result in what Bernhardt (2004) refers to as "random acts of improvement" that do not focus on guiding principles, such as vision, mission, purpose, values and beliefs, and/or standards, for the district or school.

Figure 2.8

PDSA Cycle (Deming, 2018)

Hirsh and Crow (2018)

Hirsh and Crow (2018) provide a learning team cycle in their book, *Becoming a Learning Team: a Guide to a Teacher-Led Cycle of Continuous Improvement*, which focuses on collaborative learning teams of teachers working together to produce improvements in teaching and learning. Hirsh and Crow (2018) utilize the learning team cycle, which is an inquiry cycle, as the tool to use for continuous improvement in collaborative learning for teachers. As Hirsh and Crow (2018) describe, "the standards embody a belief that a *learning team cycle* is the day-to-day means for embedding professional learning in classrooms, thus supporting teachers when they need it most" (p. 16, italics in original). Hirsh and Crow (2018) acknowledge that many models like this exist within education and provide some examples, while ultimately promoting the learning team cycle as their model of choice.

The five stages in the learning team cycle (Hirsh & Crow, 2018) consist of analyze data, set goals, learn individually and collaboratively, implement new learning, and monitor, assess, and adjust practice. Although not included as part of the learning team cycle, Hirsh and Crow (2018) state that before entering the learning team cycle, teams of teachers need to identify an area of focus, which they state should be tied to district and school goals, as well as potentially tied to strategic priorities, goals for improvement at a system-level, school-level, and individual-level, and school improvement plans (p. 17). The area of focus should be further refined by goals for grade-level or content-level teams specific to the team of teachers engaging in this process (Hirsh & Crow, 2018). Then, the team of teachers engages with the learning team cycle to further refine and address the problem of practice (Hirsh & Crow, 2018). Each stage of the learning team cycle is described below and displayed in Figure 2.9.

The first stage of the learning team cycle, Analyze data, examines student and teacher learning challenges using data (Hirsh & Crow, 2018). As Hirsh and Crow (2018) state, "team members analyze data so they can identify and better understand the exact problem they are addressing" (p. 19). Key in this stage is the identification of a more detailed problem of practice on which the team can focus; this may include identifying particular teachers who have more success in this particular area of teaching and learning from whom others can learn (Hirsh & Crow, 2018). A concrete outcome of this first stage is that teachers on the team "will access, examine, and interpret data to write data summary statements" (p. 19).

The second stage of the learning team cycle, Set goals, focuses on stating shared goals for student and teacher learning (Hirsh & Crow, 2018). A key point in this step is the need to set teacher learning goals, in addition to student learning goals, as the way to drive progress towards student learning goals. Hirsh and Crow (2018) point out that teacher learning goals are necessary at this stage to ensure that teachers do not resort to teaching in the same ways as before or implement new strategies without fully understanding how to do so. Individual or group self-assessment is suggested as the mechanism to help with setting goals (Hirsh & Crow, 2018). Student goals are written as SMART goals (specific, measurable, attainable, results-based, and timebound) in this stage, and teachers write a learning plan that includes classroom strategies and a timeline for learning about and then implementing those strategies (Hirsh & Crow, 2018).

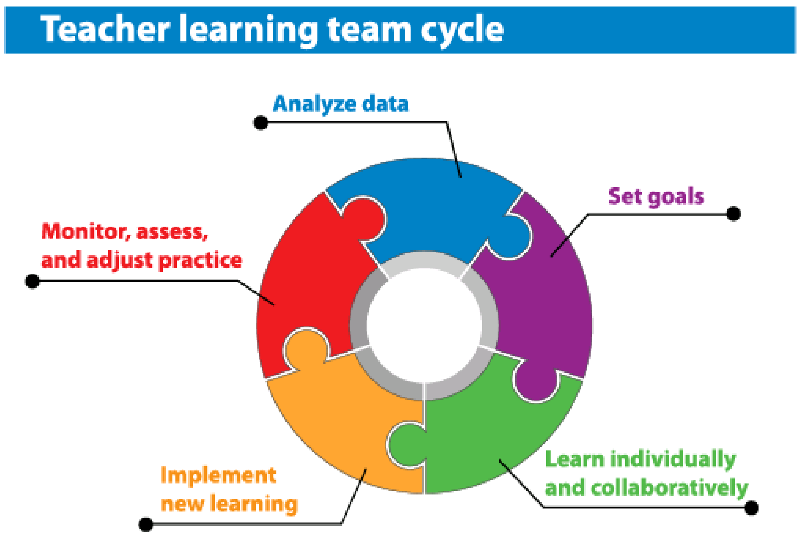
The third stage of the learning team cycle, Learn individually and collaboratively, focuses on attaining new knowledge and skills for teachers and on examining assumptions, aspirations, and beliefs for team members (Hirsh & Crow, 2018). Team members keep the learning goals for students and teachers at the front of their minds during this stage to ensure that they select experiences appropriate for the goals that they have set (Hirsh & Crow, 2018). During this stage, the team must identify learning styles and areas of expertise for each team member to allow them to learn both individually and as a group; team members must also identify areas in which the team needs support that can be obtained by reaching out to other individuals or identifying other resources (Hirsh & Crow, 2018). Hirsh and Crow (2018) point out that during this stage, team members connect what they are learning to application and think ahead to how they will implement what they are learning in their practice with students.

The fourth stage of the learning team cycle is Implement new learning and consists of the application of team members' learning in their classrooms (Hirsh & Crow, 2018). During this step, support from coaches and peers is important as teachers begin to implement new learning and strategies with students (Hirsh & Crow, 2018). Hirsh and Crow (2018) state that many teachers may experience "implementation dip", or a decline in outcomes when new strategies are first implemented (p. 22). The support and feedback from coaches and peers can help teachers work through this implementation dip (Hirsh & Crow, 2018).

The final stage of the learning team cycle, Monitor, assess, and adjust practice, focuses on gathering and examining evidence to determine if their new instructional strategies and learning has had the desired impact on student learning (Hirsh & Crow, 2018). Team members gather evidence of both the implementation of their learning, as well as the outcomes of that learning through formative and summative assessments designed to measure each student learning goal (Hirsh & Crow, 2018). The team analyzes this evidence to identify the frequency of the use of their instructional strategies, as well as whether these strategies are helping to achieve their goals (Hirsh & Crow, 2018). Based on this evidence, the team may need to adjust assumptions and revisit their instructional strategies if the evidence does not show improvement towards student learning goals, or they may engage in a new learning cycle if the evidence shows improvement towards student learning goals (Hirsh & Crow, 2018).

Figure 2.9

Teacher learning team cycle (Hirsch & Crow, 2018)



Dana and Yendol-Silva (2003)

Dana and Yendol-Silva (2003) describe teacher inquiry in their book as the means to "transforming the profession" of teaching (p. 2), with a focus on bringing teacher voices into educational reform and as a tool for professional growth. Dana and Yendol-Silva (2003) provide a framework for teachers to utilize when engaging in teacher inquiry, which consists of identifying questions or 'wonderings', collecting data, analyzing data in connection to literature, changing practice, and sharing findings. Key to their framework is the need for inquiry to be intentional and visible, in contrast to reflection which often occurs internally and on the fly (Dana & Yendol-Silva, 2003). This points to the systematization of the process. Although Dana and Yendol-Silva (2003) elaborate on each stage of their framework with text, they do not provide a graphic for this framework.

Stage one consists of identifying a question or 'finding a wondering' (Dana & Yendol-Silva, 2003). This stage requires teachers to identify questions that they can explore that come from their experiences in their own classrooms and with their own students where they have experienced "felt difficulties" (Dana & Yendol-Silva, 2003, p. 14). Dana and Yendol-Silva (2003) note that five elements of teaching must also be considered by teachers when engaging in inquiry: "the child, the context, the content, the acts of teaching, and the teacher's own beliefs or dispositions" (p. 14). The labeling of 'felt difficulties' in consideration of these five elements results in eight passions that can be explored to identify questions or 'wonderings': a child; curriculum; content knowledge; teaching strategies/techniques; beliefs about practice; personal/professional identity; social justice; and context (Dana & Yendol-Silva, 2003, p. 16). Dana and Yendol-Silva (2003) also suggest that connecting with colleagues can help teachers who are struggling to identify a 'wondering'. Finally, Dana and Yendol-Silva (2003) suggest that a question or 'wondering' should be a "real" question (meaning a question where the teacher truly does not know the answer), that teachers write open-ended questions, and that teachers make sure that the question can be explored with methods available to the teacher (p. 47).

The second stage of teacher inquiry focuses on identifying ways that a teacher can collaborate with others in the inquiry process. Dana and Yendol-Silva (2003) note that collaboration is key and that it is not a question of whether a teacher should collaborate with others, but rather how a teacher will collaborate with others in this process. They explain that collaboration is important because this type of work can be demanding on top of a teacher's existing tasks, and collaboration can provide a source of support that can be helpful (Dana & Yendol-Silva, 2003). They also note that discussing inquiry with other teachers can lead to new learning and help teachers question their beliefs (Dana & Yendol-Silva, 2003). Finally, Dana and Yendol-Silva (2003) state that collaboration can help spur change, as change can be uncomfortable and the support of others can help inspire change. Suggestions for collaboration in the inquiry process include shared inquiry (where teachers engage in inquiry together around a shared question), parallel inquiry (where teachers engage in separate inquiry questions focused on different topics, but work together during the process to provide support to each other), intersecting inquiry (where teachers engage in individual inquiry questions focused on the same topic), and inquiry support (where one teacher engages in inquiry and other teachers who are not engaging in their own inquiry questions provide support, like a critical friend) (Dana & Yendol-Silva, 2003).

The third stage of teacher inquiry is the development of a research plan (Dana & Yendol-Silva, 2003). The research plan consists of identifying a data collection strategy or strategies and deciding on a plan for the inquiry process (Dana & Yendol-Silva, 2003). Dana and Yendol-Silva (2003) note that data collection should be based on the daily classroom life. They offer seven possible strategies for data collection: fieldnotes, where teachers take notes on what they observe happening in the classroom; documents/artifacts, where teachers identify and collect specific pieces of paper or documents from their classrooms; interviews, where teachers interview students either spontaneously or in a pre-planned format; focus groups, where teachers have discussions with groups of students; reflective journals, where teachers have students write their reflections on learning and/or teachers write their own reflections on learning; surveys, where teachers survey students to capture opinions, thoughts, or knowledge; and literature as data, where literature on the ‘wondering’ is gathered to help teachers make sense of the subject (Dana & Yendol-Silva, 2003). Dana and Yendol-Silva (2003) note that most teacher inquiry cycles involve more than one data collection strategy and suggest that teachers set a specific time frame for the inquiry cycle.

The fourth stage of teacher inquiry is data analysis (Dana & Yendol-Silva, 2003). Dana and Yendol-Silva (2003) note that many teachers may feel overwhelmed by the amount of data that they have collected when they get to this stage and suggest coding and memoing as systematic ways to analyze data. Dana and Yendol-Silva (2003) suggest that teachers should begin by going through all of their data to obtain a descriptive sense of the dataset. Then, teachers move to the sense-making stage where they begin to think through patterns, outliers, and other things that stand out; this process may go through multiple iterations (Dana and Yendol-Silva, 2003). Dana and Yendol-Silva (2003) explain that the next step of the analysis process is the interpretive step, where teachers state what they have learned and what it means. The final step of the analysis process focuses on implication questions, where teachers ask questions about what they have learned about themselves, about their students, about their school(s), and about the implications for their practice, including the changes they plan to make (Dana & Yendol-Silva, 2003).

The fifth stage of teacher inquiry is the inquiry write-up, where teachers write-up the process and what they have learned, as a way to solidify knowledge and to help teachers clarify their learning (Dana & Yendol-Silva, 2003). Although they note that this write-up can take many forms, they provide an example that includes the following: background information, design of the inquiry (including the procedures, data collection strategies, and data analysis techniques), description of the resulting learning that is supported by data, and a conclusion (Dana &Yendol-Silva, 2003).

The sixth, and final, stage of teacher inquiry is making the inquiry public (Dana & Yendol-Silva, 2003). Dana and Yendol-Silva (2003) note that making inquiry public is the most likely way to spur change, and suggest that publicizing inquiry requires teachers to clarify their own thinking and allows other teachers to ask questions, which can help those who engaged in the original inquiry to further clarify and refine their own thinking. Dana and Yendol-Silva (2003) suggest sharing the final written product with colleagues, submitting the final written product to a journal focused on teacher-researchers, and sharing the final written product online. They also suggest sharing the final product through presentations, movies, posters, or at conferences.

Hamilton et al. (2009)

In a 2009 Institute of Education Sciences (IES) practice guide, Hamilton et al. (2009) describe five recommendations as a framework for effective use of data to inform instructional decisions. Of these five recommendations, two of them are specific to implementing this type of work in a classroom, while another two focus on implementing this work in a school. The fifth recommendation is focused on improving district data systems to support this type of work. In addition to these recommendations, Hamilton et al. (2009) note that to successfully engage in this type of work, a data system should pull data from multiple sources, a data team should be built in each school to encourage this work, and collaboration among teachers focused on using data to improve student achievement should be encouraged for success. Hamilton et al. (2009) also recommend that students be taught how to use their own data to set their own goals. The two recommendations specific to implementing this type of work in the classroom are "Make data part of an ongoing cycle of instructional improvement" and "Teach students to examine their own data and set learning goals" (Hamilton et al., 2009, p. 8). The first recommendation, make data part of an ongoing cycle of instructional improvement, includes a data use cycle (or cycle of inquiry), which is described here.

The data use cycle described by Hamilton et al. (2009) provides a systematic process for teachers to use data to gather evidence for their instructional decisions in an effort to "improve their ability to meet students' learning needs" (p. 10). As shown in Figure 2.10, this cycle consists of three steps. Although Hamilton et al. (2009) state that most teachers will begin this cycle by collecting and preparing data, teachers can enter this cycle at any point.

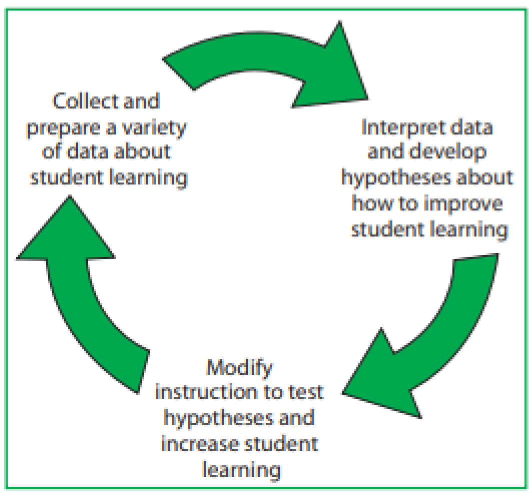
At the collecting and preparing data stage, teachers collect and prepare data from a variety of sources, which includes (but is not limited to) annual, interim, and classroom assessment data (Hamilton et al., 2009). Hamilton et al. (2009) note that it is important in this stage to identify specific questions about student achievement so that teachers can effectively determine the types of data they need to gather and prepare. The identification of specific questions allows teachers to narrow their focus on specific types of data that will help them answer those questions, and tying specific questions to schoolwide goals can be useful (Hamilton et al., 2009). Hamilton et al. (2009) note that classroom assessment data may include assessments based on the curriculum, chapter tests, and classroom-based projects, and data not specific to achievement (such as attendance, cumulative files, and behavioral data) can also be incorporated. A key point is that a variety of data should be used rather than relying on a single data source to help mitigate limitations of individual data sources (Hamilton et al., 2009). In this stage, teachers must consider the strengths and limitations of their data sources, including when the data was collected, and then prepare the data in a way that allows for interpretation. Preparation of the data includes aggregating data in ways that relate to the specific questions that teachers have about student achievement.

Once the data has been prepared, teachers interpret the data and then develop hypotheses about what is contributing to student performance and what specific actions they as teachers can take to meet students' needs (Hamilton et al., 2009). This interpretation of data can happen independently or collaboratively in teams, although it is recommended that this happens in teams to allow sharing of effective practices (Hamilton et al., 2009). Hamilton et al. (2009) note that identifying overall strengths and weaknesses for a classroom is a good place to start, as is identifying individual students' strengths and weaknesses. This identification can help teachers focus instruction and time on content where students need support. A key point in this step is that using multiple data sources for this interpretation (also called triangulation) is important to illuminate areas in which students do need support, rather than identifying something specific to the data source such as issues with particular items on an assessment (Hamilton et al., 2009). During data interpretation, teachers develop hypotheses about ways in which they can improve student achievement (Hamilton et al., 2009). Hamilton et al. (2009) note that good, testable hypotheses are based on the existing data, focus on changes to instruction or curriculum that are likely to improve students' learning, and are testable using data that can be collected during or after an intervention. Hypotheses should include identification of an intervention or instructional modification and the effect expected from that intervention, and teachers should ensure that the effect they anticipate can be measured while also identifying comparison data (Hamilton et al., 2009). Although not specifically stated by Hamilton et al. (2009), it is implied that teachers should also plan to collect the data that they will use to measure the effect of their intervention at this point if it requires data that is not typically collected.

In the third stage, teachers change their instruction to test their hypotheses and the cycle begins again when they collect and prepare student data to identify if their hypotheses were correct and their instructional changes resulted in changes in student learning (Hamilton et al., 2009). Hamilton et al. (2009) provide some suggestions for instructional changes, which include, but are not limited to, providing more time for topics where students are struggling, reordering the curriculum to focus on essential skills, grouping or regrouping students based on need, trying new ways of teaching particular concepts, realigning performance expectations, or better aligning curriculum between grade levels. They note that if the intervention was identified individually (not collaboratively), teachers may find it useful to gather feedback from peers on their chosen intervention before implementation. Teachers should keep notes during the implementation of the intervention on the response of students and their own reflections while carrying out this intervention for future reference (Hamilton et al., 2009). Once the intervention is complete, teachers return to the collect and prepare stage to evaluate the intervention. They interpret the data from the intervention and decide on their next steps, which may include continuing the current intervention, modifying the intervention, or trying a new approach (Hamilton et al., 2009). Hamilton et al. (2009) point out that change can take time and that teachers should allow sufficient time for an intervention to effect change before discarding it.

Figure 2.10

Data use cycle (Hamilton et al., 2009)



The next section synthesizes these nine inquiry cycles to identify the facets of using data to inform classroom instruction.

Facets of Using Data to Inform Classroom Instruction

As described previously, rather than utilize a specific data inquiry model as the theoretical basis for this instrument, I analyzed all of the peer-reviewed models described in this section and identified elements common to all of these models to use as the theoretical basis to develop this instrument. This process will allow this instrument to be utilized by any school or district, regardless of the inquiry cycle model that they have chosen to use. The identification of the elements common to all of these models is also important, as these are the facets, or components, of the construct of using data to inform instruction. In other words, these facets are the individual pieces that make up the construct of using data to inform instruction.

To identify the facets of using data to inform instruction, I engaged in the following steps, which are each described in more detail below. First, I created color-coded flashcards listing each step for each model/cycle described in the literature review and matched steps across models/cycles. I then used the descriptions of each model's/cycle's individual steps from the literature review to write descriptions of the new steps identified by matching steps across models. These become the facets of using data to inform instruction.

To match steps across models/cycles, I created color-coded flashcards for each model/cycle described previously (a total of 9 models/cycles). Each model/cycle was assigned its own flashcard color for organizational purposes (with the exception of two models that both received blue flashcards because there were only 8 color choices). For each model/cycle, I wrote each step in the model/cycle on an individual flashcard, as well as that step's order in the process of its own model/cycle (e.g., step 1, step 2, etc.). On the back of the flashcard, I wrote the model/cycle the step belonged to. Once all models/cycles had complete sets of flashcards, I started with one model/cycle and laid out the flashcards (steps) in that model sequentially in a vertical column on a table. I then took the next model/cycle and laid out the flashcards for that model next to the first model on the table, matching the steps in the second model to the steps in the first model that were similar in terms of the work completed in that step by placing the cards next to that step horizontally. I continued this process until all models/cycles had been laid out on the table, keeping each model/cycle in its own vertical column. This resulted in horizontal rows that contained steps similar across models/cycles. There were some models/cycles that had steps unique to that model/cycle, creating its own horizontal row. Figure 2.11 displays this organizational structure of the color-coded flashcards.

I then picked up the cards horizontally for each set of matching steps, which resulted in a stack of flashcards from each model/cycle that had similar steps. Each horizontal stack of cards, which represent a similar step across models/cycles, was labeled with the new sequential order of the combination of all models (i.e., the first set of horizontal flashcards picked up was assigned the number 1; the second set of horizontal flashcards picked up was assigned the number 2; etc.) and then each stack was paper clipped together.

Figure 2.11

Completed model flashcards

Then, each new horizontal stack of flashcards (containing similar steps across the 9 models/cycles) was analyzed to identify the facet that each stack represents. This analysis was completed by pulling the written description of each step in a horizontal stack from the literature review into a separate document and analyzing the descriptions of these similar steps across the nine models to write the description of that facet. The facets are described below using this process.

The first horizontal stack consists of steps from only two models – Data Wise and Dana & Yendol-Silva (2003). These steps focus on setting the stage for collaborative work in the inquiry process and are step one for Data Wise (Organize for Collaborative Work) and stage two of Dana & Yendol-Silva (2003) (identifying ways to collaborate with others during the inquiry process). Both steps focus on building collaboration, either with a specific team (as in Data Wise) or with other teachers in a variety of different ways (as in Dana & Yendol Silva, 2003). Although none of the other models/cycles had similar steps in their processes, most mention the collaborative nature of inquiry work and the need for this type of support in the process. In Data Wise, this step is a precursor to the Data Wise process that is typically engaged in only once prior to starting an inquiry cycle. Given that only two models specifically mention this step, and that this step is a precursor to the inquiry cycle in one of those models, this horizontal stack of cards is not included as a common element, or facet, of the construct of teachers using data to inform classroom instruction, although it is noted that a collaborative environment is identified as an important part of the process.

The second horizontal stack of cards includes a step from only one model: Data Wise (Build Assessment Literacy). This step in Data Wise focuses on learning specific principles about assessments to help teachers become responsible users of data. Interestingly, this step or type of work is not mentioned individually in any of the other inquiry cycles, possibly because it is assumed that teachers engaging in this process already have this knowledge or because it is subsumed into the data analysis portion of the inquiry cycle. In either case, again because this stack of cards contains steps from only one inquiry cycle and again because it is a precursor step in the Data Wise process to the inquiry process, it is not included as a facet of the construct of teachers using data to inform classroom instruction.

The third horizontal stack of cards consists of the first and second facet of the construct of teachers using data to inform classroom instruction: **ask questions** and **identify data**. In the ask questions facet, teachers explore interests or general questions that they have about something related to their classroom (Boudett et al., 2013; Dana & Yendol-Silva, 2003; Deming, 2018; Hamilton et al., 2009; Kekahio & Baker, 2013; Mandinach & Gummer, 2016a; National Forum on Education Statistics, 2012). These interests or general questions are then narrowed down to well-worded and defined questions that are open-ended, allowing teachers to explore them during the data inquiry cycle. These questions should be thoroughly thought through to ensure that the question is truly what the teacher wants to investigate (Deming, 2018). A key point consistent across all of the models described in this section is that the inquiry process begins with a question or questions posed by the teacher, teacher team, or school themselves. This is important, as this engenders support from the group to engage in this process and helps teams work through this process because it is specific to their experiences (Brown et al, 2017).

In the second facet, identify data, teachers assess data availability and usefulness in relation to their question(s) to determine if data exists to answer their question(s), if they have access to the data, and if they need to conduct any data collection on their own (Dana & Yendol-Silva, 2003; Deming, 2018; Hamilton et al., 2009; Kekahio & Baker, 2013; National Forum on Education Statistics, 2012). Teachers also think about how to organize and visualize these data as part of the identify data facet (Boudett et al., 2013; Mandinach, 2012). Although asking questions and identifying data were combined as one step in many of the data inquiry cycles included in this analysis, these are truly separate tasks, which is why they are separated into separate facets here. For example, a teacher may be highly skilled at asking specific, concrete questions, but struggle with identifying data sources and determining what kind of data collection they will need to do on their own.

The fourth horizontal stack of cards consists of the third facet: **examine data**. In this facet, teachers gather the data they identified in the identify data stage (Boudett et al., 2013; Dana & Yendol-Silva, 2003; Hamilton et al., 2009; National Forum on Education Statistics, 2012), ensure that they understand the data (including what the variables mean, how they were collected, and data quality) (Hamilton et al., 2009; Mandinach & Gummer, 2016a; National Forum on Education Statistics, 2012), and analyze the data to identify patterns and make observations that can be supported from the data (Hirsh & Crow, 2018; Kekahio & Baker, 2013; Mandinach, 2012). They think critically about comparing various data sources and identify any limitations in the data (Kekahio & Baker, 2013). In some cases, teachers must do their own data collection if the data they identified in the identify data stage does not exist (Mandinach & Gummer, 2016a). Teachers must also ensure that they understand the properties of the data, such as the meanings of all variable values, what missing data means, how data were collected and when, and how to assess data quality. Additionally, if planning to compare various data sources, teachers must ensure that those data sources can be accurately compared.

The fifth stack consists of the fourth facet: **interpret data to set goals**. In this stage, teachers interpret meaning from their data analysis from the third facet (Dana & Yendol-Silva, 2003; Hamilton et al., 2009; Kekahio & Baker, 2013; Mandinach, 2012; Mandinach & Gummer, 2016a; National Forum on Education Statistics, 2012) and use this interpretation to develop hypotheses about how they can improve student achievement (Hamilton et al, 2009). These hypotheses should be based on data, focus on instructional or curricular change that is likely to improve student learning, and be testable (Hamilton et al., 2009). Teachers use these hypotheses to set goals for both teacher practice and student learning (Hirsch & Crow, 2018). These goals should be clearly stated and aligned with the question identified in the first facet (asking questions), as well as aligned with the results of the data analysis from the third facet (examining data). The goals should also be constructed in consideration of the instruction that students have already received, both to understand how students have already been instructed and to ensure that teachers will employ new methods of teaching during the inquiry cycle (Boudett et al., 2013; Hirsh & Crow, 2018). An important point described by Mandinach and Gummer (2016a) is that all intended, and unintended, consequences of these goals should be thought through during this stage, with the understanding that it may not be possible to identify all unintended consequences at this point. Although the National Forum on Education Statistics (2012) and Dana and Yendol-Silva (2003) both include data analysis and interpretation as the same step, the other models separate these two steps. Given the different skill sets involved in analyzing data and then interpreting data, I have separated them into two different facets.

The sixth stack of horizontal cards consists of the fifth facet: **identify intervention**. In this stage, teachers use the goals that they set in the previous stage to identify intervention(s) to implement in their classrooms to meet the goals that they have stated (Hamilton et al., 2009; Hirsh & Crow, 2018; Kekahio & Baker, 2013; Mandinach & Gummer, 2016a). In many cases, this will require researching different pedagogies or different ways to approach student behavior, and may require identifying experts in specific areas to help plan or prepare for these interventions. Professional development may need to be planned so that all involved in the data inquiry cycle are adequately prepared to implement the intervention (Boudett et al., 2013; Hirsh & Crow, 2018). Although not explicitly stated by any model except Boudett et al. (2013) and Hamilton et al. (2009), a key activity in this stage is to identify the plan to assess progress towards the goals with each intervention (note that Boudett et al. (2013) include this as a separate step and it is the only card in the seventh stack of horizontal cards). Planning to assess progress towards goals includes identifying specific data points that indicate progress (or lack thereof) towards the goals. It is important to clarify the plan to assess progress before the intervention is implemented, especially if it will require additional data collection during the intervention.

The eighth horizontal stack of cards consists of the sixth facet: **implement intervention**. In this facet, teachers implement the intervention identified in the previous facet in their classrooms (Hamilton et al., 2009; National Forum on Education Statistics, 2012) and track the student outcomes that were identified in the previous facet (Tichnor-Wagner et al., 2017). During implementation, support from and communication with colleagues can be helpful (Boudett et al., 2013; Hirsh & Crow, 2018), both to ensure that implementation happens as intended and to provide support if it appears that student outcomes are not improving, something that Hirsch and Crow (2018) call "implementation dip", which should be expected in many cases (p. 22). Even if implementation dip happens, teachers should continue to track the student outcomes that they identified as key measures to monitor progress towards their goals. Monitoring these outcomes allows teachers to adjust strategies as needed (Boudett & Steele, 2007).

The ninth, and final, stack of horizontal cards consists of the seventh, and final, facet: **examine outcomes**. In this facet, teachers examine both the student outcome data that they collected in the previous facet, as well as data on how well the implementation of the new strategy went (including identifying if implementation did not occur as anticipated), any issues that were identified when implementing the strategy, and any possible deviations from the original strategy. If student outcomes have met the stated goals, teachers may decide to implement the strategy on a larger level within their school or district, or they may decide to begin the inquiry cycle again with a new question (Deming, 2018; Hamilton et al., 2008; Hirsh & Crow, 2018; Kekahio & Baker, 2013; Mandinach & Gummer, 2016a; National Forum on Education Statistics, 2012). If student outcomes have not met the stated goals, teachers may return to a previous step in the inquiry cycle, but with a focus on different strategies that they can implement to see improvement, as well as a detailed examination into why student outcomes did not improve as anticipated (Deming, 2018).

The construct of using data to inform classroom instruction thus consists of seven facets: ask questions, identify data, examine data, interpret data to set goals, identify intervention, implement intervention, and examine outcomes. The next section briefly describes the literature on self-efficacy, as it was utilized in the pre-dissertation instrument.

Self-Efficacy Literature

The pre-dissertation instrument that was the impetus for this dissertation was designed to measure teachers' self-efficacy for using data to inform instruction because the self-efficacy literature suggests that an individual's perceived level of self-efficacy can have an effect on that individual's behaviors. The self-efficacy literature is rooted in Bandura (1977), who describes, "The strength of people's convictions in their own effectiveness is likely to affect whether they will even try to cope with given situations" (p. 193). Levels of an individual's perceived self-efficacy are also related to the amount of effort and amount of time that the individual will allot in a specific situation (Bandura, 1977), indicating that an individual with higher levels of perceived self-efficacy may exert more time and effort in specific situations than an individual with lower levels of perceived self-efficacy. Although the instrument that will be developed in this dissertation will not focus on teachers' self-efficacy for this construct, this description is included as it was part of the preliminary instrument.

The next section describes existing instruments in the literature that measure constructs related to teachers' use of data to inform classroom instruction.

Existing Instruments

Currently, two surveys exist in the literature that measure constructs related to teacher use of data to inform classroom instruction. The first survey was developed by Dunn et al. (2013) to measure DDDM efficacy and DDDM anxiety and is called the 3D-MEA (DDDM Efficacy and Anxiety) inventory. This survey was developed to measure "teachers' sense of efficacy for DDDM and DDDM anxiety" (Dunn et al., 2013, p. 87) as part of an evaluation of job-embedded professional development related to data use for instructional decision making. Dunn et al. (2013) define efficacy in this case as a latent construct reflecting teachers' beliefs about their own abilities to perform specific tasks related to DDDM with the ultimate goal of improving student outcomes. It was developed as part of an evaluation of a state-wide professional development program in a state in the Pacific Northwest focused on increasing the use of DDDM in classrooms. The evaluation required measurement of teachers' development related to DDDM based on participation in this professional development program. Given the lack of instruments measuring teachers' efficacy with DDDM in the literature, Dunn et al. (2013) created the 3D-MEA. The instrument was developed by the two researchers who had developed the professional development program under evaluation and an educational psychologist who was the outside evaluator for the program. The instrument was developed by utilizing literature on teachers' overall sense of efficacy (not specific to DDDM) and the literature related to the specific components of DDDM.

Dunn et al. (2013) identify the four components of DDDM as (a) data identification and access, (b) data technology use, (c) anxiety, and (d) data analysis, data interpretation, and application of data to instruction. Their instrument was designed to measure teachers' efficacy with each of these four components. The pilot version of the 3D-MEA consisted of 22 items designed to measure these four components and was administered to the teachers participating in the professional development program. A total of 1,728 teachers responded to the survey. The researchers split the responses in half; they used half of the responses to conduct an exploratory factor analysis and the remaining half to conduct a confirmatory factor analysis. The final solution indicated a five factor structure; Dunn et al. (2013) describe, "The expanded five-factor structure found in this study reflects the complexity of the variables and tasks inherent to classroom level DDDM and highlights the ongoing need to better understand DDDM in the classroom" (p. 95).

Although the purpose of their instrument is similar to my intended purpose, the 3D-MEA survey consists of 3 sub-scales for efficacy with data that focus more on teacher confidence in performing a specific task rather than trying to place a teacher at a specific location along a continuum, or hierarchy, of skills. The 3D-MEA survey measures efficacy for data identification and access (3 items; Cronbach's alpha = .84), efficacy for data analysis and interpretation (3 items; Cronbach's alpha = .81), and efficacy for application of data to instruction (6 items, Cronbach's alpha = .92) (Dunn et al., 2013). While Dunn et al.'s (2013) analyses of the 3D-MEA survey results indicated favorable reliability values for the efficacy scales, the small number of items in each scale accompanied by the fact that the items were not developed based on a conceptual or theoretical framework indicates that other instruments may be useful. Additionally, the items were not developed to measure specific skills involved in each stage of using data to inform instruction; if the data from the survey is used to identify specific areas in which teachers need additional training or support, it provides a broad view of the specific areas of using data to inform instruction in which teachers need support, but does not help identify specific skills.

Validation studies have been performed with the 3D-MEA with additional populations (Walker, Reeves, & Smith, 2016), indicating that the use of an instrument measuring teachers' use of data to inform instruction is desired. My instrument can help fill this need in both research and practice and provides more detailed information on a teacher's specific location on the hierarchy of skills, providing schools and districts with a description of a teacher at that point on the continuum that can be used to better support a teacher in using data to inform instruction.

The second survey that currently exists in the literature is the Teacher Data Use Survey and was developed by Wayman et al. (2016) with the purpose of gathering information on "how teachers use data to support instruction, their attitudes toward data, and the supports that help teachers use data" (p. i). The main uses of the data gathered from this survey are described as providing an overview of how teachers are using data, providing a "comprehensive perspective on how teachers view data use", and providing "an evidence base from which to plan ongoing support, such as professional development, computer data systems, and collaborative structures" (Wayman et al., 2016, p. 1). The instrument was developed using a conceptual framework focused on how teachers use data and measures the following constructs: actions that teachers take with data, teacher competence in using data, teacher attitudes toward data, collaboration with other teachers, and organizational supports that are available to teachers (Wayman et al., 2016).

The construct from the Teacher Data Use Survey most closely aligned with my instrument is teacher competence in using data. In the Teacher Data Use Survey, this scale is described as measuring "how good teachers are at using data to inform various aspects of their practice" (Wayman et al., 2016, p. 9), and includes four items asking teachers to report their level of agreement with how good they are at using data for diagnosing learning needs, adjusting instruction, planning lessons, and setting learning goals (Wayman et al., 2016). This scale focuses more on the frequency of actions teachers take with data rather than on their perceived self-efficacy with these skills.

Additionally, although Wayman et al. (2016) state that the use of their survey is to provide evidence to plan professional development, this professional development is targeted to the whole population of teachers, not to individual teachers. My instrument focuses on specific skills utilized during the process of using data to inform instruction, which can be the specific focus for individualized targeted professional development or supports based on survey results.

The next section focuses on Classical Test Theory and Item Response Theory and their application to my instrument design.

Classical Test Theory and Item Response Theory

Research is often focused on measuring the amount of a particular attribute in individuals or groups. For example, research may focus on measuring the amount of a particular belief or ability that an individual holds, such as beliefs about gender roles. Another example may be measuring the amount of mathematics knowledge an individual has. In these cases, these attributes are not directly observable when looking at an individual, but rather must be measured using some instrument. Classical Test Theory and Item Response Theory are two theories that offer ways to measure the amount of an attribute that an individual has. They differ in the way in which the measurement model is conceptualized.

Classical Test Theory (CTT) focuses on the total score on an instrument. It assumes that a true score exists for each person and that what is measured (or observed) by an instrument is equal to that person's true score plus some amount of error. Under

CTT, the observed measurement will always include some amount of error; the goal is to reduce that error so that the observed score is as close as possible to the true score.

Item Response Theory (IRT) focuses on the individual item responses and does not assume that a true score exists for each person. The models in IRT yield the probability of a specific response to a given item. IRT models are split into person ability estimates and item difficulty estimates, which allows person ability estimates to be "free" from the item estimates and the item estimates to be "free" from the person estimates. The different IRT models are differentiated by the number of parameters associated with item-specific characteristics.

Both CTT and IRT provide item parameter statistics, including item difficulty and item discrimination, although the calculations for these parameters differ for the two theories. Item difficulty describes the difficulty level of the item, i.e., how hard it is to score the highest value for an item. For example, in CTT, item difficulty for dichotomous right/wrong responses is measured as the percentage of people answering the item correctly. This means that a higher item difficulty indicates an easier item; for example, an item with a difficulty of 0.8 would be a relatively easy item, indicating that 80% of participants answered that item correctly. In IRT, item difficulty is conceptualized as the probability of getting the item right at a given ability level. It is measured by identifying the amount of ability required in the model to have a 50% probability of answering the item correctly. In contrast to CTT, a higher item difficulty in IRT indicates a more difficult item, as this indicates that a higher level of ability is required for a 50% probability of answering the item correctly. Specific to scenario-based items, which have multiple response options and no "right" answer, item difficulty refers to the probability of choosing the highest response option given a certain ability level. Scenario-based items with a higher difficulty level indicate scenarios where it is more difficult for a participant to endorse, or choose, the highest response option.

Item discrimination describes how well the item differentiates between people with high levels of the construct and those with low levels of the construct. In CTT, item discrimination is calculated by the item-total correlation or corrected item-total correlation. The item-total correlation is the correlation between the individual scores on an item and the total scores across all items. The corrected item-total correlation is the same correlation but without the inclusion of the particular item being measured in the total score. This provides a more accurate measure of the item-total correlation, as including the particular item in the total score can inflate the discrimination indices. In IRT, item discrimination is calculated somewhat differently but is essentially the corrected item-total correlation.

IRT models can employ a variety of parameters in their estimation procedures. The family of IRT models includes models that estimate one item parameter (1PL), namely item difficulty, models that estimate two item parameters (2PL), item difficulty and item discrimination, and models that estimate three item parameters (3PL), item difficulty, item discrimination, and guessing. The Rasch measurement model is part of the family of IRT models and estimates a single item parameter: item difficulty, making it a 1PL model. The purpose of the Rasch measurement model is to measure a unidimensional variable (i.e., construct) that is hypothesized to span a continuum, meaning that the levels of the variable span a hierarchical continuum from low levels of the variable to high levels of the variable. The Rasch model assumes that item discrimination is the same across all people for whom this variable is measured, and it also assumes that guessing does not occur. For a variable that is hypothesized to span a continuum, the intention of the instrument developer is to develop a set of items that define this continuum from low levels of the variable to high levels of the variable. Additional details on Rasch measurement principles for instrument development may be found in Rasch (1966) and Ludlow et al. (2014), and are expanded upon here in the Methodology chapter (Chapter 3).

In this dissertation, I use IRT methods as part of the Rasch/Guttman Scenario (RGS) approach to develop this instrument. The pre-dissertation version of this instrument (a detailed explanation of the pre-dissertation instrument is included in Chapter 3) utilized CTT methods to analyze the factor structure of items that were written based on the DDDM framework (Mandinach, 2012) and to justify the removal of particular items that did not behave as expected (see more detailed description in Chapter 3). The use of both CTT and IRT methods in the development of an instrument can be beneficial for instrument design; Peoples et al. (2014) described the use of CTT principal components analysis in conjunction with Rasch analyses to create an "optimal instrument" (p. 55) measuring student perceptions of elementary school science classroom environments. Peoples et al. (2014) describe, "Although the primary emphasis of the study was to create a Rasch-based instrument, the perspective taken in this research was that these methodologies are not mutually exclusive and can be used synergistically to create an optimal instrument" (p. 55). Although I do not use both CTT and IRT methods with the RGS methodology, it is worth pointing out these potential synergies given that the pre-dissertation instrument that was the impetus for this dissertation was designed using CTT methods.

Chapter Summary

This chapter reviewed the literature that is relevant to my dissertation; specifically, it discussed the literature focused on teachers' use of data to inform instruction (or DDDM) and described nine peer-reviewed frameworks/models of DDDM/inquiry cycles. The overlapping steps from these nine frameworks/models were utilized to identify the facets of the construct of using data to inform classroom instruction. The seven identified facets are ask questions, identify data, examine data, interpret data to set goals, identify intervention, implement intervention, and examine outcomes.

This chapter also discussed the RGS methodology, the self-efficacy literature (as it was used in the pre-dissertation instrument), a discussion of existing instruments designed to measure the use of data to inform instruction, and descriptions of CTT and IRT. A description of CTT was included because it was utilized in the pre-dissertation instrument, while IRT methods are utilized to design and analyze the data from the dissertation instrument. The next chapter focuses on the research design and methodology for the dissertation instrument.

Chapter 3 : Research Design/Methodology

This section describes the research design and methodology for the dissertation instrument, the UDII (Using Data to Inform classroom Instruction) scale, and analysis. The construct of using data to inform classroom instruction is reviewed briefly and is followed by a brief discussion of the pre-dissertation instrument that helped spur the interest to develop the instrument for this dissertation. It then details the individual steps of the RGS approach and how they were used to develop the UDII scale. This section provides a detailed description of the overall construct, describes teachers at low, medium, and high levels of the overall construct and each individual facet, and develops a mapping sentence to be used to create the scenarios. It also details the development of the initial scenarios, response options, and instructions for the UDII scale. Finally, it describes the data collection plan, as well as the plan for the analysis of the data after data collection.

Construct of Using Data to Inform Classroom Instruction

As previously mentioned, the pre-dissertation version of this instrument was designed using CTT methods as part of my Survey Methods course in Fall 2015, utilizing Mandinach's (2012) DDDM framework to design the items. While this version of the instrument does provide some useful information about teachers' perceived self-efficacy for using data to inform their classroom instruction, I hypothesize that there is a continuum of skills (or facets) within the construct of using data to inform classroom instruction that teachers utilize when going through the process of using data to inform their classroom instruction. As described in the literature review, this continuum of facets was derived from peer-reviewed models of data inquiry cycles and frameworks. These frameworks and cycles of inquiry generally focus on the steps that individuals take when using data to inform instruction. Synthesizing these frameworks and cycles of inquiry allowed me to identify the specific skills, or facets, which are used by teachers in this process, with the goal of developing a revised scale utilizing scenarios designed to measure this hypothesized continuum of skills. The revised scale was developed using RGS methodology, utilizing Guttman's facet theory (Guttman & Greenbaum, 1998) and sentence mapping (Hackett, 2014) to create a scenario-based scale with the goal of reflecting "lived experiences" (Ludlow et al, 2014) for those responding to the scale. IRT methods, specifically a Rasch model, was utilized through Rasch analysis procedures to determine the extent to which my hypothesized continuum of skills, or facets, is supported by the empirical results.

Based on the literature review for data-driven decision making and data inquiry cycles, the following seven facets were identified as the facets in the construct of using data to inform classroom instruction: ask questions, identify data, examine data, interpret data to set goals, identify intervention, implement intervention, and examine outcomes. These facets comprise the construct of using data to inform classroom instruction and form the basis for the development of the scenario-based scale in this dissertation.

The next section describes the development of the pre-dissertation instrument that spurred my interest in pursuing this topic for my dissertation. The pre-dissertation instrument was designed using CTT methods. Table 3.1 clarifies the different terms I use when describing the pre-dissertation instrument compared to the dissertation instrument.

Table 3.1

Terminology and definitions used to describe pre-dissertation instrument and dissertation instrument

Term	Meaning	Used with pre-dissertation instrument?	Used with dissertation instrument?
Construct	This is the overarching concept that the instrument is designed to measure.	Yes	Yes
Facet	The individual pieces, or components, that make up a specific construct.	No	Yes
Instrument	The compilation of items that are designed to measure a particular construct.	Yes	Yes

Pre-Dissertation Instrument Development

My pre-dissertation instrument was initially developed as part of the Survey Methods coursework using CTT procedures to gather evidence about the hypothesized factor structure from the DDDM framework (Mandinach, 2012). As the pre-dissertation instrument was developed using CTT methods, the word "construct" as described in Table 3.1 is utilized throughout this description to describe the concept that the pre-dissertation instrument was designed to measure. Items were written to measure the individual constructs of "Data", "Information", "Knowledge", and "Implement/Impact" as depicted in the framework in Figure 2.1. The specific cognitive processes described under each construct (for example, collect and organize under "Data") were hypothesized to be so related that they were not actually separate constructs. To test this hypothesis, I wrote items to measure each cognitive process under each construct. The cognitive processes for "Implement" and "Impact" were also hypothesized to be so related that they are the same construct, rather than individual constructs.

The pre-dissertation instrument items were designed for a five point Likert scoring model with the response options of Strongly Disagree (assigned a score value of 1), Disagree (assigned a score value of 2), Neither Agree nor Disagree (assigned a score value of 3), Agree (assigned a score value of 4), and Strongly Agree (assigned a score value of 5). Likert scoring was chosen for the pre-dissertation instrument given this type of scoring's commonality in literature and in practice, increasing the likelihood that participants would be familiar with the format, and because data from a Likert format can be analyzed using CTT techniques. The decision to include the answer choice of "Neither Agree nor Disagree" was made based on previous personal experience with Likert scoring where participants truly want the option to answer Neither Agree nor Disagree to particular items where this answer choice most clearly mirrors their experience. Although the inclusion of this answer choice does provide participants with the opportunity to provide an "easy answer" because it does not force them to choose either Agree or Disagree, the benefits of including a Neither option outweighed this possibility.

Once the first draft of pre-dissertation items was complete, these items were reviewed by two sets of experts: the first group included teachers and administrators who had utilized data to inform their classroom instruction and the second group included class members of the Survey Methods course and the course professor. The first group of experts comprised three teachers who taught in a kindergarten through grade eight school in a small, suburban town in Massachusetts, as well as a central office administrator in a different small, suburban town in Massachusetts. These experts provided feedback on the wording and the content of the items based on their experience. They also provided their thoughts on any items that could be difficult for teachers to answer or that could be possibly problematic or harmful for teachers to answer, as well as whether anything was missing from the items. The majority of the feedback from this group of experts focused on the wording of the items, as well as the lack of items focused on data accessibility for teachers. This group noted that accessibility of data was an important piece of using data to inform instruction, which was not included in the first draft of items. Based on this group's description, accessibility of data captured the availability of data to teachers and the ease of accessing the data. For example, in some districts, teachers felt that data were not available to all teachers; in other districts, although data were available, teachers were either not able to access it or were not trained to access it. In both cases, this expert group felt that although teachers might be proficient with most of the skills involved in using data to inform instruction, the lack of data accessibility hindered their ability to utilize these skills and was important to measure.

The feedback on item wording and the need to include items measuring data accessibility were incorporated into the pre-dissertation set of items. The revised set of items was then shared with class members in the Survey Design course and the course professor. Their feedback focused primarily on item wording and the suggested edits were incorporated into the final set of pre-dissertation items, used in the pilot administration of the pre-dissertation instrument.

The pre-dissertation instrument was piloted in a public school district in a medium-sized city in Massachusetts using an online survey tool, Qualtrics. An introductory email describing the survey and its purpose as a pilot survey for a Survey

Methods course was emailed to all teachers employed by this district (total of 1,187 teachers), with a sentence explaining that the survey link would be emailed in a few days and that participation was completely optional. An email containing the survey link and text describing the survey's purpose was sent two days later. As an incentive to complete this optional survey, I offered a lottery of two $50 Amazon.com gift cards. Survey participants could opt to provide their email addresses at the end of the survey to be entered into the lottery, and email addresses were stored separately from participant responses to ensure anonymity in the dataset. Two participants were selected at random once survey administration closed and were emailed the gift cards via their provided email addresses.

The data collected by this pre-dissertation survey was confidential. Individual links were emailed to teachers' school email accounts, but solely to ensure that only teachers in this district responded to the survey. This also allowed for reminders to be emailed to teachers who had not yet completed the survey during the survey administration window. The data was exported from Qualtrics without identifying information, however, and, as mentioned, the email addresses for the gift card lottery were exported and stored separately from the survey response data.

A total of 410 teachers started the pre-dissertation survey and provided a response to the consent statement on the first page of the survey, resulting in an estimated response rate of 34.5%. Of the 410 teachers who responded to the consent statement, 6 teachers did not consent to participate and were removed from the dataset, resulting in a dataset of 404 responses. Of these 404 responses, approximately 30% (123 teachers) did not complete the entire survey. A useful feature of Qualtrics provides users with information on whether a participant has seen an item but left it unanswered or has not seen an item. Qualtrics also provides users with an indicator of whether or not a participant has completed the survey by clicking through every page to the end. These features were utilized to identify teachers who were able to answer the entire survey (even if they had not provided an answer to each item). Teachers who did not finish the pre-dissertation survey were removed from the dataset, resulting in a final dataset of 281 participants.

CTT methods were used to analyze the pre-dissertation data. After identifying potentially problematic items based on items with restriction of range (items where one or more answer options had no responses), low inter-item correlations (items that had low correlations with other items on the survey), and low item discrimination values (low corrected item-total correlations), an exploratory factor analysis was run to analyze the factor structure of the items. The potentially problematic items were included in the exploratory factor analysis, but particular attention was paid to their influence on each factor to determine if they should be included.

The final pre-dissertation factor solution identified 7 factors using principal axis factoring with promax rotation. A total of 11 items that were initially flagged as potentially problematic were removed from the final scales. The extracted factors were data accessibility (4 items, Cronbach's alpha = .856), data collection and organization (7 items, Cronbach's alpha = .783), positively worded information (6 items, Cronbach's alpha = .731), negatively worded information (4 items, Cronbach's alpha = .817), knowledge (9 items, Cronbach's alpha = .851), implement/impact (7 items, Cronbach's alpha = .822), and beliefs (11 items, Cronbach's alpha = .923). The factor names are based on the DDDM framework (Mandinach, 2012) and identify each stage of the DDDM framework.

The results from this pre-dissertation instrument formed the basis for this dissertation and sparked the interest to focus on the development of a scenario-based scale to measure teachers' use of data to inform their classroom instruction. The factor solution from the pre-dissertation instrument provided some evidence that measuring the use of data to inform instruction using Mandinach's (2012) framework was feasible. Given the ability of scenario-based scales to provide participants with rich descriptions of life experiences when engaging with a survey instrument, this dissertation focuses on developing an instrument using the RGS approach.

Instrument Design with RGS Methodology

As described in the literature review, scenario-based scale methodology consists of the following steps: define the construct, determine facets and generate narrative descriptions for each facet, determine the facet levels and generate descriptions to capture variation within each facet, determine the structure of the scenarios, develop the mapping sentences and construct the scenarios, decide on the response options and survey instructions, and test congruence of theory and practice (Ludlow et al, 2020). More detail on how each of these steps is used for the design of the UDII scale is described below.

Step 1: Define the construct

My research interests have always focused on teachers' use of data to inform classroom instruction. I began my doctoral program with this research interest based on my previous experience as a high school math teacher and my experience working in a district data analyst role in a large urban school system. My current work as the Director of Data Analysis and Enrollment Planning for a medium-sized city has continued to inspire my interest in this area. As part of my coursework in my doctoral program, I focused my assignments on teachers' use of data to inform classroom instruction whenever possible to begin building my literature review. Based on the body of literature that I built throughout my coursework, I identified additional resources based on reference lists. I also identified specific keywords that I utilized to expand my literature search; specifically, I used "data-driven decision making", "data-based decision making", "using data", "data-driven instruction", and "data inquiry cycle". My review of these keyword searches was restricted to articles focused on teachers and on school leaders, and included only those studies conducted in the United States specifically because of the US-based policies and focus on US teachers in this research. Although many models of cycles of inquiry exist in both literature and practice, I have limited those included in this research to models published in peer reviewed journals, books, or government guides.

Based on this body of literature, I identified the construct of "Data-Driven Decision Making", often identified in the literature by the acronym, DDDM. As previously described, although the literature most commonly refers to this practice as DDDM, I specifically and intentionally use the terminology "using data to inform classroom instruction" instead of DDDM throughout my research. Again, I believe that this terminology emphasizes that the purpose of this activity is to inform classroom instruction, while the terminology Data-Driven Decision Making emphasizes data. I believe that data is a tool to inform classroom instruction, rather than the outcome of this practice.

I synthesized this body of literature (described in the literature review), along with my personal experience, to identify a common set of skills, or facets, that teachers utilize when engaging in using data to inform classroom instruction. This common set of skills (facets) drives my definition of this construct and is often referred to as an inquiry cycle in the literature. Although there are many models of inquiry cycles, the general practices that teachers utilize are common and make up my definition of this construct. I define the construct of teachers using data to inform classroom instruction as follows:

teachers identify teaching and learning challenges in their classrooms and generate questions about student learning based on those challenges. They then identify and gather data to help answer and further refine their questions. They analyze this data and based on this analysis, they form hypotheses about how to improve student learning. Based on these hypotheses, they utilize experts or their own expertise and other external sources to identify an instructional intervention. They implement the intervention and determine its success by gathering and analyzing additional data. If the intervention was not successful, they repeat this process, either from the beginning or from a specific step in the process.

This is often referred to as the cycle of inquiry in the literature, and although there are many different models of cycles of inquiry as described in the literature review, the general process of these models follows my definition.

One key point is that the UDII scale is not designed to measure teachers' beliefs about using data to inform instruction. While I recognize that teachers' beliefs about the utility and benefits of using data to inform classroom instruction may affect their willingness to engage in this practice, it is outside the scope of this dissertation. The development of a culture of data use may be a necessary, and certainly important, step for schools and districts that are implementing or building on the use of data to inform classroom instruction in their communities, and tools exist (such as a toolkit developed by the REL Northeast & Islands) to engage in this type of work. The UDII scale is focused specifically on teachers' use of data to inform their classroom instruction as previously defined.

The construct definition above allows for the development of detailed descriptions of teachers at varying levels of the construct of using data to inform classroom instruction. The goal in Rasch measurement design is to develop a ladder-like, hierarchical continuum for the construct, describing individuals at various levels of the construct: in this case, using data to inform classroom instruction. The descriptions below describe a teacher at the low level, medium level, and high level of the construct of using data to inform classroom instruction. Teachers can progress through these levels of the construct as they get more training on the skills involved in this construct and/or as they work to develop more skills towards the use of more data in their practice. Although individual teachers may exist along the entire continuum of using data to inform classroom instruction (and are not limited to being in a low, medium, or high level), only three levels are described here for parity and to provide a rich description of what teachers look like along this continuum.

Low level: A teacher at the lowest level of the construct of using data to inform classroom instruction is very uncomfortable using data and avoids it if at all possible. This teacher cannot analyze data on their own and requires support to decide on specific interventions based on the results of the data analysis. This teacher does not have the skillset to engage in this type of work individually and avoids engaging in this practice both in formal settings, and individually in a systematic way, when at all possible.

Medium level: A teacher at the medium level of using data to inform classroom instruction is uncomfortable using data to inform instruction, but is willing to try engaging in this activity. This teacher needs help analyzing data and deciding on specific interventions based on the data analysis. This teacher may not feel confident enough in their skills for this work to believe that engaging in this practice could result in beneficial outcomes for students. This teacher generally engages in this practice only in formal settings, such as Professional Learning Communities (PLCs) or in team meetings, and avoids engaging in this practice systematically on their own.

High level: A teacher at the high level of using data to inform classroom instruction is very comfortable using data to inform instruction and is often the “go-to” person in their school building for anything related to data and for other teachers who need help engaging in this activity. This teacher feels confident in their ability to engage in this activity and is frequently asked to help other teachers or building staff with their questions related to this work. This teacher thinks that using data to inform instruction is critical to their teaching practice and regularly engages in this work in both formal settings and individually.

Figure 3.1 displays this ladder-like, hierarchical continuum for the construct of using data to inform classroom instruction with descriptions of individuals at the low, medium, and high levels.

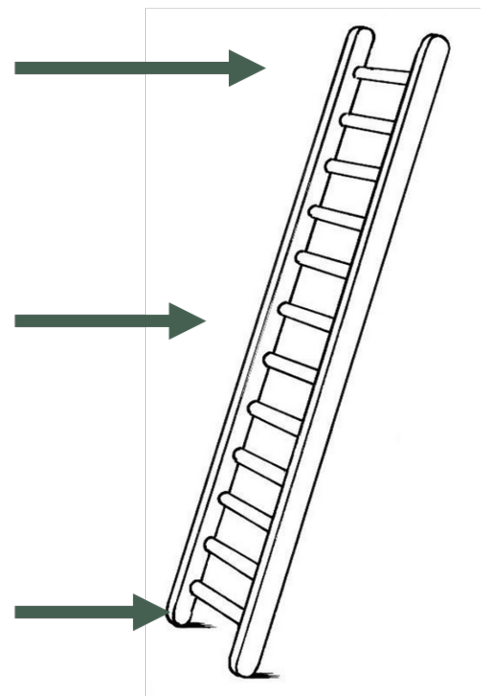
Figure 3.1

Ladder-like continuum for using data to inform classroom instruction

High level:
- Is very comfortable using data
- Is often the "go-to" person in a school building for others who need help with data or with this work
- Thinks that using data is critical for their teaching practice and engages in this work both formally and individually

Medium level:
- Is uncomfortable using data, but is willing to engage in this activity
- Needs help analyzing data and deciding on specific interventions
- Generally engages in this practice only in formal settings

Low level:
- Very uncomfortable using data and avoids it if possible
- Cannot analyze data on their own, requires support to decide on intervention
- Does not have the skillset to engage in this work individually, avoids engaging in this practice formally or individually in a systematic way when possible



Step 2: Determine facets and generate narrative descriptions for each facet

The detailed construct definition and the narrative descriptions of teachers at the low, medium, and high levels of the construct help to describe the facets detailed in Chapter Two. As described previously, facets are specific variables that, when woven together like strands of fiber to create yarn, make up the construct. For the construct of using data to inform instruction, the facets can also be conceptualized as the skills that comprise this construct. The set of skills that teachers engage in when using data to inform instruction are hypothesized to be hierarchical, meaning that in the first iteration of a data inquiry cycle, the process involves an ordered set of skills, as described in the literature review. This hierarchical nature of a data inquiry cycle, however, does not mean that all teachers learn or hone these skills in the order in which they are utilized in a data inquiry cycle. This means that some teachers may be highly skilled in skills at the bottom of the hierarchical order (i.e., ask questions) but have difficulty with other skills in the hierarchical order, while others may be highly skilled in skills at the top of the order (i.e., examine outcomes) but need support in skills at the bottom of the hierarchical order.

The literature review provided the basis for identifying the following seven facets: ask questions, identify data, examine data, interpret data to set goals, identify intervention, implement intervention, and examine outcomes. Detailed narrative descriptions of each of these seven facets are provided in the literature review.

Step 3: Determine the facet levels and generate descriptions to capture variation within each facet

The next step in the RGS approach is to write rich descriptions of individuals (in this case, teachers) at each level (low, medium, and high) of each facet. As described previously, the construct is conceptualized as a continuum spanning lower to higher levels of using data to inform instruction, and the facets, which are interwoven to form those levels of the construct, themselves consist of lower to higher levels. Just as with the continuum of the construct, teachers may exist anywhere along the continuum for each of the facets, but three levels for each facet are described to provide a rich description of this continuum, while maintaining parity. In addition, generating descriptions at three levels for each facet helps to generate a proof of concept for the RGS method by ensuring that clear descriptions of each facet can be generated, while also providing a proof of concept for the operationalization of the construct by ensuring that the construct can be truly be described as a hierarchical continuum from low to high levels (Ludlow et al., 2020).

These descriptions of each facet at the three levels follow below and were developed using the detailed facet descriptions that were written in the literature review, combined with personal professional experience, to describe a teacher at the low, medium, and high level of each facet.

As with the descriptions of teachers at low, medium, and high levels of the construct of using data to inform instruction, teachers can progress through the levels of each facet as they gain skills and grow through professional development, experiences, and practice, among other things.

Facet 1: Ask questions

Low: A teacher on the low end of the facet of ask questions struggles to identify questions that they have related to their classrooms or school. This teacher has difficulty pinpointing areas that they would like to explore. Additionally, once this teacher has identified some questions of interest, they struggle to clearly word the questions and have difficulty defining exactly what they plan to explore. This teacher also often writes close-ended questions.

Medium: A teacher at the medium level of the ask questions facet can identify general questions that they have related to their classroom or school, although they may have trouble generating some questions initially. This teacher can identify general areas that they want to explore but often struggles with narrowing down the general areas to specific questions. This teacher sometimes has difficulty writing clearly defined questions to explore and may sometimes write questions that are not open-ended.

High: A teacher at the high level of the ask questions facet can identify questions that they have related to their classroom or school. Often, a teacher at the high level may have many questions that they are interested in exploring and can successfully narrow these questions down to the one or ones that are most pressing or time sensitive to this teacher. This teacher is skilled at writing clearly defined and open-ended questions.

Facet 2: Identify data

Low: A teacher at the low level of the identify data facet may not know what data can answer their question(s); if they do have a sense of what data can help answer their question(s), they struggle to identify the availability of this data. This teacher does not know how to approach the organization of their data and typically struggles with organizing data in a clear way. A teacher at this level has difficulty thinking about how to visualize data, and often will use the same type of data visualization displays no matter what kind of data they are using, even if this visualization is not the best option to choose.

Medium: A teacher at the medium level of the identify data facet can sometimes identify what data can help answer their question(s), although they may need some support in thinking about all possible data sources. This teacher can usually identify if data is available, but again, may need some support in this area. A teacher at this level can often identify effective data visualization strategies, but may struggle to implement them without sufficient support. A teacher at this level may stick to the same type of data visualization with many data sources because they are most comfortable with this type of visualization.

High: A teacher at the high level of the identify data facet can always identify data that can help answer their question(s), although they do sometimes benefit from discussion with others to identify additional data sources. This teacher can also determine if the data they have identified is available and accessible. A teacher at this level is able to logically and effectively organize their data and is skilled at data visualization. This teacher often uses many different types of data visualization depending on the type of data that they are using and the question(s) that they are asking.

Facet 3: Examine data

Low: A teacher at the low level of the facet of examine data often struggles to gather the data they identified in the previous facet. If this teacher has to collect their own data, they struggle to collect this data in a systematic and organized way. This teacher often does not understand what their data variables mean, how the data was collected, and will not question data quality, which can lead to errors in their observations from the data. This teacher needs support to analyze their data and may often struggle to make observations that are supported by their data. Additionally, if using more than one data source, this teacher may not think critically about what, if any, comparisons may be made, which may also lead to errors in the observations they make from this data.

Medium: A teacher at the medium level of the facet of examine data can gather the data that they identified in the previous facet, although they need some support in pulling data from particular data sources and/or pulling the data together into one dataset. If required, this teacher is able to collect their own data in an organized manner. This teacher knows how to investigate and think about the data variables that they are using, think about how the data was collected, and think through data quality issues, although they may struggle to understand some of these elements. This teacher can often analyze their data without support, although they may struggle with identifying and implementing the most appropriate or efficient ways to analyze their data. Once the data are analyzed, this teacher can often make observations from the data, but may have some difficulty clearly stating these observations. If using more than data source, this teacher will think about whether the sources can be compared, but may have trouble determining if they can actually be compared.

High: A teacher at the high level of the facet of examine data can gather the data that they identified in the previous facet, can pull data from various sources efficiently, and can combine data sources into one dataset if required. If this teacher is collecting their own data, they can do it in an organized and efficient manner. This teacher will investigate and understand the data variables that they are using, think about how the data was collected, and think through data quality issues with success. This teacher can analyze their data without support, and is skilled in identifying and implementing the most appropriate and efficient ways to analyze their data. Additionally, if this teacher realizes that they need to use a different method of analysis with which they are less skilled, they are able to research and learn about this method of analysis on their own or with the support of others and implement it successfully. Once this teacher has analyzed their data, they can make observations from the data with ease. If using more than one data source, this teacher thinks critically about whether these sources can be compared and only compares them if they can be compared.

Facet 4: Interpret data to set goals

Low: A teacher at the low level of the facet of interpret data to set goals struggles to interpret meaning from their data analysis and may often make incorrect interpretations. This teacher may also struggle to use these interpretations to hypothesize about ways to improve student outcomes or be unable to generate hypotheses based on their interpretation of the data. This teacher is usually unable to write goals for both teacher practice and student learning, and when they do write goals, these goals may not be clearly stated or aligned to the overall question(s) identified at the beginning of this process once written. This teacher may also often fail to consider any intended or unintended consequences of the goals that they have written.

Medium: A teacher at the medium level of the facet of interpret data to set goals may need some support to interpret meaning from their data analysis. This teacher may also need help hypothesizing about ways in which they can improve student outcomes, or may generate hypotheses that are not based on their data interpretations or are not testable. This teacher can use their hypotheses to write goals for both teacher practice and student learning, although they may have trouble clearly stating these goals and/or ensuring that they are fully aligned with the question(s) posed at the beginning of this process. This teacher will often consider intended consequences of the goals that they write, although they may have difficulty identifying unintended consequences of these goals.

High: A teacher at the high level of the facet of interpret data to set goals can draw meaningful interpretations from their data analysis independently, although they may benefit from discussing these interpretations with others. This teacher can generate hypotheses based on their data interpretations that focus on instructional or curricular change and are testable. This teacher can use their hypotheses to write goals for both teacher practice and student learning that are clearly worded and fully aligned to the question(s) posed at the beginning of this process. This teacher also considers all intended and unintended (to the extent possible) consequences of the goals that they have written.

Facet 5: Identify intervention

Low: A teacher at the low level of identify intervention requires support to identify any interventions to meet the goals that they have written in the previous facet. Although many teachers may require support from coaches or other experts in their school to identify interventions, this teacher is not able to identify potential sources for interventions independently and is often unsure of who to ask for help. This teacher does not think about whether professional development is required for any interventions that they eventually identify. A teacher at this level fails to identify a plan to assess progress towards their goals with any intervention.

Medium: A teacher at the medium level of identify intervention may require support to identify any interventions to meet the goals that they have written in the previous facet, but they are able to identify sources for support, including both people and written materials. This teacher thinks about whether any professional development is required for the interventions that they identify and is able to seek it out. This teacher also thinks about the plan to assess progress towards their goals, although they may struggle to identify what data variables are necessary to document progress towards these goals.

High: A teacher at the high level of identify intervention can identify interventions to meet the goals that they have written in the previous facet. They may identify these interventions through their own research or through conversations with experts. This teacher thinks about any professional development that is required for the interventions that they identify and they can identify the steps necessary to ensure that the professional development occurs. This teacher thinks through their plan to assess progress towards their goals in detail and explicitly identifies the data variables that they will need to collect to document progress towards their goals.

Facet 6: Implement intervention

Low: A teacher at the low level of the facet of implement intervention has trouble implementing their intervention with fidelity. This teacher rarely seeks out support or communication with colleagues during implementation of the intervention, and often keeps information about their progress to themselves. This teacher usually does not collect or track student progress towards the goals that they set previously, or does not collect or track this data consistently.

Medium: A teacher at the medium level of the facet of implement intervention can implement their intervention, but often needs the support of colleagues to ensure that it is implemented with fidelity. This teacher will often seek out support and communication with colleagues during implementation of the intervention, but can have difficulty relaying information about their progress with the intervention to this group. This teacher plans to collect and track student progress towards their goals, but may not collect certain key data variables or may not collect them consistently.

High: A teacher at the high level of the facet of implement intervention can implement their intervention with fidelity independently, and also seeks out the support of colleagues during implementation to relay successes and failures. This teacher finds benefits in the communication with colleagues. This teacher plans for and collects student progress data towards their goals and is consistent in this data collection.

Facet 7: Examine outcomes

Low: A teacher at the low level of examine outcomes struggles to examine the student outcome data that they may have collected in the previous facet and has difficulty identifying if student outcomes meet their stated goals. This teacher also has difficulty analyzing information on the implementation of the intervention itself, including whether it was implemented with fidelity. This teacher is often unsure of how to proceed at this stage: whether they should implement the intervention at a larger level if it was successful or revisit earlier stages of the inquiry cycle again with a focus on different strategies or interventions if it was unsuccessful.

Medium: A teacher at the medium level of examine outcomes can examine the student outcome data that they collected in the previous facet, but may need some support in the analysis or the identification of whether learning goals were met. This teacher can also analyze the implementation of the intervention and identify any deviations from the intended intervention. This teacher needs some support in how to proceed at this stage: whether they should implement the intervention at a larger level if it was successful or revisit an earlier stage of the inquiry cycle again with a focus on different strategies or interventions if it was unsuccessful.

High: A teacher at the high level of examine outcomes can examine the student outcome data that they collected in the previous facet and can identify whether student learning goals were met. This teacher analyzes the implementation of the intervention to identify if it was implemented with fidelity or if any deviations from the original plan were made. This teacher is able to identify how to proceed at this point, although they benefit from discussion with colleagues on the best next steps in the process.

Step 4: Determine the structure of the scenarios

Given the need to ensure that scenarios accurately reflect lived experiences of individuals in the real world coupled with the need for parity, it is important to identify the combinations of facets and levels, also known as structs, that will be included in the scenarios before constructing the scenarios. As described in the literature review, the inclusion of all facets and all levels in the scenarios could create an extremely large number of potential scenarios. One way to identify combinations of facets and levels for inclusion in the scenarios is through an extreme groups procedure plus variation (Ludlow et al., 2014). This procedure involves creating extreme scenarios, where each scenario contains the same level of each facet (high, medium, or low), that capture the extremes, or boundaries of the construct (Ludlow et al., 2014), while also allowing coverage of the range of the construct by describing individuals at the low, medium, and high levels. Once these extreme scenarios are created, additional scenarios written specifically to describe individuals at varying levels of the construct (in between the low, medium, and high levels) are developed, to ensure that the entire range of the construct is represented in the scenarios. This results in an extreme groups plus variation design for the development of the scenarios.

Additionally, specific to my instrument, the inclusion of all seven facets in each scenario would result in long scenarios, increasing the response burden on participants. Longer scenarios also increase the cognitive load to both read the scenario and compare your own skillset and experience to all seven facets, making it potentially difficult to respond. A solution is to systematically identify subsets of facets for inclusion in a particular scenario to ensure full coverage of the construct when all scenarios are created (Chang, 2017). This systematic inclusion of subsets of facets is completed so that each scenario has overlapping facets with at least one other scenario (Chang, 2017); this ensures that each facet is represented more than once in the overall instrument. With the exception of the practice item (described below), facets for inclusion in each individual scenario were chosen in sequential order. This decision was made because of the cyclical nature of the inquiry cycle and the fact that those engaging in this cycle typically complete the steps (or facets) in sequential order, at least initially.

Table 3.2 displays the inclusion of facets and levels in each scenario using this overlapping facets procedure (Chang, 2017). Four facets are included in each scenario to reduce the overall length of the instrument by limiting the total number of items, while also ensuring that each individual scenario is not too lengthy (including more than four facets in each scenario results in scenarios that are longer and may increase the cognitive load too much). The first facet, Ask questions (Q), is included in all of the scenarios as a starting point and to ground each scenario in the beginning of the inquiry process (Chang, 2017). The other three facets in each scenario were identified to ensure that each of the other six facets appears in at least six scenarios. An additional scenario was created at the medium level to be used as the practice item in the survey (described below). This practice item is labeled P in Table 3.2. The practice item was purposefully written at the medium level to ease participants into the scenario-style items.

The levels of the facets in each scenario were chosen based on the extreme groups plus variation design (Ludlow et al., 2014), while using my own personal experience to ensure that the combination of facet levels in each scenario is plausible. A scenario at each extreme level (low, medium, and high) was designed first, where all facets in the scenario are at the same level (i.e., a 1 for the extreme low scenario). This results in a total scenario score span from 4 (with four low facets each with an individual score of 1) to twelve (with four high facets each with an individual score of 3). The extreme scenarios have total scores of four (low), eight (medium), and twelve (high). Then, six scenarios were written to fill in the additional total scores remaining in the score span (i.e., 11, 10, 9, 7, 6, and 5). The levels of individual facets were varied in these six scenarios to allow for the creation of scenarios that span the entire construct. The decision of which level to assign to each facet in these six scenarios was determined based on my real-world experience and what made sense logically within each scenario. Three additional scenarios were written at the extreme levels to ensure full coverage of the construct and to provide some alternative options for these extreme scenarios. Table 3.2 lists the levels of each facet that are included in each scenario, denoting a high level with an H (and a facet score of 3), a medium level with an M (and a facet score of 2) and a low level with an L (and a facet score of 1). The total scenario score is also displayed in Table 3.2 for each scenario and is calculated by summing the individual facet scores in each scenario.

Table 3.2

Inclusion of facets by scenario for extreme groups plus variation procedure (Ludlow et al., 2014)s

Scenario Label	Facet 1: Ask questions (Q)	Facet 2: Identify data (D)	Facet 3: Examine data (E)	Facet 4: Interpret data to set goals (G)	Facet 5: Identify intervention (I)	Facet 6: Implement intervention (II)	Facet 7: Examine outcomes (O)	Total scenario score
A	H (3)	H (3)	H (3)	H (3)				12
B	H (3)	H (3)				H (3)	H (3)	12
C	H (3)			H (3)	H (3)	M (2)		11
D	H (3)		H (3)		M (2)		M (2)	10
E	H (3)	M (2)	M (2)	M (2)				9
F	M (2)			M (2)	M (2)	M (2)		8
G	M (2)		M (2)		M (2)		M (2)	8
P	M (2)	M (2)		M (2)			M (2)	8
H	H (3)	M (2)				L (1)	L (1)	7
I	H (3)	L (1)	L (1)	L (1)				6
J	M (2)		L (1)		L (1)		L (1)	5
K	L (1)			L (1)	L (1)	L (1)		4
L	L (1)	L (1)				L (1)	L (1)	4

Step 5: Develop the mapping sentences and construct the scenarios

In the next step of RGS development, a mapping sentence is developed which provides the layout for the scenarios. Modeled after Borg and Shye (1995) and Hackett (2014), the mapping sentence allows for the definition of the specific facets and levels included in the scenario, called structs, as well as the logical linkages between the facets using everyday language. The mapping sentence then provides a framework for the creation of the scenarios (Hackett, 2014) by combining individual structs within the mapping sentence into structuples (unique combinations of structs).

To create the mapping sentence, a template for the content in the mapping sentence was constructed first, which was then used to populate the mapping sentence and develop the scenarios. The template for the mapping sentence content, displayed in Table 3.3, is a matrix that includes all of the facets as rows and the three levels as columns. Each cell in the template includes a sentence stem, which helps describe the specific level of that facet (indicated in italics in Table 3.3), and the specific activities associated with that facet (indicated in non-italic font in Table 3.3). The narrative descriptions of an individual at each level of each specific facet that were written previously in step 3 of the RGS approach were utilized to construct this template. Note that the columns indicating the level of each facet include a code value of 1, 2, or 3. These code values are the facet scores and will be used to assign total scenario scores; as previously mentioned, the values were assigned so that a low facet is equivalent to a score of 1, a medium facet is equivalent to a score of 2, and a high facet is equivalent to a score of 3.

Table 3.3

Template for the content in the mapping sentence for the construct of using data to inform classroom instruction

	High: Code = 3	Medium: Code = 2	Low: Code = 1
Facet 1: Ask questions (Q)	*Is successful, successfully navigates* generating questions, narrowing down to specific questions, writing clearly defined questions, writing open-ended questions	*Sometimes has difficulty with, occasionally struggles with* generating questions, narrowing down to specific questions, writing clearly defined questions, writing open-ended questions	*Has difficulty with, struggles with* identifying questions, pinpointing areas for exploration, writing clearly worded questions, defining their area of exploration
Facet 2: Identify data (D)	*Is successful, does not need support* understanding what data can answer their question, identifying availability of data, organizing data clearly, thinking about data visualization, utilizing various data visualization techniques	*Often needs support with, occasionally struggles with* understanding what data can answer their question, identifying availability of data, organizing data clearly, thinking about data visualization, utilizing various data visualization techniques	*Struggles with, is unsuccessful with* understanding what data can answer their question, identifying availability of data, organizing data clearly, thinking about data visualization, utilizing various data visualization techniques
Facet 3: Examine data (E)	*Is always independent, is successful* gathering data, collecting data systematically, combining data sources, collecting data in an organized way, thinking critically about their data, analyzing their data, making observations about their data, comparing multiple data sources	*May need some support with, occasionally struggles with, is sometimes independent* gathering data, collecting data systematically, collecting data in an organized way, thinking critically about their data, analyzing their data, making observations about their data, comparing multiple data sources	*Struggles with, does not understand, needs support with* gathering data, collecting data systematically, thinking critically about their data, analyzing their data, making observations about their data, comparing multiple data sources
Facet 4: Interpret data to set goals (G)	*Can successfully, can independently* interpret meaning from their data analysis, seek support of others in validating their interpretations, generate hypotheses about how to improve student learning based on their interpretations, write clear and aligned goals from their data analysis, consider all intended and unintended consequences of their goals	*Occasionally struggles with, may need support with* interpreting meaning from their data analysis, accurate interpretation of meaning from their data analysis, generating hypotheses about how to improve student learning based on their interpretations, writing clear or aligned goals from their data analysis, considering intended and unintended consequences of their goals	*Struggles with, is unsuccessful with* interpreting meaning from their data analysis, accurate interpretation of meaning from their data analysis, generating hypotheses about how to improve student learning based on their interpretations, writing goals from their data interpretations, writing clear or aligned goals, considering

			intended or unintended consequences of their goals
Facet 5: Identify intervention (I)	*Can independently, can successfully* identify interventions to meet their goals, identify and plan appropriate professional development required for any interventions, plan to assess progress towards their goals with any intervention, identify the data variables necessary to document progress towards their goals	*May require some support to, sometimes struggles to* identify interventions to meet their goals, identify appropriate professional development required for any interventions, plan to assess progress towards their goals with any intervention	*Requires support to, cannot independently* identify interventions to meet their goals, identify potential sources of support to help plan interventions, determine if professional development is required for any interventions, plan to assess progress towards their goals with any intervention
Facet 6: Implement intervention (II)	*Actively seeks support with, is independent with* implementing intervention with fidelity, seeking support of colleagues during implementation of intervention, sharing progress of intervention, planning for student progress towards goals, collecting data about student progress towards goals	*Often needs support with, struggles with consistently* implementing intervention with fidelity, seeking support of colleagues during implementation of intervention, sharing progress of intervention, collecting or tracking student progress towards goals	*Has difficulty with, rarely seeks support with* implementing intervention with fidelity, seeking support of colleagues during implementation of intervention, sharing progress of intervention, collecting or tracking student progress towards goals
Facet 7: Examine outcomes (O)	*Can independently, can successfully* examine student outcome data previously collected, identify if student outcomes meet their stated goals, analyze the implementation of the intervention, determine if the intervention was implemented with fidelity, identify next steps in the inquiry cycle	*Benefits from support with, occasionally struggles with* examining student outcome data previously collected, identifying if student outcomes meet their stated goals, analyzing the implementation of the intervention, determining if the intervention was implemented with fidelity, identifying next steps in the inquiry cycle	*Has difficulty with, struggles with, is unsure about* examining student outcome data previously collected, identifying if student outcomes meet their stated goals, analyzing the implementation of the intervention, determining if the intervention was implemented with fidelity, identifying next steps in the inquiry cycle

The mapping sentence for the construct of using data to inform classroom instruction is presented below and was constructed using the mapping sentence template in Table 3.3. The facets are included in sequential order in the mapping sentence, as teachers generally proceed through an inquiry cycle in the order of the facets, at least initially. Although teachers may return to different facets in the cycle as they continue to engage in it, the first time through the cycle is generally in sequential order. Based on prior research utilizing RGS methodology (Reynolds, 2020), names and gender-specific pronouns are not utilized in the mapping sentence to reduce any possible bias that might be introduced by the use of names and/or gender-specific pronouns.

Mapping Sentence

Facet Q (Ask questions)

{high [is successful, successfully navigates]}

Teacher X *{medium [sometimes has difficulty with, occasionally struggles with]}*

{low [has difficulty with, struggles with]}

generating questions and narrowing down and writing clearly defined questions.

Facet D (Identify data)

{high [are successful, do not need support]}

They *{medium [often need support with, occasionally struggle with]}*

{low [struggle with, are unsuccessful with]}

understanding what data can answer their question, identifying and organizing their data clearly, and utilizing various data visualization techniques.

Facet E (Examine data)

{high [is always independent, is successful]}

Teacher X *{medium [may need some support with, occasionally struggles with]}* *{low [struggles with, needs support with]}*

gathering data systematically, thinking critically about and analyzing data, making observations about their data, and comparing multiple data sources.

Facet G (Interpret data to set goals)

{high [can successfully, can independently]}

They *{medium [occasionally struggle to, may need support to]}*

{low [struggle to, unsuccessfully]}

interpret meaning from their data analysis, seek the support of others to validate their interpretations, generate hypotheses about how to improve student learning based on their interpretations, write clear and aligned goals, and consider all consequences (intended and unintended) of their goals.

Facet I (Identify intervention)

{high [can independently, can successfully]}

Teacher X *{medium [may require some support to, sometimes struggles to]}*

{low [requires support to, cannot independently]}

identify interventions to meet their goals, identify and plan necessary professional development, plan to assess progress towards their goals, and identify the data required to document progress towards their goals.

Facet II (Implement intervention)

{high [actively seek support with, are independent with]}

They *{medium [often need support with, struggle with consistently]}*

{low [have difficulty with, rarely seek support with]}

implementing their intervention with fidelity, seeking the support of colleagues and sharing progress during implementation, planning for student progress towards goals, and collecting data about student progress towards these goals.

Facet O (Examine outcomes)

{high [is independent, is successful]}

Teacher X *{medium [benefits from support with, occasionally struggles with]}* *[low [has difficulty with, struggles with, is unsure about]}*

examining student outcome data collected from the intervention, identifying if these outcomes meet their stated goals, analyzing the implementation of the intervention for fidelity, and identifying next steps in the inquiry cycle.

The mapping sentence was then used to create the initial scenarios. Although the mapping sentence includes all seven facets, the scenarios were created utilizing only the facets from the mapping sentence that correspond to the facet inclusion plan for each scenario, as displayed in Table 3.2. Given the detailed description of each facet that is included in the mapping sentence, a straight translation of the mapping sentence into scenarios would result in long scenarios (even with only four facets per scenario), as well as repetition of the content in the scenarios across facet levels. The only variation in each scenario would be the specific facets included in each scenario and the verbs used to describe the level of those facets, and the repetition would likely be very taxing for participants to read.

Given the repetition that a straight translation of the mapping sentence into scenarios would create, a modified procedure was used to write the initial scenarios. This modified procedure involved selecting a subset of the activities described for each facet (from the mapping sentence) for each scenario to allow for variation in the description of each scenario. When selecting the subset of activities described for each facet from the mapping sentence, I ensured that each activity described for a particular facet is included in at least one scenario. For example, facet D (Identify data) has three separate activities: understanding what data can answer their question, identifying and organizing their data clearly, and utilizing various data visualization techniques. When writing the scenarios that contain facet D, I ensured that each of these three activities was represented in at least one scenario. Activities from a particular facet in a particular scenario were selected for coherency across facets to ensure that each scenario made sense when compared to my real-world experiences. As part of this process, some specific words in each scenario were modified from the mapping sentence to make the scenario more readable and grammatically correct. Please note that the activity of generating hypotheses about how to improve student learning based on teacher interpretations from facet G was not added until after the pilot administration of the survey, and thus does not appear in any scenarios until the full administration.

The initial scenarios, constructed using an extreme groups procedure plus variation (Ludlow et al., 2014) with four facets per scenario and selecting a subset of the activities described for each facet for a particular scenario, are presented in Table 3.4. The inclusion of four facets per scenario results in twelve scenarios total plus one practice item, with a word count range of 62-75 words in an individual scenario. Again, a value of 3 for a facet equals the high level, a value of 2 equals the medium level, and a value of 1 equals the low level. The order in which the initial scenarios are presented in Table 3.4 (from highest total score to lowest total score) represents the hypothesized continuum of the initial scenarios. I hypothesize that the low level scenarios will cluster at the bottom of the construct of using data to inform instruction, medium level scenarios will cluster in the middle, and high level scenarios will cluster at the top.

Table 3.4

Initial scenarios constructed using overlapping facets and an extreme groups plus variation procedure (Ludlow et al., 2014)

Scenario Label	Facet Levels (Total score)	Scenario
A	Q3, D3, E3, G3 (12)	Teacher A is successful generating and writing clearly defined questions. They do not need support understanding what data can help answer their question or utilizing various data visualization techniques. This teacher can always independently gather data systematically, think critically about and analyze data, and compare multiple data sources. They can successfully interpret meaning from their data analysis and write clear and aligned goals.
B	Q3, D3, II3, O3 (12)	Teacher B is successful generating and writing clearly defined questions. They do not need support understanding what data can answer their question or identifying and organizing their data clearly. They independently implement their intervention with fidelity and collect data about student progress towards their goals. This teacher can successfully analyze the implementation of the intervention for fidelity and identify next steps in the inquiry cycle.
C	Q3, G3, I3, II2 (11)	Teacher C successfully navigates generating and narrowing down to clearly defined questions. They can independently write clear and aligned goals from their data and consider all consequences (intended and unintended) of their goals. Teacher C can successfully identify and plan required professional development for their intervention and identify the data required to document progress towards their goals. They often need support with sharing progress during intervention implementation and planning for student progress towards their goals.
D	Q3, E3, I2, O2 (10)	Teacher D successfully navigates generating and writing clearly defined questions. This teacher can always independently think critically about and analyze data, as well as make observations about their data. Teacher D sometimes struggles to identify interventions to meet their goals and identify the data required to document progress towards these goals. This teacher benefits from support with examining student outcome data collected from the intervention and identifying if these outcomes meet their stated goals.

E	Q3, D2, E2, G2 (9)	Teacher E is successful generating and writing clearly defined questions. They often need support identifying and organizing their data clearly and utilizing various data visualization techniques. This teacher occasionally struggles with thinking critically about and analyzing data, making observations about their data, and comparing multiple data sources. They may need support with interpreting meaning from their data analysis and considering all consequences (intended and unintended) of their goals.
F	Q2, G2, I2, II2 (8)	Teacher F occasionally struggles with generating questions and narrowing down to clearly defined questions. They may need support with interpreting meaning from their data analysis and writing clear and aligned goals. Teacher F may require some support to identify interventions to meet their goals and to plan to assess progress towards these goals. They struggle with consistently implementing their intervention with fidelity and collecting data about student progress towards goals.
G	Q2, E2, I2, O2 (8)	Teacher G occasionally struggles with generating questions and narrowing down to clearly defined questions. This teacher may need some support with thinking critically about and analyzing data and making observations about their data. Teacher G sometimes struggles to identify interventions to meet their goals and to plan to assess progress towards these goals. This teacher benefits from support with examining student outcome data collected from the intervention and identifying next steps in the inquiry cycle.
P	Q2, D2, G2, O2 (8)	Teacher P sometimes has difficulty generating and writing clearly defined questions. They often need support with understanding what data can answer their question and utilizing various data visualization techniques. This teacher occasionally struggles with interpreting data from their data analysis and writing clear and aligned goals. They benefit from support with examining student outcome data collected from their intervention and identifying next steps in the inquiry cycle.
H	Q3, D2, II1, O1 (7)	Teacher H successfully navigates generating and writing clearly defined questions. They often need support with understanding what data can answer their questions and utilizing various data visualization techniques. They have difficulty with implementing their intervention with fidelity and planning for student progress towards goals. This teacher struggles with examining student outcome data collected from the intervention and identifying if these outcomes meet their stated goals.

I	Q3, D1, E1, G1 (6)	Teacher I is successful pinpointing areas for exploration and writing clearly defined questions. They struggle with understanding what data can answer their question and utilizing various data visualization techniques. This teacher needs support with thinking critically about and analyzing data and comparing multiple data sources. They struggle to interpret meaning from their data analysis, to seek the support of others to validate their interpretations, and to consider all consequences (intended and unintended) of their goals.
J	Q2, E1, I1, O1 (5)	Teacher J sometimes has difficulty narrowing down to and writing clearly defined questions. They need support with gathering data systematically and comparing multiple data sources. This teacher cannot independently identify interventions to meet their goals or plan to assess progress towards their goals. They struggle with analyzing the implementation of their intervention for fidelity and identifying next steps in the inquiry cycle.
K	Q1, G1, I1, II1 (4)	Teacher K has difficulty with narrowing down and writing clearly defined questions. They struggle with interpreting meaning from their data analysis and writing clear and aligned goals. Teacher K cannot independently identify interventions to meet their goals or identify the data required to document progress towards their goals. This teacher has difficulty planning for student progress towards goals and collecting data about student progress towards these goals.
L	Q1, D1, II1, O1 (4)	Teacher L struggles with generating and writing clearly defined questions. They are unsuccessful understanding what data can answer their question or identifying and organizing their data clearly. This teacher has difficulty implementing their intervention with fidelity and collecting data about student progress towards their goals. They struggle to identify if student outcomes meet their stated goals or to identify next steps in the inquiry cycle.

Step 6: Decide on the response options and survey instructions

The next step in RGS development is creating the response options for the scenarios and the instructions that participants will receive when engaging with the scenarios. Given that a scale of scenarios is likely new to most, if not all, participants, the instructions need to be specific and clear. Additionally, the use of a start-up, or practice, item to allow participants to practice with this new format can help reduce confusion on how to interact with these types of items for participants (Ludlow et al., 2014).

The response options for this instrument were developed in an effort to reduce social desirability bias in the responses (i.e., the desire to provide 'correct' responses to make an individual look better, or more socially desirable, than they may actually be). Participants are asked to compare their own experience and skillset to Teacher X; response options all start with "Teacher X is..." rather than "I am..." in an effort to reduce the desire to respond more positively to a particular scenario than is warranted (Antipkina & Ludlow, 2020). The use of comparison of a teacher's own experiences and skills in the response options follows Ludlow et al. (2014) by utilizing a comparative scenario response format. The initial instructions and response options, as well as the instructions for the practice item and the practice item itself, are presented below.

Instructions: These scenarios describe different teachers' experiences and skillsets when using data to inform their classroom instruction. Think about your own experience using data to inform your classroom instruction while reading each scenario. After reading each scenario, you will be asked to compare your own experience and skillset to the teacher's experience and skillset in that scenario.

Response options: How does your experience and skillset using data to inform classroom instruction compare to Teacher X?

- Teacher X is much less skilled than me
- Teacher X is slightly less skilled than me
- Teacher X is just like me
- Teacher X is slightly more skilled than me
- Teacher X is much more skilled than me

Start-up practice item: This item is presented to participants after the instructions and labeled "Practice Item: As the scenario-type survey may be a new experience, this practice item allows you to engage with a scenario. Please read the following practice scenario and compare your own experience to Teacher P".

Teacher P sometimes has difficulty generating and writing clearly defined questions. They often need support with understanding what data can answer their question and utilizing various data visualization techniques. This teacher occasionally struggles with interpreting data from their data analysis and writing clear and aligned goals. They benefit from support with examining student outcome data collected from their intervention and identifying next steps in the inquiry cycle.

Step 7: Testing congruence of theory and practice

The next step in RGS development is gathering expert feedback on the instrument and then the pilot of the instrument with a small group of individuals to field test the items. The pilot of this instrument was broken into a pre-pilot administration and a pilot administration. The pre-pilot administration utilized a convenience sample of family and friends who are current or retired teachers and focused on gathering feedback on the face validity of the survey (Ludlow et al., 2020). This feedback was mainly focused on whether the survey format, including the instructions and response options, make sense, as well as whether the actual scenarios are plausible to current and retired teachers. The pilot administration utilized Amazon Mechanical Turk (MTurk) as a sample and focused on testing whether the empirical results provided proof of concept for this methodology (Ludlow et al., 2020). After the pilot and revisions, the instrument was administered to a large sample of current teachers as a full survey administration to collect data. More details on the pre-pilot, pilot, and full administrations are described later in this section. The data were then analyzed using the Rasch rating scale model (Andrich, 1996; Wright & Masters, 1982) to identify if the responses to the scenarios by teachers supported the hypothesized continuum of the construct. The Rasch rating scale model provides a variable map (discussed below in the data analysis section) that places both items and people on the same scale and allows for the empirical assessment of the hypothesized continuum of the construct.

Expert feedback protocol: A group of master's and doctoral students in the Measurement, Evaluation, Statistics, and Assessment program at Boston College was asked to provide expert feedback on the scenarios in the instrument in mid-August 2021. Given that this expert feedback was gathered during August 2021 and that the COVID-19 pandemic remained a concern, this feedback was gathered remotely via email. In addition, three recent doctoral graduates from Boston College who employed the RGS methodology in their dissertations were also asked for feedback via email. This group of students and graduates was sent an email that contained the request for feedback, along with a Word document attachment that included my description of the construct, overall descriptions of teachers at high, medium, and low levels of the construct, and the thirteen initial scenario items. Please see Appendix A for the text of the email.

Feedback was also gathered from two district-based curriculum coordinators who work directly with teachers after the modifications from the master's and doctoral student/graduate comments were made to gather information on the face validity of the revised instrument. This feedback was focused on ensuring that the language in the scenarios reflects teachers' experiences and that the scenarios both make sense and are plausible for an audience of teachers.

Feedback received: The main pieces of feedback received from the group of master's and doctoral students/graduates focused on the length and repetition of the scenarios. Comments noted that the length of the scenarios with four facets was too long, making it difficult to maintain focus throughout the entire survey and prompting concerns that participants would stop reading each scenario in detail given the length of each one. In addition, the repetition of content and vocabulary used in the scenarios was also noted as a potential problem that could deter participants from reading each scenario in detail. Particularly, there were comments about the use of facet Q (Ask questions) at the beginning of every scenario and the repetition that this caused in reading each scenario. Commenters noted that this repetition could generate participant fatigue or prompt participants to skim over the content in each scenario, rather than reading it in detail, because each one started so similarly.

The order of the scenarios presented in the instrument itself is chosen so that the scenarios are not presented in ascending or descending order based on total scenario score, but rather vary the order of the scenario scores to present scenarios with a higher score, followed by one with a lower score, followed by one with a medium score, and so forth. Because the scenarios were named alphabetically based on total scenario score (see Table 3.4), this means that when the scenarios are presented to respondents, they are not presented in alphabetical order. One of the curriculum coordinators wondered why the "names" of the teachers in each scenario (i.e., Teacher A, Teacher B, etc.) were not in alphabetical order when scenarios were presented. This coordinator noted that they kept trying to identify a pattern in the teacher names and that it detracted from their focus on the content of the scenarios.

In response to this feedback, the scenarios were redesigned with three facets per scenario and without forcing facet Q to be in every scenario, using the same methodology described for the development of the initial scenarios. The initial decision to include facet Q in every scenario was based on methodology described in Chang (2017); however, Chang (2017) had a specific theoretical reason to ensure that one specific facet was included in all scenarios. For my instrument, no one facet is more important in the inquiry cycle than any other facet; therefore, requiring one facet to be included in every single scenario is not necessary. By allowing all of the facets to vary in each scenario and using three facets per scenario, each facet will appear at least three times in the entire instrument.

Table 3.5 displays the revised facet inclusion for each scenario with three facets per scenario. When three facets are included in each scenario, the total scenario scores range from 3 (three low facets) to 9 (three high facets). The medium level scenario has a total score of 6. To ensure that the entire range of scores is reflected in the scenarios, a minimum of seven scenarios is required. In response to the comments about the overall length of the instrument, no additional scenarios were constructed with the same total score (as was done in the initial set of scenarios). The practice item, however, was retained, resulting in a total of eight scenarios. Facets were identified for inclusion in a particular scenario using a spiral-type method (Chang, 2017), where three adjacent facets are identified for a particular scenario, and then the next three adjacent facets are identified for the next scenario, ensuring that one facet overlaps between both scenarios. Similar to the initial round of development, I identified the level of each facet in each scenario based on my real world experience. In addition, the scenario labels and teacher names were renamed so that when the survey is presented to participants, the teacher names are in alphabetical order. Table 3.5 displays the scenarios by total score, not alphabetically, although respondents will be presented the scenarios in alphabetical order.

Table 3.5

Revised inclusion of facets by scenario utilizing extreme groups plus variation (Ludlow et al., 2014)

Scenario Label	Facet 1 (Q)	Facet 2 (D)	Facet 3 (E)	Facet 4 (G)	Facet 5 (I)	Facet 6 (II)	Facet 7 (O)	Total scenario score
C	H (3)	H (3)	H (3)					9
G			H (3)	H (3)	M (2)			8
A					H (3)	M (2)	M (2)	7
E	M (2)	M (2)					M (2)	6
B			M (2)	M (2)		L (1)		5
D	M (2)				L (1)	L (1)		4
F		L (1)		L (1)			L (1)	3
P	M (2)		M (2)		M (2)			6

Additionally, more specific feedback was received from the master's and doctoral students/graduates about the content of some of the scenarios and the answer choices.

The wording “next steps in the inquiry cycle” in facet O was identified as potentially problematic because the inquiry cycle is not discussed anywhere else in the content of the survey or the instructions. Given that the terminology of inquiry cycle could be considered jargon, it was removed from all scenarios and replaced with “next steps”. In addition, the description of writing clear and aligned goals in facet G was identified as confusing in the scenarios that contained it because it seems to “come out of nowhere” and the purpose of these goals was not clear as it was written in the scenarios. Scenarios containing this skill were reworded to make it clear that the goals were specific to student learning. Facet Q was also identified as needing further detail to clarify that the questions teachers were thinking about and writing were specific to student learning; scenarios containing facet Q were reworded to reflect this comment. Comments also noted that the vocabulary and sentence structure in the individual scenarios should be more varied to make them more interesting to read, as well as to ensure that they reflect the lived experiences of teachers; scenarios were reworded to vary both the vocabulary and sentence structure. Finally, the use of “always” in some scenarios was identified as potentially problematic because participants might interpret this too literally and assess themselves incorrectly when reading this scenario. The use of “always” before specific descriptors in the scenarios was removed.

Both coordinators noted that they believed that each scenario was plausible for teachers once the particular elements that they identified as potentially confusing were edited, providing some evidence of face validity. Additionally, both coordinators noted that in facet I, teachers would not be planning the professional development required for their chosen intervention, but rather would identify and seek out whatever professional development was required. Based on this feedback, any reference to planning professional development in the scenarios was removed. One of the coordinators thought the terminology "data visualization techniques" in facet D was possibly confusing, as some teachers would not know what this means. The scenarios that mention this skill were reworded to clarify what data visualization techniques means.

Feedback was also received on the instructions and answer choices. A suggestion was made from the master's and doctoral students/graduates to bullet the instructions to make it easier for participants to read and understand the instructions and to understand how to interact with scenario-style items; the instructions were changed to bullets. Additionally, the use of "experience and skillset" in the response options directions was identified as confusing, especially by the curriculum coordinators, as these can be interpreted differently, and the answer choices specify skills only. Both curriculum coordinators also noted that the word "experience" could be interpreted as both an adjective and a noun, and that the answer choices are really looking for more of a descriptor of proficiency or savvy with the descriptions in the scenarios. Based on this feedback, the word experience was removed from the response options directions and the directions were modified to describe the scenarios as depicting teachers' skillsets. The answer choices were reworded to ask teachers to compare their skills to the skills of the teacher in each scenario.

The feedback on the overall design of the scenarios and the specific feedback on content from both the master's and doctoral students/graduates and the curriculum coordinators were utilized to revise the initial scenarios to the pre-pilot set of scenarios, displayed in Table 3.6. The scenarios are displayed in order of total score in Table 3.6, not alphabetically. As previously noted, scenarios have been renamed alphabetically in the order in which they will appear to participants based on feedback from the curriculum coordinators. The pre-pilot scenarios have a word count range of 61 words to 83 words. The modified response options and instructions from this feedback follow Table 3.6.

Table 3.6

Pre-pilot scenarios constructed using overlapping facets and an extreme groups plus variation procedure (Ludlow et al., 2014)

Scenario label	Facet levels (Total score)	Scenario
C	Q3, D3, E3 (9)	Teacher C is successful generating and writing clearly defined questions that clarify their thoughts on student learning. They do not need support understanding what data can help answer their questions or identifying different techniques to display their data. This teacher can independently gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.
G	E3, G3, I2 (8)	Teacher G is able to think critically about data that can help answer their questions about student learning and successfully make observations about their data while comparing multiple data sources. They can also independently interpret meaning from their data analysis to develop student learning goals and consider all potential consequences of these goals. Teacher G may require some support to both identify interventions to help attain these goals and identify the data required to document progress towards these goals.
A	I3, II2, O2 (7)	Teacher A can successfully identify interventions to meet their student learning goals and plan to assess progress towards these goals. However, they sometimes struggle to implement their chosen intervention with fidelity and to collect data about student progress towards their goals during the intervention. Teacher A benefits from support with examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.
E	Q2, D2, O2 (6)	Teacher E occasionally needs help distilling their thoughts and wonderings about student learning into clearly defined questions. They often need support understanding what data can help answer these questions and organizing their data clearly once they have identified it. This teacher occasionally struggles with analyzing the implementation of the intervention they have chosen for fidelity and identifying their next steps once the intervention is complete.
B	E2, G2, II1 (5)	Teacher B may need some support thinking critically about and analyzing data that helps answer their questions about student learning. Once they complete their data analysis focused on these questions, they sometimes have trouble extracting meaning from this

		analysis and often do not seek the support of others to validate their interpretations. When they choose an intervention for student learning, Teacher B struggles to both plan for student progress towards their student learning goals and to share progress with colleagues during their implementation.
D	Q2, I1, II1 (4)	Teacher D sometimes has difficulty narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they require support to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D often holds back on sharing progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.
F	D1, G1, O1 (3)	Teacher F struggles to understand what data can answer their questions about student learning and to organize that data clearly. This teacher has trouble interpreting results from their data analysis related to their questions about student learning and struggles to write clear and aligned goals for student learning based on their analysis. Once they have implemented a student learning intervention, they need support to examine student outcome data collected from this intervention and to identify their next steps.
P	Q2, E2, I2 (6)	Teacher P sometimes has difficulty pinpointing and writing clearly defined questions that describe their thoughts about student learning. They may need some support systematically gathering data related to these questions and comparing multiple data sources. This teacher sometimes has trouble identifying interventions to meet their goals related to student learning or planning to assess progress towards these goals on their own and looks to others for support.

Pre-pilot response options: How do your skills using data to inform classroom instruction compare to Teacher X?

- Teacher X is much less skilled than me
- Teacher X is slightly less skilled than me
- Teacher X is just like me
- Teacher X is slightly more skilled than me
- Teacher X is much more skilled than me

Pre-pilot instructions: As previously mentioned, feedback on the instructions included bulleting the list of directions to make it easier for participants to read, particularly because this is a new type of scale format for many participants and the use of bullets makes it less likely that participants will just skim through the directions without reading them carefully. Additionally, a sentence was added describing that the word 'data' in the context of these scenarios covers all types of data, not only assessment data. This addition was based on some feedback received during the expert review, as well as feedback from the pre-dissertation instrument, where some participants emailed me to say that they could not complete the survey without knowing what kind of data was being referenced in the survey. As described in Chapter 2, many types of data should be utilized when using data to inform classroom instruction and this practice should not be focused solely on assessment data (Brown et al., 2017; Marsh et al, 2006; Mandinach & Gummer, 2016b). The revised set of instructions based on this feedback is presented below.

These scenarios describe different teachers' skillsets when using data to inform their classroom instruction. Data refers to a wide variety of data sources in these scenarios, not solely assessment data.

- *Reflect on your own skills using data to inform your classroom instruction while reading each scenario.*
- *Compare your own skillset to the teacher's experience in the scenario.*
- *Choose the response that most accurately reflects your comparison.*

The next sections describe the population, sampling procedures, and details for the pre-pilot, pilot, and full administrations.

Population Definition

As described previously, the construct of using data to inform classroom instruction focuses on teachers, as they are the ones who use data to inform their classroom instruction. In addition, administrators may also use data to help inform classroom instruction, although this work may be more targeted at a grade or school level. Thus, this instrument is targeted to teachers and school administrators. In addition, because the federal and state laws that require this type of work focus on public school teachers and administrators, the population definition is further refined to public school teachers and administrators working in public education at the time of the full administration.

Sampling Procedures

The sampling procedures specific to each administration (pre-pilot, pilot, and full administration) are described in each administration section below. All administrations were administered via Qualtrics, an online survey administration software. Identifying the target sample size for each administration is important to ensure that the estimates obtained from the data from each administration are stable and accurate. Although there is no specific rule for sample size with the Rasch model, there are some guidelines in the literature that can help guide sample size determination. Lord (1980) notes that the Rasch model, in comparison to other IRT models, can be utilized with small sample sizes, because estimates from the Rasch model are relatively accurate with small sample sizes. For pilot studies that are well-designed in terms of sampling from the population, sample sizes of 30 can be enough to produce item and person estimates that are stable within plus or minus one logit (Linacre, 1994). A sample size of 100 utilizing the Rasch model should be sufficient for person and item estimates to be stable within plus or minus one-half a logit (Linacre, 1994). In general, a sample size of 100 is usually acceptable (Chen et al., 2014; Linacre, 1994; Wright, 1977) and a sample size of 400 is generally sufficient (Wright, 1977). In addition, for a Rasch rating scale model, Linacre (2002a) recommends that there are least 10 observed responses for each rating scale category of each item on the survey.

The pre-pilot, pilot, and full instruments each have eight scenario items (one practice item and seven scenarios) with five rating scale categories (response categories). The pre-pilot utilized a convenience sample and was focused on providing face validity and proof of concept, so a smaller sample size of 20 respondents was planned for the pre-pilot. For the pilot and full administrations, a sample size of at least 150 respondents was targeted. The target sample size of at least 150 respondents follows the sample size recommendations discussed here, and allows for the possibility of more than ten observed responses for each rating scale category of each item on the survey. For the full administration, although there was a target of 150 respondents, it was anticipated that the actual number of respondents would be higher, given the larger target population (more detail is provided in the full administration details below).

Pre-Pilot Administration

The pre-pilot administration utilized a convenience sample of my friends and family members who are current or retired teachers and was administered in early September 2021 (September 10-15). The pre-pilot recruitment email (shown in Appendix B) was sent to nine current or former teachers, and some of these people forwarded the email to other teachers that they know. The goal of the pre-pilot was to gather information on the face validity of the survey, although the data from the pre-pilot was also analyzed using a Rasch rating scale model to evaluate if there was some proof of concept based on this small sample. For the pre-pilot, a sample size of twenty (20) participants was the goal.

Feedback and survey responses were received from twenty-two (22) respondents on the pre-pilot. The feedback received from the pre-pilot provided evidence of face validity. Multiple respondents replied that the scenarios were plausible and that they had identified themselves or other colleagues when reading each scenario. One respondent noted that this actually made responding to the survey easier, because they were able to connect the scenario to an individual that they knew, and then assess their own skillset compared to that individual's skillset. Additional feedback noted that the scenarios required a larger cognitive load than a typical survey and took longer to respond to, which is to be expected when using the RGS approach to design a survey. One respondent noted that they had to look up the word fidelity to understand what it meant in the context of this survey. Removing this word from the scenarios and replacing it with a more frequently used word in education may be worthwhile. Finally, another respondent noted that they felt it was easy to mark themselves as a higher skillset than the teachers in the highest (or hardest) scenarios and thought others may feel the same way. This may be some evidence of social desirability bias; to address this concern, the pilot administration included a social desirability scale to assess the extent of social desirability bias in the responses to the scenario items.

As mentioned previously, a higher cognitive load and a longer time to take this type of survey is expected. Although this requires more work from respondents, the comments that many were able to identify real-life individuals in each scenario and that some were then able to compare themselves to that real-life individual should make the task of responding to this survey slightly easier. Correctly estimating the length of time that it will take individuals to respond to the survey and providing that information in advance can also help with this concern. Results from the Rasch analysis of the pre-pilot data, as well modifications to the scenario items based on this analysis, are presented in the Results section in Chapter 4.

Pilot Administration

The pilot administration was administered from September 22-25, 2021. The pilot administration utilized Amazon Mechanical Turk (MTurk) for sampling. Amazon MTurk is an online crowdsourcing marketplace where requestors can publish tasks (in this case, a survey) and individual workers can choose to accept and complete these tasks. Requestors can set parameters on the individuals who respond to their task. Individual workers who complete tasks on Amazon MTurk are paid for their work. For the pilot administration, I required that respondents worked in the field of education and compensated workers $2 per survey (respondents could only respond to the survey once). My goal was a sample size of 150 respondents. A total of 169 responses were received from the pilot.

In addition to the revised scenario items and demographic questions from the pre-pilot, the pilot survey included a social desirability scale to measure the extent of social desirability bias in the responses to the scenario items. The social desirability scale utilized was the Marlowe-Crowne (M-C) Form C (Reynolds, 1982), a shortened version of the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960). The items included on the Marlowe-Crowne Social Desirability Scale and on the M-C Form C were chosen for those scales because they focus on behaviors that are culturally approved but are relatively infrequent in the general population, and have minimal implications of psychopathology (Crowne & Marlowe, 1960; Reynolds, 1982). The Marlowe-Crowne Social Desirability Scale includes 33 items, which can be lengthy for respondents to respond to in addition to the other items included on the particular survey instrument. Reynolds (1982) investigated a variety of shortened forms that utilize some of items from the Marlowe-Crowne Social Desirability Scale; the M-C Form C was the preferred form, both because of its strong psychometric properties and the short length of the form (13 items). The results from the pilot administration, as well as modifications to the scenario items based on those results, are presented in the Results section in Chapter 4.

Full Administration

The full administration was administered from November 15-26, 2021 in one medium-sized city (approximately 1,200 teachers) in Massachusetts. The full administration consisted of an email that was sent to all teachers two days before the survey link was sent to introduce myself and the survey (please see Appendix F for the text of this email). The survey link was sent via email two days later and the survey window was open for two weeks (please see Appendix G for the text of this email). To encourage participation, four $50 Amazon.com gift cards were raffled off to respondents who provided their email address at the end of the survey. The survey was anonymous and the email addresses were used only to conduct the gift card raffle. Email addresses were stored separately from survey responses. The results from the full administration are presented in the Results section in Chapter 4.

Data Analysis

As previously described, I synthesized the existing literature on frameworks and cycles of inquiry for DDDM to develop a continuum of skills, or facets, that teachers utilize while using data to inform their classroom instruction and I utilized a Rasch model to analyze the results. As described in the literature review, the Rasch model conceptualizes a construct (in this case, using data to inform classroom instruction) as a unidimensional construct that spans a continuum from low to high levels. The items written to measure this unidimensional construct are hypothesized to be placed along this hierarchical construct, as shown by the scenario scores on the person-item variable map (described later in this section).

The Rasch model measures item difficulty and assumes that item discrimination is the same for all people for whom the variable is measured and also assumes that guessing does not exist. Although multi-parameter models (which measure discrimination and/or guessing in addition to item difficulty) will always fit the data better than a one-parameter model (i.e., a Rasch model), a Rasch model can be very useful for diagnostic purposes when the purpose of measuring a variable is to place a person on the continuum or hierarchy for that variable with the intention of describing that person's location on that continuum. This is particularly useful for instruments that place a person on the continuum based on their responses. The description of that person's location on the continuum describes what a person at that location looks like and/or can do, and can be utilized to identify the needs for a person at that location in terms of growth on that scale. When used as part of the RGS methodology, the detailed descriptions of the facet levels are utilized for this detailed description of a person at a particular location on the continuum.

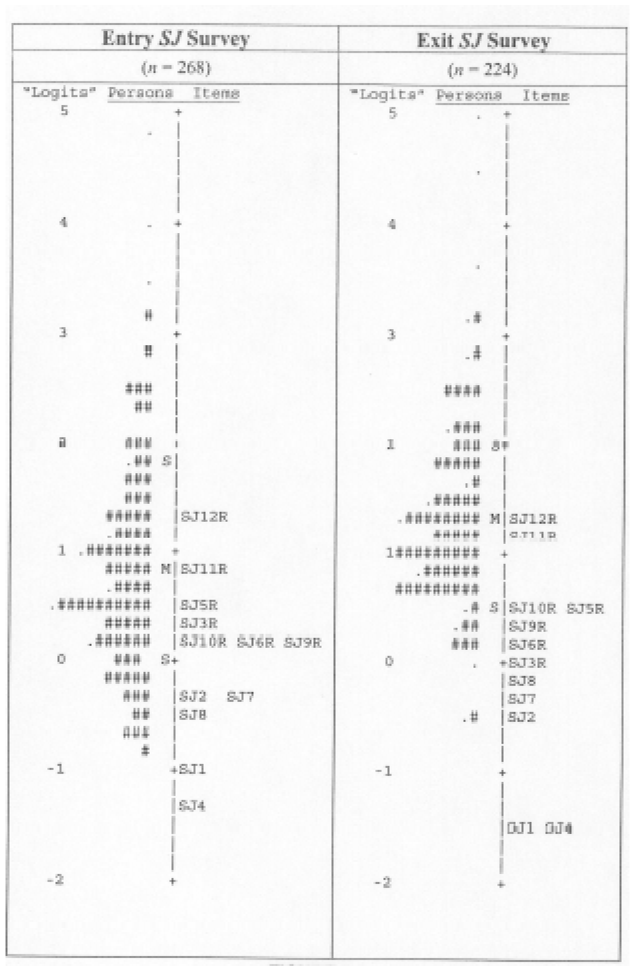
The principles of Rasch measurement, which are utilized when designing a Rasch measurement instrument (including the RGS methodology), include the following: unidimensionality, variation, the uniform spread of items along the continuum, a hierarchical nature of item progression along the continuum, equally discriminating items, independent items, and well-fitting items as a match between theory and data (Ludlow et al., 2014; Rasch, 1966). Unidimensionality means that the items on the scale measure a single construct, while variation means that the items vary from easy to difficult with an appropriate spread of difficulty in between easy and difficult. The uniform spread of items along the continuum refers to the representation of items spread along the continuum of easy to difficult like a ladder, without large gaps between item locations, which would indicate a missing “ladder rung” or level of the construct. The hierarchical nature of item locations along the continuum means that items are spread along this ladder, or continuum, from relatively easier to more difficult to endorse tasks consistent with the hypothesized progressive structure of the construct. Equal discrimination means that all items differentiate equally well between people with low ability and people with high ability. Item independence indicates that the probability of answering an item “correctly” is not dependent on the specific answer to a prior item. Finally, item fit between theory and data means that the resulting data fit with the theory about the construct under investigation (meaning that the principles described here are all supported by the data).

Some important points about the use of these Rasch measurement principles in instrument design include the fact that the construct under investigation should be clearly defined by theory and experience and that this clear definition should be utilized when designing items. Both of these points are key steps in the RGS methodology and are crucial for providing evidence of content validity. Additionally, part of the process of designing these items requires an a priori hypothesis about the hierarchical structure of the items based on the clear definition of the construct. Once the instrument is administered, the empirical results are compared to the a priori hypothesis of the structure of the items, which helps provide content validity if the empirical structure matches the a priori hypothesis. The model is also tested for statistical goodness of fit by comparing the observed and expected response patterns using residual analyses, where a residual is the difference between the observed response that a person gave and their expected response based on the Rasch rating scale model. Residual analyses can provide suggestions for revisions as needed.

Figure 3.2 provides an example of the utilization of the Rasch model to analyze responses from a particular scale. It displays the Learning to Teach for Social Justice-Beliefs Scale for students entering a teacher preparation program and the same students exiting the program (Ludlow, Enterline, & Cochran-Smith, 2008). The person-item variable maps displayed in Figure 3.2 place the respondents on the left hand side of the map and the individual items on the scale on the right hand side of the map. The scale is in logits (described later in this section). These variable maps are used to identify where an individual teacher candidate lies in comparison to the items on the "entry to the program" survey and they provide information on the items that student teachers need additional support on in order to progress upwards along the continuum at the time of exit from the program. These maps also help evaluate the a priori hypothesis about the structure of the learning to teach for social justice construct by displaying the items from "easiest" (at the bottom of the map) to "hardest" (at the top of the map), while also providing evidence of measurement invariance for the scale based on the similarity of the maps at the times of entry and exit from the program.

Figure 3.2

Variable maps from the Learning to Teach for Social Justice-Beliefs Scale (entry and exit administrations). (Ludlow, Enterline, & Cochran-Smith, 2008)



I used a rating scale Rasch model for this research, as opposed to a simple dichotomous Rasch model or a partial credit Rasch model, given the structure of the answer choices for this instrument, which require participants to rate how similar their skillsets are to the skills described in the scenario. A simple dichotomous Rasch model is not appropriate for this analysis, given that there are more than two response options. Although a partial credit Rasch model will always fit the data better than a rating scale Rasch model, a partial credit model should only be used in situations where participants must get portions of an item right to receive the partial credit available for the item. For example, a partial credit Rasch model could be used to model responses for math items where a participant must get the first part of an item correct to receive credit on the second part of the item. This is not applicable for scales developed with response options where participants do not receive partial credit for selecting a particular response and where items are designed so that the amount of the construct required to move from a response of 2 to a 3 (for example, from Disagree to Neither Agree nor Disagree) is the same across all items in the instrument.

The Rasch rating scale model equation is shown in Equation 3.1. This equation shows that the probability of person *n* answering in category *x* to item *I* (π_{nix}) is a function of the location (or difficulty) of item *I* (which is represented by δ) and the location of the kth step in each item relative to that item's scale value (which is represented by τ), known as the threshold parameter. The threshold parameter can be thought of as the difficulty of moving from one category (or score) to the next category (or score) on an item (for example, from a 1 to a 2). The Rasch rating scale model indicates that the value of one answer option (for example, the value of "scoring" a 3) is the same across all items in the instrument and that the increase from one answer option to the next is monotonic and the same across all items.

$$\pi_{nix} = \frac{e^{\sum_{j=0}^{x_{ni}}\left[\beta_n - \left(\delta_i + \tau_j\right)\right]}}{\sum_{k=0}^{m} e^{\sum_{j=0}^{k}\left[\beta_n - \left(\delta_i + \tau_j\right)\right]}} \qquad \text{Eq. 3.1}$$

IRT models, including the Rasch model, estimate both person ability and item difficulty in logits. Logits are the natural logarithm of the odds of an event occurring.

Logits are used because a logistic function is the most reasonable depiction of person ability, as it is bounded by zero and one. A logistic function becomes asymptotic when the probability approaches zero or one. In addition, a logistic function does not assume that an increase or decrease in person ability is associated with the same amount of change in the probability of getting an item right, as would be assumed in a linear function (such as linear regression).

Person ability is calculated by estimating the odds of an event occurring (which is the probability of an event occurring divided by the probability of the event not occurring) and then taking the natural logarithm of the odds. This results in a person ability estimate that is in logits and is the natural log of the odds of getting an item right that has a zero logit value. Values over zero for people indicate higher levels of ability and values below zero indicate lower levels of ability. Item difficulty is calculated in the same way and results in an item difficulty estimate that is in logits and is the natural log of the odds of a person with a zero value for ability getting the item correct. Values over zero for items indicate more difficult items and values under zero for items indicate easier items.

In addition to examining the person-item variable maps to evaluate the a priori hypothesis of the item structure, descriptive statistics, person and item separation statistics, Andrich thresholds, the rating scale category structure, Category Characteristic Curves (CCCs), and fit statistics were also examined as part of the data analysis plan. Descriptive statistics include the item mean and standard deviation for the items. The descriptive statistics can help provide some initial information about the ordered structure of the items by ordering items by their mean values. The person and item separation statistics provide some indication of how well the people are separated on the instrument and how well the items are separated on the instrument. This indicates how well the items differentiate between people of low and high ability on the construct. The Andrich thresholds indicate the point at which an individual has a 50% probability of choosing the next highest response option for the item. Andrich thresholds can be plotted alongside the respondents on the survey on a map, similar to the person-item variable map. Examining the Andrich thresholds can be useful to identify which response options respondents are most likely to select for particular items. The rating scale category structure is examined by looking at the order of the observed average of person estimates by item response category and evaluating if these averages are ordered as intended, while increasing monotonically in category order. The Category Characteristic Curves (CCCs) for the entire instrument are also examined to evaluate if the Andrich thresholds increase monotonically as the categories increase and to determine if the categories are in the correct, ordered progression. The CCCs also provide information on the probability of each response category across the entire instrument, which provide an indication of how easy or difficult it is to respond to each category in the rating scale structure.

Finally, fit statistics were examined to identify items and people with inconsistencies in their responses, which can indicate problems, or misfit. These statistics provide a way to analyze model fit by looking at person and item residuals. Fit statistics utilize the variance in the responses and the magnitude of the residuals and include both INFIT (information weighted fit statistic) and OUTFIT (unweighted fit statistic) statistics. Both INFIT and OUTFIT are mean-square fit statistics, which indicate the size of randomness (or distortion) in the responses; the expected values for each are 1

(Linacre, 2002b). A mean-square value close to 1 indicates that there is not much distortion in the measurement system (Linacre, 2002b). The INFIT statistic can provide an indication that responses to an item are problematic across all people on the scale; although both person and item INFIT can be calculated, it is generally used to assess the quality of items across people by looking at item INFIT. The INFIT statistic is a weighted mean square residual, weighted by the variance of getting the item right. In general, a value greater than 1.4 for an INFIT statistic indicates that there are a relatively large number of unexpected responses. For a person, this means that there are consistent inconsistencies within their responses. For an item, this means that there are consistent inconsistencies across people in the dataset for the item. These people and/or items should then be investigated to identify why these consistent inconsistencies are occurring. The OUTFIT statistic looks at outlier responses across the dataset to identify unusual responses and is an unweighted mean square residual. It can provide an indication of items that have at least one highly unexpected response across all respondents or people that have highly unexpected responses on at least one item. Similar to the INFIT statistic, a value larger than 1.4 (or sometimes 1.3) can indicate misfit for the OUTFIT statistic. Although the OUTFIT statistic is calculated for both items and people, it is often used to identify people with unexpected responses.

Standardized fit statistics (ZSTD) are also examined for both the INFIT and OUTFIT statistics. The ZSTD statistics are *t*-tests to test the hypothesis of whether the data fit the model (Linacre, 2002b) and are treated as z-scores. They show the significance of whether the data fit the model; the expected value is zero, and values less than around -2.0 indicate the response data are too predictable (meaning small residual variation) while values greater than around +2.0 indicate lack of predictability (meaning large residual variation) (Linacre, 2002b).

In addition to the Rasch analysis described above, a Principal Components Analysis (PCA) was performed on the residuals from the pilot and the full administration (Ludlow, 1983); given the small sample size from the pre-pilot, the residual PCA was not performed for the pre-pilot results. The purpose of the residual PCA is to evaluate whether there is an unexplained, unintended construct remaining in the residual correlation matrix. Ideally, there will be no evidence of an unexplained construct in the correlation matrix, which provides evidence that the residuals are random and that the assumption of unidimensionality for the Rasch model is met. The residual PCA can be compared to a PCA of randomly generated data to help evaluate whether there is an unexplained construct in the residual data.

Chapter Summary

This chapter began with a refresher on the construct of using data to inform classroom instruction that was developed and detailed in Chapter 2. It then described the instrument development process for the pre-dissertation instrument to provide background on the work that spurred the focus of this dissertation. Next, it detailed the steps of the RGS methodology in the context of the construct of using data to inform classroom instruction. It described the process that was used to develop the scenarios for the pre-pilot administration. Then, the population definition and sampling procedures for the pre-pilot, pilot, and full administrations were described. Finally, the chapter closed with a detailed description of the data analysis plan and provided background on the Rasch model.

Chapter 4 : Results

This chapter provides the results from the pre-pilot administration, the pilot administration, and the full administration of the UDII scale. It discusses descriptive statistics and the Rasch analyses for all three administrations, as well as the Principal Components Analysis (PCA) of the residuals and randomly generated data for the pilot and the full administrations. It ends with an interpretative person-item variable map to aid in the interpretation of the raw scores for survey respondents, making the survey results more user-friendly for schools and districts.

Pre-Pilot Administration Results

The data from the pre-pilot administration were analyzed to provide initial proof of concept for the hypothesized hierarchical structure of the scenarios, as well as to make any necessary modifications to the scenario items before the pilot administration (please see Appendix C for the pre-pilot survey). This hypothesized hierarchical structure is displayed in Table 4.1, with the hypothesized most difficult scenario listed first (C) and the hypothesized easiest scenario listed last (F). Responses were received from 22 current or retired teachers and there were no missing data in the pre-pilot data. The answer choices were coded as follows:

- Teacher X is much less skilled than me: 5
- Teacher X is slightly less skilled than me: 4
- Teacher X is just like me: 3
- Teacher X is slightly more skilled than me: 2
- Teacher X is much more skilled than me: 1

Based on this coding structure, items with a higher mean value were "easier" to respond to, meaning that it was easier for a respondent to identify themselves as a higher skill level than the teacher in the scenario. Items with a lower mean value were “harder” or more difficult to respond to, meaning that it was harder for a respondent to identify themselves as a higher skill level than the teacher in the scenario. The hypothesized difficulty of a particular scenario item is denoted by the combination of the level of the facets in the scenario. Table 4.1 displays the scenarios labeled by their teacher name first (a letter), followed by the levels of the facets in the scenario (where H equals high, M equals medium, and L equals low), and then the actual facet and level combinations in the scenario. For example, scenario C is hypothesized to be the most difficult scenario for a respondent to rate themselves as a higher skillset than the teacher in the scenario, denoted by HHH (high levels for all three facets included in the scenario).

Descriptive Statistics

Table 4.1 displays descriptive statistics from the pre-pilot administration. The items are ordered by their hypothesized hierarchical continuum in this table, with the hypothesized hardest item first and the hypothesized easiest item last. Although the hypothesized continuum will be evaluated using a Rasch model, looking at the descriptive statistics in this way can provide some evidence to evaluate the hypothesized structure to provide proof of concept for the construct of using data to inform classroom instruction.

The descriptive statistics provide some evidence of proof of concept of the hypothesized scenario structure. Scenario P (MMM), which is the practice item, has a higher mean value than scenarios B (MML) and D (MLL), although scenario P should be slightly more difficult to respond to than B and D given the levels of its facets (which should result in a lower mean value for scenario P). However, this could be some evidence of a start-up effect since P is the first item that respondents encounter and many (if not all) respondents have not engaged in a scenario-style survey before. Scenarios B (MML) and D (MLL) have very similar means, although scenario B should be more difficult to respond to at a high level than scenario D given the levels of the facets in these scenarios.

Table 4.1

Descriptive statistics from the pre-pilot ordered by the hypothesized scenario structure

Item	Mean	Standard Deviation
C: HHH (Q3, D3, E3)	2.64	0.58
G: HHM (E3, G3, I2)	3.27	0.94
A: HMM (I3, II2, O2)	3.68	0.65
E: MMM (Q2, D2, O2)	4.27	0.70
P: MMM (Q2, E2, I2)	4.59	0.67
B: MML (E2, G2, II1)	4.45	0.60
D: MLL (Q2, I1, II1)	4.41	0.73
F: LLL (D1, G1, O1)	4.59	0.67

Rasch Analysis

The person-item variable map from the Rasch analysis is presented below in Figure 4.1. Please note that scenario P is included in the Rasch analysis for the pre-pilot to identify the location of this scenario in relation to the other seven scenarios, but scenario P will not be included in the full administration analysis because this is a practice item. The people who responded to the pre-pilot are plotted on the left-hand side of the middle line and the items are plotted on the right-hand side of the middle line. The plotted scenarios in Figure 4.1 generally follow the hypothesized structure, with the exception of scenarios B, D, and P (which were also noted as potentially problematic based on their descriptive statistics). In Figure 4.1, scenario D (MLL) is plotted above scenario B (MML), indicating that scenario D was slightly more difficult for a respondent to identify themselves as a higher skill level than the teacher in the scenario than in scenario B, although B was hypothesized to be more difficult to respond to than D. Additionally, scenario P (MMM) is plotted at the bottom of the item scale with scenario F (LLL). Scenario P should be more difficult for a respondent to identify themselves as a higher skill level than the teacher in scenario F, but it does not appear that way in the variable map. As noted in the discussion of the descriptive statistics, this may be due to a start-up effect where respondents are learning how to interact with the scenario-style items when responding to scenario P, and this possibility provides evidence for the importance of retaining this practice item in the scale.

Additionally, the item estimates tend to be below the person estimates in the variable map (the mean for item estimates is denoted by the M on the right side of the line and the mean for the person estimates is denoted by the M on the left side of the line). This makes it difficult, based on pre-pilot data, to provide a detailed text description of a person at the very high end of the scale (for example, person 17) because there are no scenarios hard enough on the construct for that person.

Based on this analysis, scenarios B and D need to be revised so that scenario B is more difficult for a respondent to rate themselves as a higher skillset than the teacher in that scenario than for scenario D. Scenarios B and D are more closely clustered together than the other scenarios as well and would benefit from revisions to separate them. Additionally, scenario P needs to be revised to be more difficult for a respondent to rate themselves as a higher skillset than the teacher in the scenario, in an effort to move its position up on the scale (although as noted, the placement of scenario P at the bottom of the item scale may be due to a start-up effect, in which case any revisions may not help move this scenario).

Figure 4.1

Person-item variable map from the pre-pilot

```
MEASURE         PERSON - MAP - ITEM
                    <more>|<rare>
    7                     +
                      17  |
                          |
                         T|
    6                     +
                          |
                          |
                          |
    5         18  4   7   +
                         S|
                      1   |T C
                          |
    4                     +
                  14  6   |
                          |
      12  13  15  3   5  M|
    3             11  8   +
                          |
                          |
                      22  |S G
    2                     +
                  10  19  |
              16  21  9  S|
                      20  |
    1                     +  A
                          |
                      2   |
                          |
    0                    T+M
                          |
                          |
                          |  E
   -1                     +
                          |  D
                          |  B
                          |
   -2                     +  F      P
                    <less>|<freq>
```

The person separation from the pre-pilot was 1.73 with a reliability of 0.75. The person separation value can be used to calculate the number of statistically distinct strata in the data, which can be compared to the number of hypothesized strata (Wright &

Masters, 1982); for the construct of using data to inform classroom instruction, the number of hypothesized strata is three (high, medium, and low). The formula in Equation 4.1 is used to calculate the number of statistically distinct strata in the data, where H_P is the number of statistically distinct strata in the data and G_P is the person separation value.

$$H_P = (4G_P + 1)/3 \qquad \text{Eq. 4.1}$$

For the pre-pilot, the number of statistically distinct strata in the data is equal to 2.64 using Equation 4.1 and the hypothesized number of strata is 3; these two values are relatively similar. However, the pre-pilot was a very small sample and a larger sample in the pilot may provide better person separation, increasing the number of statistically distinct strata in the data. The item separation from the pre-pilot was 5.07 (reliability of 0.96), indicating that the items likely differentiate between high and low levels of skillsets of using data to inform classroom instruction.

Figure 4.2 displays the Andrich thresholds for the pre-pilot data. In this variable map, people are still represented on the left-hand side of the line, and the Andrich thresholds for each item are represented on the right-hand side of the line. The Andrich threshold indicates the place where a respondent has a 50% probability of choosing the next highest response option for the item. The Andrich thresholds for each item are shown by the item label and the value: for example, G.4 indicates the place where a respondent has a 50% probability of choosing a 3 or a 4 for scenario G. Respondents in this map are plotted directly across the center line from the response option they were likely to select on the survey. Based on this variable map, respondents were most likely to select the highest two answer options which were coded with a 4 or a 5 (Teacher X is much less skilled than me (5) or Teacher X is slightly less skilled than me (4)) than the other answer options, with the exception of scenario C (which was constructed as the most difficult scenario). Based on these results, revisions to increase the difficulty of most of the items would be useful.

Figure 4.2

Andrich thresholds from the pre-pilot

```
MEASURE  PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
         <more>|<rare> 2        3        4        5
    9          +
               |
               |                                 C.5
               |
    8          +
               |
               |
               |
    7          +
            X  |
               |                        C.4      G.5
              T|
    6          +
               |
               |
               |                                 A.5
    5     XXX  +
              S|
            X  |T
               |                        G.4
    4          +
           XX  |
               |
        XXXXX M|                                 E.5
    3      XX  +                        A.4
               |                                 D.5
                                                 B.5
               |
            X  |S                                P.5
                                                 F.5
    2          +               C.3
           XX  |
          XXX S|
            X  |                        E.4
    1          +
               |                        D.4
            X  |      C.2               B.4
               |
    0         T+M              G.3      P.4
                                        F.4
               |
               |
               |
   -1          +
               |
               |      G.2      A.3
               |
   -2          +
               |S
               |
               |      A.2
   -3          +
               |               E.3
               |
               |               D.3
                               B.3
   -4          +
               |
               |T              P.3
                               F.3
               |      E.2
   -5          +
               |      D.2
                      B.2
               |
               |      P.2
                      F.2
   -6          +
         <less>|<freq> 1        2        3        4
```

The rating scale category structure is displayed in Table 4.2. The observed average of person estimates is ordered as intended as shown in Table 4.2, with the average increasing monotonically, as expected, as the category labels increase. The Andrich thresholds also increase monotonically as the categories increase as shown in Table 4.2, as intended. Response category one has an INFIT statistic of 1.74 and an OUTFIT statistic of 2.20, indicating that there is some misfit for this response category. When examining the misfitting responses for response category one, the two people who responded with a value of one (Teacher X is much more skilled than me) were both expected to respond with a higher response category. One of these people responded with a value of one to the last scenario (scenario G) and the other person responded with a value of one to the fourth scenario (scenario C). Scenarios G and C were hypothesized to be the most difficult for respondents to rate themselves as a higher skill level than the teacher in the scenario, and these misfitting responses could be related to the difficulty of these scenarios; however, this is based on only two people who gave a response of "1" and must be interpreted with caution.

Table 4.2

Observed averages of person estimates and Andrich thresholds from the pre-pilot

Response label	Response frequency	Observed Average	Infit	Outfit	Andrich threshold
1 (Teacher X is much more skilled than me)	2	-2.05	1.74	2.20	N/A
2 (Teacher X is slightly more skilled than me)	8	-1.53	1.06	0.97	-3.82
3 (Teacher X is just like me)	46	0.60	1.09	1.00	-2.35
4 (Teacher X is slightly less skilled than me)	54	3.39	0.81	0.69	2.01
5 (Teacher X is much less skilled than me)	66	5.42	1.01	1.07	4.17

The Category Characteristic Curves (CCCs) in Figure 4.3 provide evidence of a typical rating scale structure for all of the scenario items on the survey. The categories are in the correct ordered progression. The probability of responding with a 5 (Teacher X is much less skilled than me) is relatively high in the CCCs and provides additional evidence that some items may need to be made more difficult. In addition, there are many people responding with a 3, which is “Teacher X is just like me”. Making some of the scenarios more difficult to respond to at a high level may reduce the probability of people responding with a 3.

Figure 4.3

Category Characteristic Curves (CCCs) from the pre-pilot

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----------+-----------+-----------+-----------+-----------+-
R  1.0 +                                                        55555+
O      |                                                   55555     |
B      |1                                                0 55          |
A      | 11                                            55            |
B   .8 +   1                 333333                   5              +
I      |    1               3      3                 5               |
L      |     1             3        3               5                |
I      |      1           3          3             5                 |
T   .6 +      1          3            3    444    5                  +
Y      |       1        3              3  4   44  5                  |
    .5 +        1 2222 3               3 4      45                   +
O      |         *    *2                *       54                   |
F   .4 +        21    3 2              4 3     5  4                  +
       |       2  1  3   2            4   3    5   4                 |
R      |     22    13     2          4     3  5     4                |
E      |    2      13      2        4      3 5       4               |
S   .2 +   2       31       2      4        *         4              +
P      | 22       3  1       22  44        5 33        44            |
O      |2       33    11       **        55    3         44          |
N      |     333        1114444  222  555       333        44444     |
S   .0 +*************************************************************+
E      -+-----------+-----------+-----------+-----------+-----------+-
       -6          -3           0           3           6           9
        PERSON [MINUS] ITEM MEASURE
```

Table 4.3 displays misfit results for the items on the pre-pilot. The INFIT statistics (information-weighted fit statistics) can provide an indication that an item is problematic across all people on the scale. Values over 1.4 for INFIT indicate a problematic item for all people. In this case, scenario G has an INFIT value of 1.65, which provides an indication that it is problematic (some people gave either higher or lower responses than expected). The OUTFIT statistics (unweighted fit statistics) can indicate items that have at least one highly unexpected response across all respondents where the value is larger than 1.4. Again, scenario G has an OUTFIT value over 1.4 (the value is 1.74). Looking at the residuals for respondents to scenario G, respondents generally selected a higher response category than they were expected to for this scenario (i.e., they were selecting responses indicating that they had higher levels of the skills than the teacher in the scenario than they were expected to). This could be because it was the last scenario on the survey and respondents were experiencing survey fatigue, or it could indicate that the scenario needs to be reworded to make it more difficult to respond to than its current form.

Table 4.3

Fit statistics for the pre-pilot

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic (INFIT)		Unweighted Fit Statistic (OUTFIT)	
		MNSQ	ZSTD	MNSQ	ZSTD
C	4.44 (0.37)	1.16	0.60	1.00	0.13
G	2.32 (0.40)	1.65	1.72	1.74	1.80
A	0.96 (0.38)	0.52	-1.88	0.49	-1.84
E	-0.86 (0.38)	0.83	-0.53	0.76	-0.52
D	-1.32 (0.40)	1.04	0.22	0.86	-0.15
B	-1.49 (0.41)	1.15	0.57	1.32	0.74
P	-2.03 (0.44)	0.99	0.07	0.69	-0.31
F	-2.03 (0.44)	0.89	-0.23	0.65	-0.39

Based on the analysis of the pre-pilot data, scenarios B and D need to be revised so that scenario B is more difficult to respond to than scenario D. In addition, revisions to increase the difficulty of some of the items would be useful to reduce the probability of respondents selecting the highest response option and/or the middle response option. Scenario G needs to be made more difficult to address the misfitting responses. Finally, scenario P should be made slightly more difficult to resolve its placement in the continuum of scenarios. These revisions to the scenarios are displayed in Table 4.4 as the pilot scenarios, which were administered in the pilot administration. The bolded text in the pilot scenario column in Table 4.4 indicates the specific revisions that were made to the scenarios for the pilot administration.

Table 4.4

Pre-pilot scenarios and revised pilot scenarios

Scenario Label (Total score)	Pre-Pilot Scenario	Pilot Scenario	Reason for revision
C (9)	Teacher C is successful generating and writing clearly defined questions that clarify their thoughts on student learning. They do not need support understanding what data can help answer their questions or identifying different techniques to display their data. This teacher can independently gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.	Teacher C is successful generating and writing clearly defined questions that clarify their thoughts on student learning. They do not need support understanding what data can help answer their questions or identifying different techniques to display their data. This teacher can independently gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.	No revisions.
G (8)	Teacher G is able to think critically about data that can help answer their questions about student learning and successfully make observations about their data while comparing multiple data sources. They can also independently interpret meaning from their data analysis to develop student learning goals and consider all potential consequences of these goals. Teacher G may require some support to both identify interventions to help attain these goals and identify the data required to document progress towards these goals.	Teacher G **can** critically **examine** data **to** help answer their questions about student learning, **while independently** comparing multiple data sources **to make observations about their data**. **Interpreting** meaning from their data analysis to develop student learning goals **is a strength of Teacher G's** and **they** consider all potential consequences of these goals. Teacher G may require some support to both identify interventions to help attain these goals and identify the data required to document progress towards these goals.	Scenario G produced misfitting responses where many respondents provided higher than expected responses. It has been revised to be harder to endorse, specifically in facets E and G by changing the words in bold in an effort to make it more difficult for people to rate their own skillset as higher than Teacher G.
A (7)	Teacher A can successfully identify interventions to meet their student	Teacher A can **independently** identify interventions to meet their student	A was revised to be slightly more difficult based on the

	learning goals and plan to assess progress towards these goals. However, they sometimes struggle to implement their chosen intervention with fidelity and to collect data about student progress towards their goals during the intervention. Teacher A benefits from support with examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.	learning goals, **while also** plan**ning** to assess progress towards these goals. They sometimes **have trouble** implement**ing** their chosen intervention **as intended,** and **sometimes have difficulty** collect**ing** data about student progress towards their goals during the intervention. Teacher A **finds support helpful when** examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.	Andrich thresholds. There were no respondents on the scale at the thresholds for responses 1 to 2 or 2 to 3, so the revisions are intended to make it slightly more difficult to respond in the higher scoring response categories. In addition, the word fidelity was changed to "as intended" so it is easier to interpret.
E (6)	Teacher E occasionally needs help distilling their thoughts and wonderings about student learning into clearly defined questions. They often need support understanding what data can help answer these questions and organizing their data clearly once they have identified it. This teacher occasionally struggles with analyzing the implementation of the intervention they have chosen for fidelity and identifying their next steps once the intervention is complete.	Teacher E **sometimes benefits from** help **to** distill their thoughts and wonderings about student learning into clearly defined questions. **Once they have defined their questions, this teacher occasionally seeks support to** understand what data can help answer these questions and to organize this data clearly. This teacher may struggle **to** analyze how well the implementation of their intervention **has gone** and **is unsure of** their next steps once the intervention is complete.	E was revised to be slightly more difficult to respond to in an effort to spread out the items at the bottom of the scale. This was done by changing the words in bold in an effort to make it slightly more difficult for respondents to rate their own skillset as higher than Teacher E. In addition, the word fidelity was removed based on feedback received during the pre-pilot to make it easier to interpret.
B (5)	Teacher B may need some support thinking critically about and analyzing data that helps answer their questions about student learning. Once they complete their data analysis focused on these questions, they sometimes have trouble extracting meaning from this analysis and often do not seek the support of others to validate their	Teacher B **sometimes struggles to** think critically about and analyze data **to** help answer their questions about student learning. Once they complete their data analysis focused on these questions, they **may look for support to** extract meaning from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher	B was revised to be more difficult to respond to than D by changing the words in bold so it is slightly more difficult for respondents to rate their skillset as higher than Teacher B.

	interpretations. When they choose an intervention for student learning, Teacher B struggles to both plan for student progress towards their student learning goals and to share progress with colleagues during their implementation.	B **has difficulty** plan**ning** for student progress towards their student learning goals and **often holds back on** shar**ing** progress with colleagues during their implementation.	
D (4)	Teacher D sometimes has difficulty narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they require support to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D often holds back on sharing progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.	Teacher D sometimes **needs help** narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they **need the** support **of others** to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D **rarely** shar**es** progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.	D was revised to be slightly easier to respond to than it was in the pre-pilot by changing the words in bold to make it slightly easier for respondents to rate their skillset as higher than Teacher D.
F (3)	Teacher F struggles to understand what data can answer their questions about student learning and to organize that data clearly. This teacher has trouble interpreting results from their data analysis related to their questions about student learning and struggles to write clear and aligned goals for student learning based on their analysis. Once they have implemented a student	Teacher F struggles to understand what data can answer their questions about student learning and to organize that data clearly. This teacher has trouble interpreting results from their data analysis related to their questions about student learning and struggles to write clear and aligned goals for student learning based on their analysis. Once they have implemented a student learning	No revisions.

	learning intervention, they need support to examine student outcome data collected from this intervention and to identify their next steps.	intervention, they need support to examine student outcome data collected from this intervention and to identify their next steps.	
P (6)	Teacher P sometimes has difficulty pinpointing and writing clearly defined questions that describe their thoughts about student learning. They may need some support systematically gathering data related to these questions and comparing multiple data sources. This teacher sometimes has trouble identifying interventions to meet their goals related to student learning or planning to assess progress towards these goals on their own and looks to others for support.	Teacher P **occasionally** has difficulty pinpointing and writing clearly defined questions that describe their thoughts about student learning. **Once they have defined their questions, they** may need some support **to** systematically gather data related to these questions and compare multiple data sources. This teacher **can sometimes benefit from help to** identify interventions to meet their goals related to student learning **and to** plan to assess progress towards **their** goals.	P was revised to be slightly more difficult to respond to by changing the words in bold, although as the practice item, the issues identified for this item in the pre-pilot may be related to a start-up effect with scenario-type items and changing the words in bold may not result in the hypothesized changes.

Pilot Administration Results

Descriptive Statistics

Responses were received from 169 workers on Amazon Mechanical Turk (MTurk) who are employed in the education industry (please see Appendix D for the pilot survey). One of these respondents did not answer any items beyond scenario E, and upon examination of this individual's response in Qualtrics, this respondent did not view any items on the survey after scenario E. This respondent must have abandoned the survey and, given that they did not respond to any social desirability items, was removed from the analysis, bringing the pilot total dataset to 168 respondents.

None of the scenario items had missing data in the pilot dataset. Six of the thirteen social desirability scale items had between 1-3 missing responses (a percentage of 0.6% to 1.8% missing data for these individual items). The question on the number of years of experience for the respondents had 1 missing response. Across individuals, nine respondents were missing data on only 1 item, for a missing percentage of 0.6% for each individual. Overall, 5.4% of respondents had missing data; however, all of this missing data was for the social desirability scale and no individual item on the social desirability scale had more than 1.8% missing data. Because the social desirability scale was included in the pilot administration only to assess the extent to which respondents provided more socially desirable answers to the scenario items, and the fact that the overall percentage of missing data by individuals and by item is relatively low (close to 5%), listwise deletion of respondents with missing data was utilized in the following analyses.

Demographic information was also collected from the pilot administration and is displayed in Table 4.5. Fifty-two percent (52%) of respondents were female and 48% were male. Sixty-two percent (62%) were White, 13% were Asian, 12% were Black/African American, 8% were Hispanic/Latinx, 2% were more than one race, 2% were Native American/Alaskan Native, and 1% preferred not to answer this item. When asked how many years of teaching experience they had, 8% had 0-3 years, 29% had 3-5 years, 37% had 5-10 years, 14% had 10-15 years, and 11% had more than 15 years. Please note that the overlapping years in the response categories for this question were revised in the full administration so that they do not overlap. When asked for their primary teaching role, 55% responded that they were secondary general content teachers (English/ELA, History/Social Studies, Math, or Science), 13% were elementary specialist teachers, 11% were secondary specialist teachers, 10% were administrators, 7% were elementary homeroom teachers, and 5% were in some other kind of role (including English Language Learner (ELL) teachers, computer teachers, and university-level educators).

Table 4.5

Demographics of respondents from the pilot

Demographic group	Number of respondents	Percentage of respondents
Gender		
Female	87	51.8%
Male	81	48.2%
Race/Ethnicity		
African American/Black	20	11.9%
Asian	21	12.5%
Hispanic/Latinx	14	8.3%
More than one race	4	2.4%
Native American/Alaskan Native	3	1.8%
White	104	61.9%
Prefer not to answer	2	1.2%
Years of teaching experience		
0-3 years	14	8.4%
3-5 years	49	29.3%
5-10 years	62	37.1%
10-15 years	24	14.4%
More than 15 years	18	10.8%
Primary teaching role		
Administrator	17	10.1%
Elementary homeroom teacher	11	6.5%
Elementary specialist teacher	21	12.5%
Secondary general content teacher	93	55.3%
Secondary specialist teacher	18	10.7%
Other	8	4.8%

Answer choices for the scenario items were coded as follows (the same coding structure as the pre-pilot administration):

- Teacher X is much less skilled than me: 5
- Teacher X is slightly less skilled than me: 4
- Teacher X is just like me: 3
- Teacher X is slightly more skilled than me: 2
- Teacher X is much more skilled than me: 1

As in the pre-pilot, scenario items with a higher mean value were "easier" to respond to, meaning that it was easier for a respondent to identify themselves as a higher skill level than the teacher in the scenario. Scenario items with a lower mean value were "harder" to respond to, meaning that it was harder for a respondent to identify themselves as a higher skill level than the teacher in the scenario. For example, scenario F was hypothesized to be the easiest scenario to respond to, meaning it should be the easiest scenario for a respondent to rate themselves as a higher skill level than the teacher in the scenario. This is denoted by the levels of the facets in scenario F (all three of which were low, or LLL). Scenario C was hypothesized to be the hardest scenario to respond to, meaning it should be the hardest scenario for a respondent to rate themselves as a higher skill level than the teacher in the scenario. This is also denoted by the levels of the facets in scenario C (all three of which were high, or HHH).

Table 4.6 displays descriptive statistics for the scenario items from the pilot administration. The items are ordered by their hypothesized hierarchical continuum in this table, with the hypothesized hardest item first (C) and the hypothesized easiest item last (F). The items are labeled by their teacher name first (a letter), followed by the levels of the facets in the scenario (where H equals high, M equals medium, and L equals low) and then the actual facet and level combinations in the scenario. Although the hypothesized continuum was evaluated using a Rasch model, as in the pre-pilot, looking at the descriptive statistics in this way can provide some evidence to evaluate the hypothesized structure to provide proof of concept for the construct of using data to inform classroom instruction.

The descriptive statistics for the scenario items provide some evidence of proof of concept of the hypothesized scenario structure. Scenarios B and P have similar item means, although B (MML) should be slightly easier to respond to than P (MMM), meaning B should have a higher mean value than P. Scenario P, however, is the practice scenario and this may be some evidence of a start-up effect, as respondents may be learning how to interact with the scenario-style item with their responses to scenario P. Scenario F (LLL) has a similar mean value to scenario E (MMM), although scenario F should be the easiest to respond to, meaning it should have the highest mean value. Scenarios B (MML) and D (MLL) had very similar means in the pre-pilot and were revised in an attempt to make scenario B more difficult to respond to than scenario D. Based on the descriptive statistics, B does appear to be more difficult to respond to than scenario D.

Table 4.6

Descriptive statistics for the scenario items from the pilot ordered by the hypothesized scenario structure

Item	Mean	Standard Deviation	Range
C: HHH (Q3, D3, E3)	2.72	1.18	1-5
G: HHM (E3, G3, I2)	3.07	1.03	1-5
A: HMM (I3, II2, O2)	3.36	1.11	1-5
E: MMM (Q2, D2, O2)	3.62	1.17	1-5
P: MMM (Q2, E2, I2)	3.68	1.18	1-5
B: MML (E2, G2, II1)	3.67	1.15	1-5
D: MLL (Q2, I1, II1)	3.79	1.16	1-5
F: LLL (D1, G1, O1)	3.64	1.30	1-5

Note. The following abbreviations in parentheses refer to the specific facets in each scenario. Q: Ask questions. D: Identify data. E: Examine data. G: Interpret data to set goals. I: Identify intervention. II: Implement intervention. O: Examine outcomes. The number after each facet label indicates the level of the facet in the scenario (1=Low; 2=Medium; 3=High).

Table 4.7 displays descriptive statistics for the social desirability scale items from the pilot administration. The social desirability scale items come from the Marlowe-Crowne social desirability scale, short form C, abbreviated as M-C Form C (Reynolds, 1982). These items are presented as True/False items to respondents. The response (True or False) that is scored with a one for the item, also known as the keyed response, is identified in parentheses next to the item in Table 4.7, where F means False and T means True. This keyed response in the table indicates the socially desirable answer to the question; the other response option was coded with a zero for that item. The majority of the items have a mean value around 0.40, indicating that slightly under half of the respondents chose the socially desirable response for the item. Two items have higher mean values of 0.80, indicating that 80% of the respondents chose the socially desirable response for these items (No matter who I'm talking to, I'm always a good listener; I'm always willing to admit it when I make a mistake). The overall scale average is also shown in Table 4.7.

Table 4.7

Descriptive statistics for the social desirability scale (M-C Form C) items from the pilot (keyed response in parentheses)

Item	N	Mean	Standard Deviation	Range
It is sometimes hard for me to go on with my work if I am not encouraged. (F)	168	0.43	0.50	0-1
I sometimes feel resentful when I don't get my way. (F)	167	0.40	0.49	0-1
On a few occasions, I have given up doing something because I thought too little of my ability. (F)	168	0.43	0.50	0-1
There have been times when I felt like rebelling against people in authority even though I knew they were right. (F)	168	0.47	0.50	0-1
No matter who I'm talking to, I'm always a good listener. (T)	167	0.80	0.40	0-1
There have been occasions when I took advantage of someone. (F)	168	0.43	0.50	0-1
I'm always willing to admit it when I make a mistake. (T)	166	0.80	0.40	0-1
I sometimes try to get even rather than forgive and forget. (F)	168	0.39	0.49	0-1
I am always courteous, even to people who are disagreeable. (T)	168	0.72	0.45	0-1
I have never been irked when people expressed ideas very different from my own. (T)	168	0.60	0.49	0-1
There have been times when I was quite jealous of the good fortune of others. (F)	165	0.41	0.49	0-1
I am sometimes irritated by people who ask favors of me. (F)	167	0.42	0.49	0-1
I have never deliberately said something that hurt someone's feelings. (T)	167	0.62	0.49	0-1
Total scale	159	6.91	2.88	0-13

The correlations between each scenario item and the M-C Form C scale responses from the pilot administration are presented in Table 4.8. As displayed, the correlations between scenario C and the M-C Form C scale and scenario G and the M-C Form C scale are statistically significantly different from zero ($p < .05$). These correlations indicate that there may be some social desirability bias in the responses to these two scenarios. This means that respondents may have been responding to these scenarios in what they believe is the socially desirable way, rather than the way that reflects their actual experience. These two scenarios (C (HHH) and G (HHM)) were written to be the most difficult for respondents to respond to and it is possible that respondents could identify that these two scenarios had teachers with the highest skillsets in the set of scenarios. Scenarios C and G could be revised slightly to make it less obvious that these teachers have the highest levels of skillsets in the set of scenarios in an attempt to resolve this issue.

Table 4.8

Correlations of scenario items and the M-C Form C scale from the pilot

	Scenario							
	P	A	B	C	D	E	F	G
M-C Form C	0.08	-0.10	0.10	-0.17*	0.09	0.03	0.14	-0.16*

*$p < .05$.

Rasch Analysis

The person-item variable map from the Rasch analysis is presented below in Figure 4.4. Please note that scenario P is included in the Rasch analysis for the pilot administration to identify the location of this scenario in relation to the other seven scenarios, but scenario P will not be included in the full administration analysis because this is a practice item. The people who responded to the pilot are plotted on the left-hand side of the middle line and the items are plotted on the right-hand side of the middle line. As a reminder, the hypothesized continuum of scenarios (displayed in Table 4.6) is as follows:

C
G
A
E/P
B
D
F

The plotted scenarios in Figure 4.4 generally follow the hypothesized structure, with the exception of scenarios P and F. In Figure 4.4, scenario F (LLL) is plotted with scenario E (MMM), indicating that scenario F has a similar difficulty level to scenario E. Scenario F, however, is meant to be the easiest scenario to respond to and should be located at the bottom of the scale. Additionally, scenario P (MMM) is plotted with scenario B (MML), although P should be more difficult for respondents than B. This finding again may be due to a start-up effect and provides evidence for the importance of retaining this practice item in the scale.

In the pre-pilot administration, scenario D (MLL) was plotted above B (MML) and scenario D was revised to make it easier for a respondent to rate themselves as a higher skillset than the teacher in the scenario. They are now ordered as hypothesized, with scenario D below scenario B.

As shown in Figure 4.4, the item estimates are slightly below the person estimates in the variable map (again, the mean for item estimates is denoted by the M on the right side of the line and the mean for the person estimates is denoted by the M on the left side of the line). This makes it difficult, based on pilot data, to describe a person at the very high end of the scale or very low end of the scale because there are no scenarios as high or as low as the person estimates at these ends of the scale.

Based on these results, scenario F needs to be revised so that it is easier for respondents to rate their skill level as higher than the teacher in the scenario. Scenarios B, D, E, and F are more closely clustered together than the other scenarios as well and would benefit from revisions to separate them further. Additionally, scenario P needs to be revised to be more difficult for respondents to rate their skill level as higher than the teacher in the scenario in an effort to move its position on the scale (although as noted, the placement of scenario P near the bottom of the item scale may be due to a start-up effect). Details on these revisions are described later in this section in Table 4.13.

Figure 4.4

Person-item variable map from the pilot

```
MEASURE                                      PERSON - MAP - ITEM
                                                 <more>|<rare>
    4                                           14 16  +
                                                       |
                                                       |
                                                   38  |
                                                       |
                                                       |
                                                       |
                                                       |
                                                       |
    3                                                  +
                                             14 32 76  |
                                                       |
                                                       |
                                                      T|
                                             14 46 91  |
                                                       |
                                                       |
                                             11 16 28  |
    2                                                  +
                                 12 15 16 16 29 40 53  |
                                                       |
                                             14 26 93  |
                                                      S|
      10 11 13 14 15 15 21 33 34 5  59 65 71 74 78 92  |
                                                       |
                                    12 13 15 16 4  57  |
            10 11 11 11 13 14 14 37 45 61 82 85 90 94  |
    1                                                  +
            10 11 12 13 16 27 52 6  64 68 75 88 9  98  |T C
                           10 12 13 15 16 19 20 51 89  |
            1  11 15 15 17 18 24 35 36 41 55 58 96 97  |
                                                      M|
                     10 10 11 14 14 14 16 22 44 70 73  |S G
                                 13 13 13 13 15 16 99  |
                     10 10 10 12 12 12 2  23 31 62 63  |
                                 10 13 3  43 66 69 95  |  A
    0                                                  +M
                                    10 11 15 25 56 77  |
                                          14 54 79 87  |  E     F
                                                72 8   |  B     P
                                                      S|S D
                                          12 12 60 80  |
                                          39 42 49 86  |
                                             12 67 7   |
                                                       |T
   -1                                           15 16  +
                                                       |
                                    11 15 30 50 81 83  |
                                                       |
                                                   12 T|
                                                       |
                                                       |
                                                       |
                                                13 84  |
   -2                                                  +
                                                       |
                                                   48  |
                                                       |
                                                       |
                                                       |
                                                   47  |
                                                       |
                                                       |
   -3                                              11  +
                                                 <less>|<freq>
```

The person separation from the pilot was 2.01 with a reliability of 0.80. As described for the pre-pilot, the person separation value can be used to calculate the number of statistically distinct strata in the data, which can be compared to the number of hypothesized strata in the data (Wright & Masters, 1982), which in this case is three (high, medium, and low). The formula in Equation 4.1 was used to calculate the number of statistically distinct strata in the data. Based on this formula, the number of statistically distinct strata in the data is 3.01, compared to 3 hypothesized strata. This is an improvement from the pre-pilot and matches the hypothesized strata. The item separation from the pilot was 4.74 (reliability of 0.96), indicating that the items likely differentiate between high and low levels of skillsets of using data to inform classroom instruction.

Figure 4.5 displays the Andrich thresholds for the pilot data. In this variable map, people are still represented on the left-hand side of the line, and the Andrich thresholds for each item are represented on the right-hand side of the line. The Andrich threshold indicates the place where a respondent has a 50% probability of choosing the next highest response option for the item. The Andrich thresholds for each item are shown by the item label and the value: for example, G.4 indicates the place where a respondent has a 50% probability of choosing a 3 or a 4 for scenario G. Respondents in this map are plotted directly across the center line from the response option they were likely to select on the survey. Based on this variable map, there is more variation in responses on the pilot compared to the pre-pilot (where respondents were most likely to select the highest two answer options, except for scenario C). For the pilot, respondents were less likely to select the lowest response option (Teacher X is much more skilled than me). Based on Figure 4.5, there is a clear pattern of the expected responses for people at different ability levels. There is some overlap between answer options 3 and 4 (for example, see A.3, G.3, and C.3 below, which overlap with some of the thresholds for the response values of 4 for some scenarios).

Figure 4.5

Andrich thresholds from the pilot

```
MEASURE    PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
               <more>|<rare> 2        3        4        5
    4             #  +
                     |
                     |
                  .  |
                     |
                     |
                     |
                     |
                     |
    3                +
                 .#  |
                     |
                     |
                    T|
                 .#  |                                 C.5
                     |
                     |
                 .#  |
    2                +                                 G.5
               .###  |
                     |
                 .#  |                                 A.5
                    S|
           ########  |
                     |                        C.4      E.5
                                                       F.5
                ###  |                                 B.5
                                                       P.5
            #######  |                                 D.5
    1                +
            #######  |T                       G.4
              .####  |               C.3
            #######  |
                    M|                        A.4
             .#####  |S
               .###  |               G.3
             .#####  |                        E.4
                                              F.4
               .###  |                        B.4
                                              P.4
    0                +M              A.3      D.4
                ###  |
                 ##  |
                  #  |               E.3
                                     F.3
                    S|S              B.3
                                     P.3
                 ##  |               D.3
                 ##  |
                 .#  |
                     |T     C.2
   -1             #  +
                     |
                ###  |
                     |      G.2
                  . T|
                     |
                     |
                     |      A.2
                  #  |
   -2                +
                     |      E.2
                            F.2
                            B.2
                            P.2
                  .  |
                     |      D.2
                     |
                     |
                  .  |
                     |
                     |
   -3             .  +
               <less>|<freq> 1        2        3        4
EACH "#" IS 2: EACH "." IS 1
```

Table 4.9 displays statistics related to the rating scale category structure. The observed average of person estimates is ordered as intended, with the average increasing monotonically, as expected, as the category labels increase. The Andrich thresholds also increase monotonically as the categories increase, as expected. There is again some slight misfit associated with response category 1 as there was in the pre-pilot (Teacher X is much more skilled than me; INFIT value of 1.29 and OUTFIT value of 1.22), indicating there were some responses of 1 from people who were expected to score higher. The misfit for category 1, though, has improved from the pre-pilot. The Category Characteristic Curves (CCCs) in Figure 4.6 provide evidence of a typical rating scale structure for all of the scenario items on the survey. The categories are in the correct ordered progression. The probability of responding with a 5 (Teacher X is much less skilled than me) is still relatively high in the CCCs from the pilot administration and provides additional evidence that some items may need to be revised to be more difficult for respondents to rate themselves as a higher skill level than the teacher in the scenario. The probability of responding with a 3 (Teacher X is just like me) is lower than in the pre-pilot and fits the rating scale category structure better than in the pre-pilot.

Table 4.9

Observed averages of person estimates and Andrich thresholds from the pilot

Response label	Response frequency	Observed Average	INFIT	OUTFIT	Andrich threshold
1 (Teacher X is much more skilled than me)	85	-0.88	1.29	1.22	N/A
2 (Teacher X is slightly more skilled than me)	250	-0.36	0.84	0.83	-1.85
3 (Teacher X is just like me)	304	0.38	0.87	0.82	-0.11
4 (Teacher X is slightly less skilled than me)	394	0.99	1.05	1.13	0.43
5 (Teacher X is much less skilled than me)	311	1.55	1.03	1.03	1.53

Figure 4.6

Category Characteristic Curves (CCCs) from the pilot

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----------+-----------+-----------+-----------+-----------+-
R  1.0 +                                                        55555+
O      |                                                 5555555     |
B      |11                                            555            |
A      |  11                                       555               |
B   .8 +    11                                    5                  +
I      |      1                                 55                   |
L      |       11                              5                     |
I      |         1                            5                      |
T   .6 +          1                          5                       +
Y      |           1                        5                        |
    .5 +            1 2222222              5                         +
O      |            2*       2        444 5                          |
F   .4 +           2  1       2     44   *44                         +
       |         22    1      3*333*    5   4                        |
R      |        2       1    3  2 4 3  5     44                      |
E      |      22         1 33   4*   3*        44                    |
S   .2 +    22           3*1   4  22 5 3         44                  +
P      |  22            3   1 4    5*   33         44                |
O      |22           333    4*1   5  22   33         4444            |
N      |         3333   4444   ***1    222  33333        4444444     |
S   .0 +***************5555555    1111111***************************+
E      -+-----------+-----------+-----------+-----------+-----------+-
       -4          -2           0           2           4           6
        PERSON [MINUS] ITEM MEASURE
```

Table 4.10 displays misfit results for the items on the pilot. As noted for the pre-pilot, the INFIT statistics can provide an indication that an item is problematic across all people on the scale. Values over 1.4 for INFIT indicate a problematic item for all people. For the pilot, none of the items have an INFIT value over 1.4, although scenarios F and C are close (with INFIT values of 1.35 and 1.32, respectively). On the pre-pilot, scenario G had an INFIT value of 1.65; on the pilot, it has an INFIT value of 1.00, indicating improvement.

The OUTFIT statistics (where the value is larger than 1.4) indicate items that have at least one highly unexpected response across all respondents. None of the scenarios have OUTFIT statistics with a value larger than 1.4, although again, scenarios F and C are close (with OUTFIT values of 1.35 and 1.33, respectively). On the pre-pilot, scenario

G had an OUTFIT value of 1.74; on the pilot, it has an OUTFIT value of 1.03, indicating improvement.

Table 4.10

Fit statistics for the pilot ordered by logit estimate

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic (INFIT)		Unweighted Fit Statistic (OUTFIT)	
		MNSQ	ZSTD	MNSQ	ZSTD
C	0.94 (0.09)	1.32	2.85	1.33	2.89
G	0.49 (0.09)	1.00	0.07	1.03	0.33
A	0.12 (0.09)	0.75	-2.61	0.74	-2.66
E	-0.22 (0.09)	1.00	0.02	0.92	-0.67
F	-0.25 (0.09)	1.35	3.06	1.35	2.90
B	-0.29 (0.09)	0.85	-1.43	0.91	-0.76
P	-0.32 (0.09)	0.75	-2.54	0.72	-2.71
D	-0.47 (0.09)	0.89	-1.02	0.87	-1.14

Looking at misfitting individuals and their residuals shows observed and expected responses and can help identify response patterns for individuals that are unexpected (please see Appendix E for the table of misfitting individuals and their residuals from the pilot). From an examination of the residuals from the pilot, the majority of the misfitting individuals had higher than expected responses to scenarios C and G, which were the two scenarios that had correlations with the social desirability scale that were statistically significantly different from zero ($p < .05$). For these individuals, it seems plausible that their responses to scenarios C and G were influenced by wanting to provide the socially desirable response instead of their lived experience response (i.e., providing a higher response than may actually reflect their experience). Misfitting individuals also had lower than expected responses to scenario F, which provides some additional evidence that scenario F should be revised to be easier to respond to with a higher response option. In addition, there were some misfitting individuals who had unexpected responses to scenarios P and A, which could be some evidence of a start-up effect, as P is the practice item and A is the first scenario item presented to respondents. Finally, there were a few unexpected responses to other scenarios, which is to be expected with a larger sample of respondents. There were no evident patterns to those other unexpected responses.

Principal Components Analysis

A principal components analysis (PCA) was run on the Rasch residuals from the pilot administration. The purpose of the PCA for the residuals is to look for evidence of an unidentified construct in the residual data, as one of the assumptions of the Rasch model is unidimensionality. As described previously, unidimensionality means that the items on the scale measure a single construct. The PCA on the Rasch residuals can be compared to a PCA on random data in a parallel analysis to assess unidimensionality. In this analysis, unlike a typical PCA, the goal is to identify zero patterns in the residual data. The residual data should be similar to randomly generated data to provide evidence that there is no unidentified construct in the residual data.

Table 4.11 displays the eigenvalues and the percentage of variance explained for each eigenvalue from the PCA for the pilot residuals and randomly generated data. Although a component with an eigenvalue slightly over 2 was extracted from the pilot residuals, which could indicate an unexplained construct in the residuals, the difference in the variance explained by the first eigenvalue for the pilot residuals and the randomly generated data is just around 10%, which is not a large difference. When the scree plots for the residuals and the randomly generated data are examined (in Figures 4.7 and 4.8, respectively), although there appears to be a break between components one and two in the residual data PCA, this break is also apparent in the scree plot of the randomly generated data. Both scree plots appear similar. Finally, Figures 4.9 and 4.10 display the component loading plots for the PCA for the residuals and the randomly generated data, respectively. The components for the residuals do not appear clustered into groups and they approximate random data (which would be plotted in a circular pattern) (Ludlow, 1983). These combined results provide some evidence that there is no unidentified construct present in the residuals.

Table 4.11

Principal Components Analysis results for the residuals from the pilot compared to random data

Pilot residuals			Random data		
Component number	Eigenvalue	% Variance explained	Component number	Eigenvalue	% Variance explained
1	2.181	27.266	1	1.347	16.838
2	1.266	15.823	2	1.139	14.241
3	1.181	14.765	3	1.106	13.822
4	1.065	13.308	4	1.054	13.176
5	.880	10.998	5	1.016	12.697
6	.743	9.283	6	.847	10.587
7	.671	8.385	7	.823	10.293
8	.014	0.172	8	.668	8.347

Figure 4.7

Scree plot for residuals from the pilot

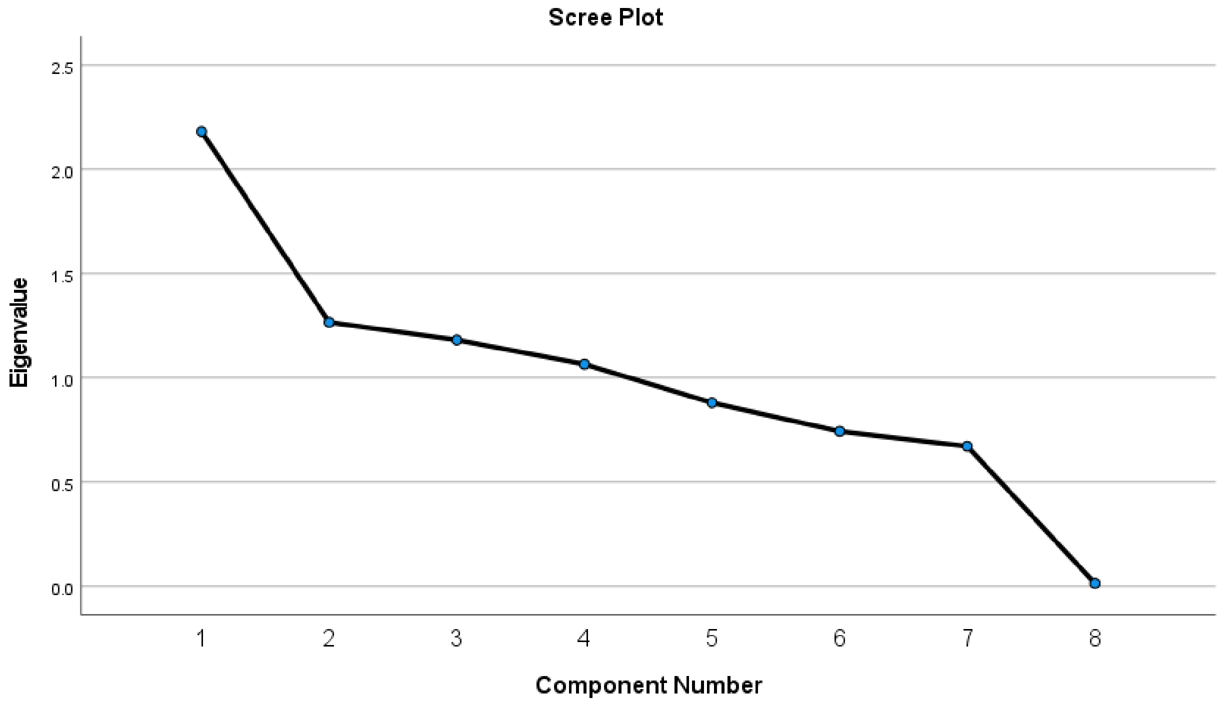


Figure 4.8

Scree plot for randomly generated data

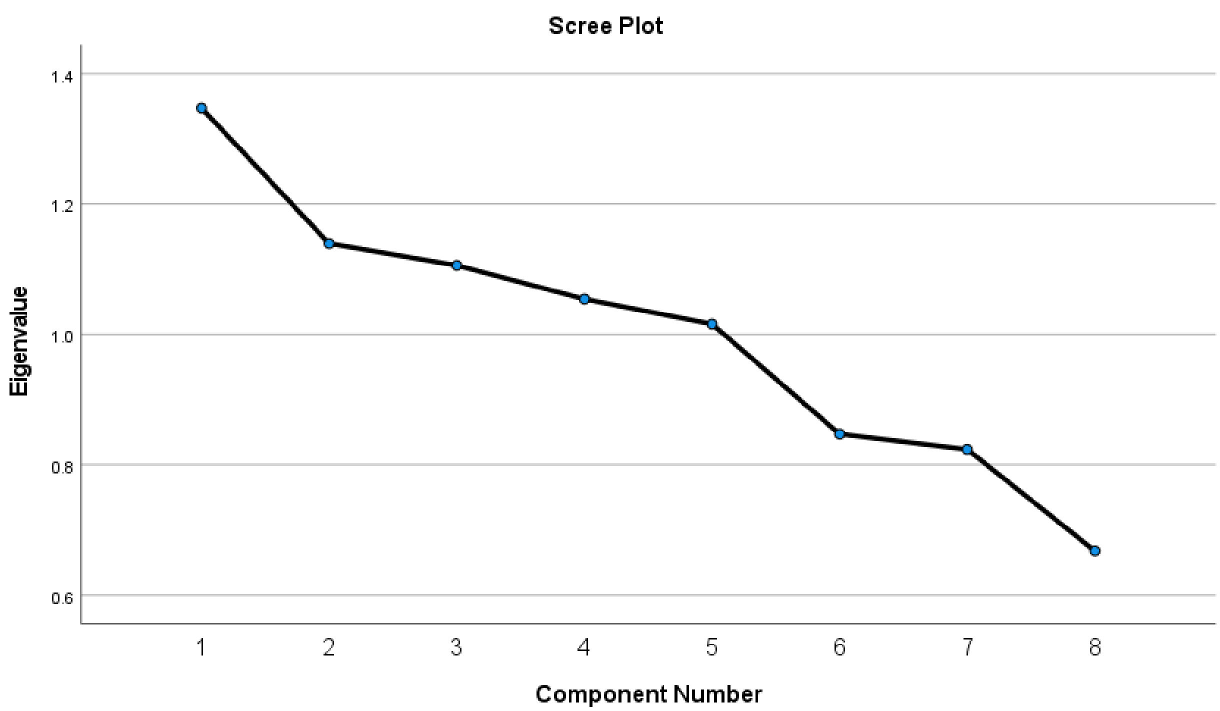


Figure 4.9

Component loading plot for residuals from the pilot

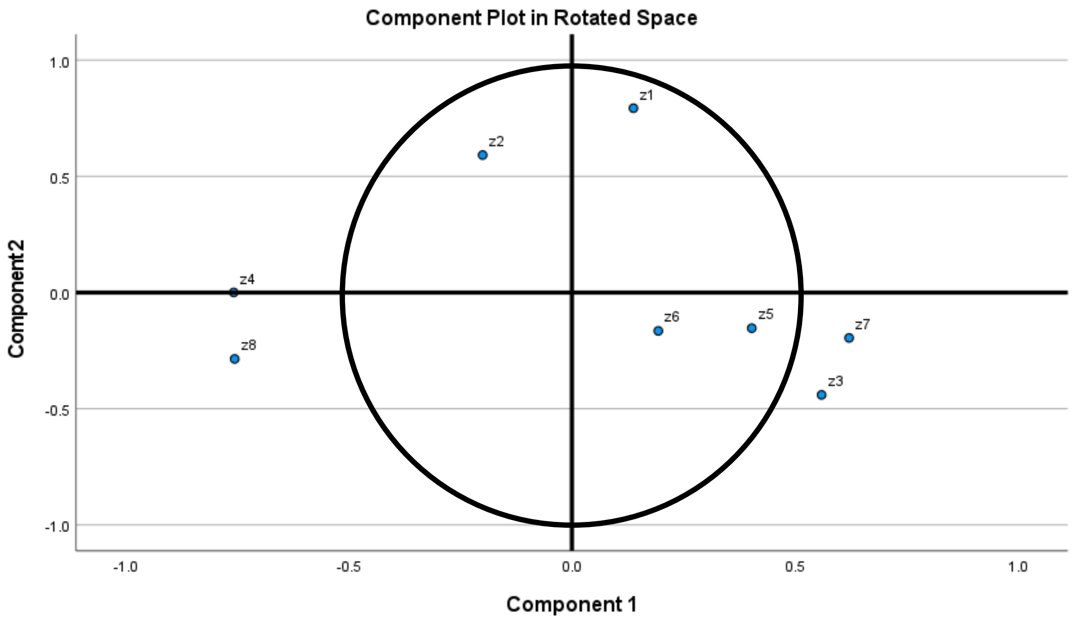
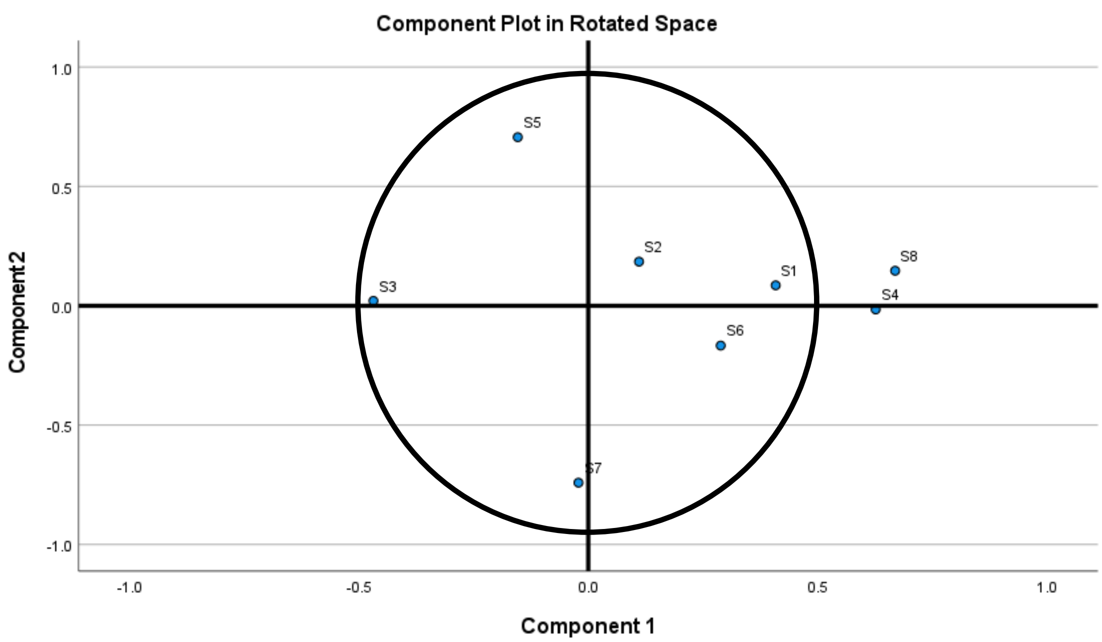


Figure 4.10

Component loading plot for randomly generated data



Based on these results from the pilot, some revisions to most items are suggested. Specifically, scenarios C and G should be revised slightly because of the social desirability bias that may be affecting responses to these two highest scenarios on the scale. Evidence of possible social desirability bias for these two scenarios was identified in both the correlations of the M-C Form C scale and these two scenarios, as well as the responses from misfitting individuals. These scenarios should be revised to make it less obvious that these teachers have the highest skillsets in the set of scenarios.

Scenario F also needs to be revised to make it easier to respond to, as it was not ordered as hypothesized in the person-item variable map. Misfitting individuals also provided lower than expected responses to scenario F, providing more evidence that revisions to make it easier should be attempted.

Scenarios B, D, E, and F are clustered together at the end of the scale. Revisions to separate these items would be useful. Scenario P could also be revised to make it more difficult to respond to, although as the practice item on the scale, it is possible that any revisions will have little effect, as people are learning how to interact with these types of scenario items.

Finally, revisions to the language utilized for scenarios that contain facet G (Interpret data to set goals) should be made for the full administration (this includes scenarios C, G, and B). Specifically, the language should be altered to include the use of hypotheses as part of this facet, as described by Hamilton et al. (2009), as the model described by Hamilton et al. (2009) was added after the pilot administration. This model focuses on the development of hypotheses about student learning to set goals as part of the model and thus needs to be incorporated into the scenarios that contain facet G.

Table 4.12 displays the revised inclusion of facets by scenario for the full administration. Part of the revisions to scenarios C and F includes changing the composition of the facets in these scenarios. This was done in an effort to make it less obvious that the teacher in scenario C has the highest skillset in the group of scenarios. Scenarios/facets in bold in Table 4.12 indicate a change in the facet composition in the scenario for the full administration from the pre-pilot and pilot administrations; scenarios/facets that are not in bold remain the same in the full administration as in the pre-pilot and pilot administrations. Table 4.13 displays the scenario items as presented on the pilot administration, along with the revisions based on this analysis for the full administration. Revisions to the scenarios for the full administration are noted in bold text in the full administration scenario column in Table 4.13.

Table 4.12

Revised inclusion of facets by scenario for the full administration

Scenario Label	Facet 1 (Q)	Facet 2 (D)	Facet 3 (E)	Facet 4 (G)	Facet 5 (I)	Facet 6 (II)	Facet 7 (O)	Total scenario score
C		**H (3)**		**H (3)**			**H (3)**	**9**
G			H (3)	H (3)	M (2)			8
A					H (3)	M (2)	M (2)	7
E	M (2)	M (2)					M (2)	6
B			M (2)	M (2)		L (1)		5
D	M (2)				L (1)	L (1)		4
F	**L (1)**	**L (1)**	**L (1)**					**3**
P	M (2)		M (2)		M (2)			6

Table 4.13

Scenarios from the pilot and full administration scenarios after pilot revisions

Scenario label (Total score)	Pilot scenario	Full administration scenario	Reason for revision
C (9)	Teacher C is successful generating and writing clearly defined questions that clarify their thoughts on student learning. They do not need support understanding what data can help answer their questions or identifying different techniques to display their data. This teacher can independently gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.	**Teacher C can identify data to answer their questions about student learning and consistently organizes that data clearly. This teacher independently interprets results from their data analysis related to their questions about student learning, while writing clear and aligned goals for student learning based on hypotheses about how to improve student learning that come from their analysis. Once they have implemented a student learning intervention, they can successfully examine student outcome data collected from this intervention and are able to identify their next steps.**	Scenario C was correlated with the M-C Form C (social desirability scale) and misfitting individuals provided higher than expected responses to it. Scenario C was composed of the first three facets (Q, A, and E), which may make it easier for respondents to provide high responses, as it is focused on writing questions, understanding data, and collecting data (the beginning of the inquiry cycle). The facets for C and F were swapped in an attempt to make C more difficult to respond to and F easier to respond to.
G (8)	Teacher G can critically examine data to help answer their questions about student learning, while independently comparing multiple data sources to make observations about their data. Interpreting meaning from their data analysis to develop student learning goals is a strength of Teacher G's and they consider all potential consequences	Teacher G **successfully compares multiple data sources while** critically examin**ing** data to help answer their questions about student learning. Interpreting meaning from their data analysis to develop student learning goals is a strength of Teacher G's, and they consider all potential consequences of these goals **before moving forward**.	Scenario G was correlated with the M-C Form C (social desirability scale) and misfitting individuals provided higher than expected responses to it. Scenario G was revised to make it slightly more difficult to respond to by altering the language so it

	of these goals. Teacher G may require some support to both identify interventions to help attain these goals and identify the data required to document progress towards these goals.	Teacher G **benefits from** some support to both identify interventions to help attain these goals and **to** identify the data required to document progress towards these goals.	sounds less positive.
A (7)	Teacher A can independently identify interventions to meet their student learning goals, while also planning to assess progress towards these goals. They sometimes have trouble implementing their chosen intervention as intended, and sometimes have difficulty collecting data about student progress towards their goals during the intervention. Teacher A finds support helpful when examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.	Teacher A can independently identify interventions to meet their student learning goals, while also planning to assess progress towards these goals. They sometimes have trouble implementing their chosen intervention as intended, and sometimes have difficulty collecting data about student progress towards their goals during the intervention. Teacher A finds support helpful when examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.	No revisions.
E (6)	Teacher E sometimes benefits from help to distill their thoughts and wonderings about student learning into clearly defined questions. Once they have defined their questions, this teacher occasionally seeks support to understand what data can help answer these questions and to organize this data clearly. This teacher may struggle to analyze how well the implementation of their intervention has gone and is unsure of their next steps once the intervention is complete.	Teacher E sometimes benefits from help to distill their thoughts and wonderings about student learning into clearly defined questions. Once they have defined their questions, this teacher occasionally seeks support to understand what data can help answer these questions and to organize this data clearly. This teacher may struggle to analyze how well the implementation of their intervention has gone and is unsure of their next steps once the intervention is complete.	No revisions.
B (5)	Teacher B sometimes struggles to think	Teacher B **occasionally has difficulty**	Revised to be slightly easier to

	critically about and analyze data to help answer their questions about student learning. Once they complete their data analysis focused on these questions, they may look for support to extract meaning from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher B has difficulty planning for student progress towards their student learning goals and often holds back on sharing progress with colleagues during their implementation.	think**ing** critically about and analyz**ing** data to help answer their questions about student learning. Once they complete their data analysis focused on these questions, they may **need** support to **generate hypotheses about how to improve student learning** from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher B has **trouble** planning for student progress towards their student learning goals and often holds back on sharing progress with colleagues during their implementation.	respond to in an effort to separate it from scenario E in the person-item variable map.
D (4)	Teacher D sometimes needs help narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they need the support of others to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D rarely shares progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.	Teacher D **occasionally** needs help narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they **require** the support of others to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D rarely shares progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.	Revised to be slightly easier to respond to in an effort to separate it from scenario B in the person-item variable map.
F (3)	Teacher F struggles to understand what data can answer their questions about student learning and to organize that data clearly. This teacher has trouble	**Teacher F struggles to generate and write clearly defined questions that clarify their thoughts on student learning. They have trouble**	The facets (but not the levels) from scenario C and F in the pilot were swapped in an attempt to make F easier to

	interpreting results from their data analysis related to their questions about student learning and struggles to write clear and aligned goals for student learning based on their analysis. Once they have implemented a student learning intervention, they need support to examine student outcome data collected from this intervention and to identify their next steps.	**understanding what data can help answer their questions or identifying different techniques to display their data. This teacher needs the support of others to gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.**	respond to.
P (6)	Teacher P occasionally has difficulty pinpointing and writing clearly defined questions that describe their thoughts about student learning. Once they have defined their questions, they may need some support to systematically gather data related to these questions and compare multiple data sources. This teacher can sometimes benefit from help to identify interventions to meet their goals related to student learning and to plan to assess progress towards their goals.	Teacher P **sometimes** has **trouble clarifying** and writing clearly defined questions that describe their thoughts about student learning. Once they have defined their questions, they **may benefit from** some support to systematically gather data related to these questions and compare multiple data sources. **When identifying interventions to meet their student learning goals, the support of others is helpful, and this teacher finds working with others** to plan to assess progress towards their goals **is useful.**	Revised to be slightly harder to respond to.

Full Administration Results

Descriptive Statistics

Responses were received from 287 teachers in a public school district in a medium-sized city in Massachusetts (please see Appendix H for the full administration survey). Three of these teachers did not provide consent and were removed from the dataset. Another 53 teachers provided consent but did not answer any of the questions on the survey and were also removed from the dataset. An additional 15 teachers responded only to the practice item, which was presented first, and then abandoned the survey. Given that these teachers did not respond to any additional items on the survey, they were removed from the dataset. Eight teachers responded only to the practice item and scenario A, and then abandoned the survey and five teachers responded only to the practice item, scenario A, and scenario B, and then abandoned the survey. Given that these teachers only viewed and responded to the practice item and one or two additional scenarios, they were removed from the dataset. Teachers that responded to the practice item and at least three additional scenarios were retained in the dataset, given that they responded to at least half of the scenarios (all of the teachers who responded to the practice scenario and only three or four additional scenarios (a total of 7 teachers) responded only to the scenarios presented in sequential order and then abandoned the survey). This results in a dataset of 203 responses.

Of the 203 responses, 190 teachers (94%) answered all of the items, including the demographic questions. One hundred and ninety-two (192) respondents (95%) answered all of the scenario items. Individual items had varying levels of missing data, as displayed in Table 4.14 below. The percentage of missing data increases as the order of the scenarios increases, as some teachers abandoned the survey after answering the practice scenario and then three or four additional scenarios. This is an acceptable amount of missing data, given that each individual item is missing 5% or less. Listwise deletion was utilized in the following analyses.

Table 4.14

Missing data from the full administration

Item	Number of missing responses	Percentage missing
Scenario items		
P	2	1.0%
A	0	0.0%
B	1	0.5%
C	0	0.0%
D	5	2.5%
E	7	3.4%
F	9	4.4%
G	9	4.4%
Demographic items		
Gender	11	5.4%
Race/Ethnicity	10	4.9%
Years of teaching experience	10	4.9%
Primary teaching role	10	4.9%

Demographic information from the full administration is displayed in Table 4.15. Eighty-one percent (81%) of respondents were female, 17% were male, and 2% were nonbinary (neither female nor male). Eighty-six percent (86%) were White, 8% were Asian, 4% were more than one race, 1% were Black/African American, 1% were Hispanic/Latinx, 1% selected other, and 1% preferred not to answer. When asked how many years of teaching experience they had, 6% had 0-3 years, 11% had 4-6 years, 16% had 7-10 years, 16% had 11-15 years, and 52% had more than 15 years. When asked for their primary teaching role, 25% responded that they were secondary general content teachers (English/ELA, History/Social Studies, Math, or Science), 18% were secondary specialist teachers, 18% were elementary homeroom teachers, 10% were elementary specialist teachers, 2% were administrators, and 28% were in some other kind of role (including English Language Learner (ELL) teachers, special education teachers, interventionists, literacy coaches, and preschool teachers).

Table 4.15

Demographics of respondents from the full administration

Demographic group	Number of respondents	Percentage of respondents
Gender		
Female	156	81.3%
Male	32	16.7%
Nonbinary (neither female nor male)	3	1.6%
Prefer not to answer	1	0.5%
Race/Ethnicity		
African American/Black	1	0.5%
Asian	15	7.8%
Hispanic/Latinx	1	0.5%
More than one race	7	3.6%
White	165	85.5%
Other	2	1.0%
Prefer not to answer	2	1.0%
Years of teaching experience		
0-3 years	11	5.7%
4-6 years	21	10.9%
7-10 years	30	15.5%
11-15 years	30	15.5%
More than 15 years	101	52.3%
Primary teaching role		
Administrator	4	2.1%
Elementary homeroom teacher	34	17.6%
Elementary specialist teacher	19	9.8%
Secondary general content teacher	49	25.4%
Secondary specialist teacher	34	17.6%
Other	53	27.5%

Answer choices for the scenario items were coded as follows (the same as the coding for the pre-pilot and pilot administrations):

- Teacher X is much less skilled than me: 5
- Teacher X is slightly less skilled than me: 4
- Teacher X is just like me: 3
- Teacher X is slightly more skilled than me: 2
- Teacher X is much more skilled than me: 1

As with the pre-pilot and pilot administrations, scenario items with a higher mean value were “easier” to respond to, meaning that it was easier for a respondent to identify themselves as a higher skill level than the teacher in the scenario. Scenario items with a lower mean value were “harder” to respond to, meaning that it was harder for a respondent to identify themselves as a higher skill level than the teacher in the scenario. Table 4.16 displays descriptive statistics for the scenario items from the full administration. The items are ordered by their hypothesized continuum in this table, with the hypothesized hardest item first and the hypothesized easiest item last. The items are labeled by their teacher name first (a letter), followed by the levels of the facets in the scenario (where H equals high, M equals medium, and L equals low) and the actual facet and level combinations in the scenario. Although the hypothesized continuum will be evaluated using a Rasch model, looking at the descriptive statistics in this way can provide some evidence to evaluate the hypothesized structure for proof of concept for the construct of using data to inform classroom instruction.

The descriptive statistics for the scenario items provide some evidence of proof of concept of the hypothesized scenario structure, and provide some evidence of success for the changes to some scenarios from the pilot. In contrast to the pre-pilot and pilot results, the item means are now ordered and consistent with their hypothesized structure, where scenario C has the lowest item mean and scenario F has the highest item mean. Scenarios E and P should have similar item means based on the hypothesized structure, and although scenario P's mean is slightly lower than scenario E's mean, both fall into their hypothesized order in comparison to the other scenarios. All scenarios had at least one response in each response category, except for scenario G, which had no responses in the "Teacher X is much less skilled than me" (score of 5) answer category.

Table 4.16

Descriptive statistics for the scenario items from the full administration ordered by the hypothesized scenario structure

Item	Mean	Standard Deviation	Range
C: HHH (D3, G3, O3)	2.45	0.758	1-5
G: HHM (E3, G3, I2)	2.72	0.842	1-4
A: HMM (I3, II2, O2)	3.54	0.845	1-5
E: MMM (Q2, D2, O2)	4.03	0.740	1-5
P: MMM (Q2, E2, I2)	3.79	0.799	1-5
B: MML (E2, G2, II1)	4.27	0.745	1-5
D: MLL (Q2, I1, II1)	4.47	0.785	1-5
F: LLL (Q1, D1, E1)	4.55	0.698	1-5

Rasch Analysis

The person-item variable map from the Rasch analysis including scenario P is presented below in Figure 4.11 to identify the location of scenario P on the hypothesized continuum from the full administration. The person-item variable map from the Rasch analysis without scenario P is presented in Figure 4.12. The teachers who responded to the survey are plotted on the left-hand side of the middle line and the items are plotted on the right-hand side of the middle line in both figures. All remaining analyses will be conducted without scenario P, as it was included solely as a practice scenario in the survey instrument. As a reminder, the hypothesized continuum of scenarios (displayed in Table 4.16) is as follows:

C
G
A
E/P
B
D
F

The plotted scenarios in Figure 4.11 and Figure 4.12 follow the hypothesized structure, indicating that the changes made to the scenarios from the pilot administration were successful. In the pilot administration, scenarios P and F were not in the hypothesized order: scenario F was located with scenario E, when F should have been located at the bottom of the scale, and scenario P was located with scenario B, when P should have been above B. In the full administration, scenario F is located at the bottom of the scale and scenario P is located above scenario B. In the pilot, scenarios B, D, E, F, and P were also closely clustered together; changes to those scenarios appear to have been successful in separating E and B, although scenarios D and F are still clustered closely together at the bottom of the scale.

Similar to the pilot, the item estimates are still slightly below the person estimates in the variable map (the mean for item estimates is denoted by the M on the right side of the line and the mean for the person estimates is denoted by the M on the left side of the line). This can make it difficult to describe a person at the very high or very low end of the scale because there are no scenarios as high or as low as those most extreme person locations, although there is only one person (based on the full administration) below scenario F.

Figure 4.11

Person-item variable map from the full administration including scenario P

```
MEASURE                                                        PERSON - MAP - ITEM
                                                                   <more>|<rare>
    6                                                                    +
                                                                         |
                                                                         |
                                                         68 78 10 12 16  |
    5                                                                    +
                                                                        T|
                                                   28 55 90 10 11 15 19  |
                                                                         |
    4                               2  3  19 29 56 75 87 88 12 13 15 20  +T
                                                                         |
                              4  9  15 30 32 73 89 10 10 11 12 16 18 18 S|  C
                                                                         |
    3                      7  14 21 52 57 65 11 12 12 13 15 17 17 18 19  +
                              6  38 39 53 72 85 91 92 98 14 14 15 17 17  |  G
                                                               17 19 20  |
      8  12 24 41 42 45 51 67 82 93 97 99 10 11 13 13 13 13 14 14 19 20  |
    2 37 54 58 61 64 74 83 94 95 10 11 11 12 16 16 16 18 18 19 19 19 20 M+S
                                                                         |
         5  20 22 31 36 46 49 63 77 96 11 11 11 12 12 14 14 14 15 16 16  |
                              17 18 23 33 35 43 66 86 10 11 13 14 15 17  |
    1           25 27 34 44 50 60 62 71 84 10 14 15 16 17 17 18 18 18 18  +
                                             13 48 70 79 10 14 15 15 19 S|
                                                                         |  A
                                          11 16 12 13 13 13 15 16 18 19  |
    0                                                       80 10 17 17  +M P
                                                               26 47 16  |
                                                                     19  |
                                                            10 59 76 81 T|  E
   -1                                                                    +
                                                                     69  |
                                                                         |  B
                                                                     12  |
   -2                                                                40  +S
                                                                         |  D
                                                                         |  F
                                                                         |
   -3                                                                    +
                                                                         |
                                                                         |
                                                                         |
   -4                                                                    +T
                                                                         |
                                                                         |
                                                                         |
   -5                                                                    +
                                                                         |
                                                                         |
                                                                         |
   -6                                                                    +
                                                                         |
                                                                         |
                                                                         |
   -7                                                                 1  +
                                                                   <less>|<freq>
```

Figure 4.12

Person-item variable map from the full administration without scenario P

```
MEASURE                                      PERSON - MAP - ITEM
                                                 <more>|<rare>
    6                                                  +
                                                       |
                                                       |
                                          6 7 1 1 1 1  |
    5                                                  +
                                                    2 T|
                                3 1 2 5 5 9 1 1 1 1 1  |T
                                                       |
    4                                                  +
                              2 9 2 3 7 8 8 1 1 1 1 1  |
                                                      S|  C
          4 7 1 1 3 3 5 5 6 7 8 1 1 1 1 1 1 1 1 1 1 1  |
    3                                                  +
          6 2 2 3 4 7 8 9 9 9 9 1 1 1 1 1 1 1 1 1 1 1  |  G
                                                  1 2  |
        8 1 3 4 4 5 5 8 8 9 9 9 9 1 1 1 1 1 1 1 1 1 2  |S
    2                                               1 M+
                3 5 6 6 6 6 7 1 1 1 1 1 1 1 1 1 1 1 1 2  |
      5 1 2 2 3 3 4 4 5 6 6 7 8 9 1 1 1 1 1 1 1 1 1 1  |
                                                       |
    1         1 2 3 4 4 6 7 8 1 1 1 1 1 1 1 1 1 1 1 1  +
                                  2 2 3 4 7 7 1 1 1 1  |
                                                      S|  A
                            1 1 1 5 1 1 1 1 1 1 1 1 1  |
    0                                         8 1 1 1  +M
                                              2 4 1 1  |
                                                    1  |
                                              1 5 7 8 T|  E
   -1                                                  +
                                                       |
                                                    6  |  B
                                                       |
   -2                                               1  +
                                                       |S D
                                                    4  |  F
                                                       |
   -3                                                  +
                                                       |
                                                       |
                                                       |
   -4                                                  +
                                                       |
                                                       |T
                                                       |
   -5                                                  +
                                                       |
                                                       |
                                                       |
   -6                                                  +
                                                       |
                                                       |
                                                       |
   -7                                               1  +
                                                 <less>|<freq>
```

The person separation from the full administration is 1.79 with a reliability of 0.76. As described for the pre-pilot and pilot administrations, using Equation 4.1, the number of statistically distinct strata in the data is 2.72, which is similar to the hypothesized number of strata, which is 3. The item separation from the full administration is 16.6 (reliability of 1.00), indicating the items differentiate well between high and low level of skillsets for using data to inform instruction.

Figure 4.13 displays the Andrich thresholds for the full administration data. The thresholds indicate the place where a respondent has a 50% probability of choosing the next highest response option for the item. For example, G.4 indicates the place where a respondent has a 50% probability of choosing a 3 or a 4 for scenario G. Respondents in this map are plotted directly across the center line from the response option they were likely to select on the survey. Based on this variable map, there is slightly less variation in response choices on the full administration compared to the pilot, but more variation than on the pre-pilot. On the full administration, respondents were less likely to select response options 1, 2 and 3 (Teacher X is much more skilled than me, Teacher X is slightly more skilled than me, and Teacher X is just like me) than they were to select those choices on the pilot.

Figure 4.13

Andrich thresholds from the full administration

```
MEASURE    PERSON - MAP - ITEM - Andrich thresholds (modal categories if ordered)
               <more>|<rare> 2        3        4        5
    8                +
                     |
                     |
                     |                                 C.5
    7                +
                     |
                     |                                 G.5
                     |
    6                +
                     |
                     |
                ###  |
    5                +                        C.4
                  . T|
             .#####  |T
                     |                        G.4      A.5
    4                +
             ######  |
                    S|
        ###########  |
    3                +                                 E.5
        ###########  |
                  #  |
       .###########  |S                                B.5
    2             . M+               C.3      A.4
         .#########  |
       ############  |                                 D.5
                     |               G.3               F.5
    1     ##########  +
              #####  |                        E.4
                    S|
            .######  |
    0            ##  +M     C.2               B.4
                 ##  |
                  .  |
                 ## T|      G.2               D.4
   -1                +               A.3      F.4
                     |
                  .  |
                     |
   -2             .  +
                     |S              E.3
                  .  |
                     |
   -3                +      A.2      B.3
                     |
                     |
                     |               D.3
   -4                +               F.3
                     |      E.2
                     |T
                     |
   -5                +      B.2
                     |
                     |
                     |      D.2
   -6                +      F.2
                     |
                     |
                     |
   -7             .  +
               <less>|<freq> 1        2        3        4
EACH "#" IS 2: EACH "." IS 1
```

Table 4.17 displays statistics related to the rating scale category structure. The observed average of person estimates is ordered as intended, with the average increasing monotonically, as expected, as the category labels increase. The Andrich thresholds also increase monotonically as the categories increase, as intended. Response option 1 (Teacher X is much more skilled than me) was not selected as frequently as expected and was selected by some overall high-scoring teachers who were not expected to give such a low response on a hypothesized easier scenario, as indicated by the INFIT and OUTFIT values over 1.4 in Table 4.17. This will be discussed in more detail when examining misfit data next. The Category Characteristic Curves (CCCs) in Figure 4.14 provide evidence of a typical rating scale structure for all of the scenario items on the survey. The categories are in the correct ordered progression. The probability of responding with a 5 (Teacher X is much less skilled than me) is lower than on the pre-pilot and pilot administrations, which provides some evidence of successful edits to the scenarios to make them more difficult. The probability of responding with a 3 (Teacher X is just like me) is higher than on the pilot; a lower probability of responding with a 3 would fit the rating scale structure better. The CCCs from the full administration, however, appear to fit the rating scale structure overall better than in the pre-pilot or pilot administrations.

Table 4.17

Observed averages of person estimates and Andrich thresholds from the full administration

Response label	Response frequency	Observed Average	Infit	Outfit	Andrich threshold
1 (Teacher X is much more skilled than me)	48	-2.67	1.49	1.54	N/A
2 (Teacher X is slightly more skilled than me)	150	-1.55	1.08	1.08	-3.56
3 (Teacher X is just like me)	359	0.30	0.94	0.93	-1.56
4 (Teacher X is slightly less skilled than me)	427	2.67	0.92	0.83	1.41
5 (Teacher X is much less skilled than me)	406	4.62	0.95	0.95	3.71

Figure 4.14

Category Characteristic Curves (CCCs) for the full administration

```
        CATEGORY PROBABILITIES: MODES - Andrich thresholds at intersections
P      -+-----------+-----------+-----------+-----------+-----------+-
R  1.0 +                                                             +
O      |                                                             |
B      |                                                             |
A      |                                                             |
B   .8 +1                                                           5+
I      | 11                                                        5 |
L      |   1                         333                         55  |
I      |    11                    333   33                      5    |
T   .6 +      1                 33        33       444444      5     +
Y      |       1   22222222    3            3    44      44   5      |
    .5 +        122        22 3              3344          445       +
O      |        21          3*                43            54       |
F   .4 +      22  1        3  2              4  3          5  44     +
       |     2     1      3    22          44    3       55     4    |
R      |   22       1    3       2        4       33    5        4   |
E      |  2          11 3         22    44          3  5          44 |
S   .2 +22            3*            2  4             *5             4+
P      |            33  11           **            55 33             |
O      |          33      11      444  222       55     333          |
N      |     33333          1****4        222**55          3333      |
S   .0 +*********************5555************11**********************+
E      -+-----------+-----------+-----------+-----------+-----------+-
       -5          -3          -1           1           3           5
        PERSON [MINUS] ITEM MEASURE
```

Table 4.18 displays misfit results for the items on the full administration. The items are ordered by logit estimate in this table. Table 4.19 displays the same misfit information for the items on the full administration, but in item entry order (i.e., the order in which the items are presented on the instrument). Table 4.19 can be useful to identify any start-up effects for the first item or any fatigue effects with the last item. In both tables, values over 1.4 for INFIT indicate a problematic item. For the full administration, none of the items have an INFIT value over 1.4, although scenario D has an INFIT value of 1.30. On the pilot administration, scenarios F and C had INFIT values over 1.3 (1.35 and 1.32, respectively). On the full administration, they have lower INFIT values (1.24 and 0.94, respectively).

The OUTFIT statistics (where the value is larger than 1.4) indicate items that have at least one highly unexpected response across the respondents. None of the scenarios have OUTFIT statistics with a value larger than 1.4. On the pilot, scenarios F and C had OUTFIT values over 1.3 (with OUTFIT values of 1.35 and 1.33, respectively); on the full administration, they had OUTFIT values of 1.10 and 0.93, respectively, indicating improvement.

Based on Table 4.19, it does not appear that there are any start-up effects or fatigue effects across respondents as a whole, as the INFIT and OUTFIT values for the first scenario (A) and the last scenario (G) are below the threshold of 1.4.

Table 4.18

Fit statistics for the full administration ordered by logit estimate

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic (INFIT)		Unweighted Fit Statistic (OUTFIT)	
		MNSQ	ZSTD	MNSQ	ZSTD
C	3.50 (0.11)	0.94	-0.57	0.93	-0.72
G	2.78 (0.12)	1.07	0.68	1.06	0.61
A	0.60 (0.12)	0.91	-0.93	0.91	-0.94
E	-0.73 (0.12)	0.76	-2.73	0.73	-2.89
B	-1.46 (0.13)	0.95	-0.50	0.92	-0.65
D	-2.20 (0.14)	1.30	2.67	1.10	0.69
F	-2.48 (0.15)	1.24	2.08	1.10	0.64

Table 4.19

Fit statistics for the full administration ordered by item entry order

Item	Logit Estimate (S.E.)	Information-Weighted Fit Statistic		Unweighted Fit Statistic	
		MNSQ	ZSTD	MNSQ	ZSTD
A	0.60 (0.12)	0.91	-0.93	0.91	-0.94
B	-1.46 (0.13)	0.95	-0.50	0.92	-0.65
C	3.50 (0.11)	0.94	-0.57	0.93	-0.72
D	-2.20 (0.14)	1.30	2.67	1.10	0.69
E	-0.73 (0.12)	0.76	-2.73	0.73	-2.89
F	-2.48 (0.15)	1.24	2.08	1.10	0.64
G	2.78 (0.12)	1.07	0.68	1.06	0.61

Looking at misfitting individuals and their residuals can help identify response patterns that are unexpected (please see Appendix I for the person-response table from the full administration). The majority of individuals with misfitting responses from the full administration had unexpected responses to scenarios F and G. Scenario F was hypothesized to be the easiest scenario, and the majority of individuals with misfitting data for scenario F gave lower than expected responses to scenario F. This could be some evidence of confusion on how to respond to scenario F. Scenario G is at the end of the survey and could be some evidence of survey fatigue for individuals, especially given the higher cognitive load required when responding to scenario-based items, or it could be related to social desirability bias, as scenario G was hypothesized to be the second most difficult scenario. It appears that both explanations are plausible for scenario G, as there are some individuals who had lower than expected responses to scenario G and other individuals who had higher than expected responses to scenario G. In addition, some of the individuals with misfitting responses had unexpected responses to scenario A, which could be evidence of a start-up effect for individuals. This pattern was also evident in the residuals from the pilot administration. There were also a small number of individuals who had lower than expected responses to scenario D, which could be related to confusion, and a small number of individuals with unexpected responses to scenario C (the most difficult scenario) which could be related to social desirability bias. Finally, there were a few unexpected responses to other scenarios, which is to be expected with a larger sample of respondents. There were no evident patterns to these other unexpected responses.

Principal Components Analysis

A principal components analysis (PCA) was run on the Rasch residuals from the full administration. As described previously, the purpose of the PCA for the residuals is to look for evidence of an unidentified construct in the residual data. In this analysis, unlike a typical PCA, the goal is to identify zero patterns in the residual data to meet the assumption of unidimensionality. The residual data should be similar to randomly generated data to provide evidence that there is no unidentified construct in the residual data.

Table 4.20 displays the eigenvalues and the percentage of variance explained for each eigenvalue from the PCA for the full administration residuals and randomly generated data. Although a component with an eigenvalue slightly over 2 was extracted from the full administration residuals, which could indicate an unexplained construct in the residuals, the difference in the variance explained by the first eigenvalue for the full administration residuals and the randomly generated data is just over 10%, which is not a large difference. This same pattern was evident in the pilot administration data. When the scree plots for the residuals and the randomly generated data are examined (in Figures 4.15 and 4.16, respectively), there does appears to be a break between components one and two in the residual data PCA; a break between components one and two is somewhat apparent in the scree plot of the randomly generated data, although not as pronounced. Figures 4.17 and 4.18 display the component loading plots for the residual PCA and the randomly generated data PCA. The components in the residual data plot do not appear clustered into groups and approximate random data (which would be plotted in a circular pattern). When compared to the component loading plot of the randomly generated data PCA, the residual PCA component plot appears similar, although the location of the randomly generated data plot is shifted to the right. These combined results provide some evidence that there is no unidentified construct present in the residuals.

Table 4.20

Principal Components Analysis results for the residuals from the full administration compared to random data

Full administration residuals			Random data		
Component number	Eigenvalue	% Variance explained	Component number	Eigenvalue	% Variance explained
1	2.086	29.803	1	1.265	18.074
2	1.216	17.374	2	1.111	15.871
3	1.109	15.839	3	1.063	15.191
4	1.010	14.427	4	1.033	14.756
5	0.811	11.580	5	0.947	13.524
6	0.725	10.363	6	0.855	12.213
7	0.43	0.615	7	0.726	10.372

Figure 4.15

Scree plot for residuals from the full administration

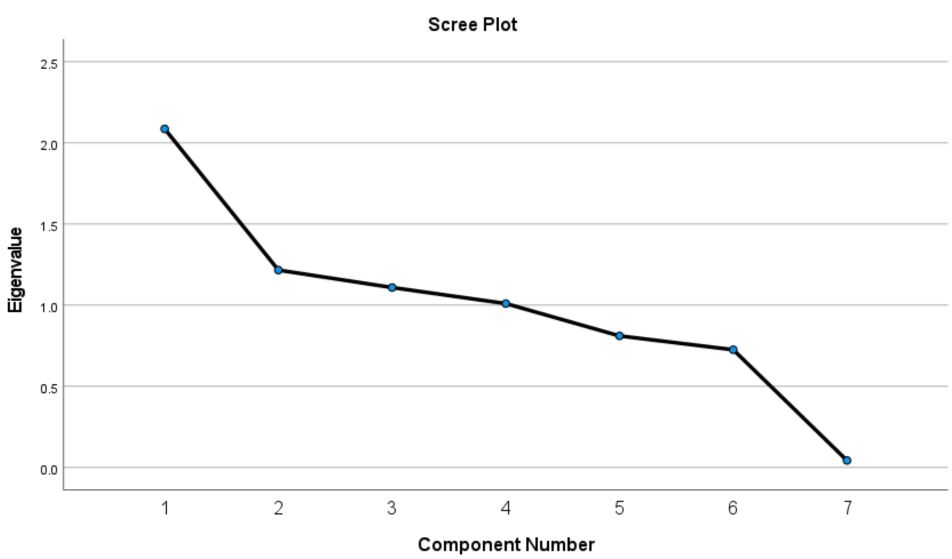


Figure 4.16

Scree plot for randomly generated data

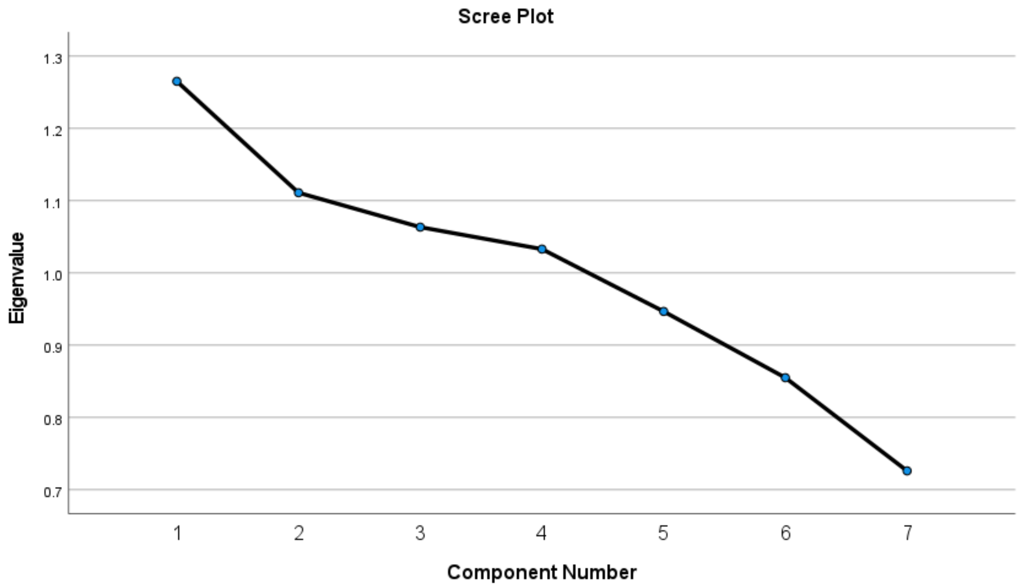


Figure 4.17

Component loading plot for residuals from the full administration

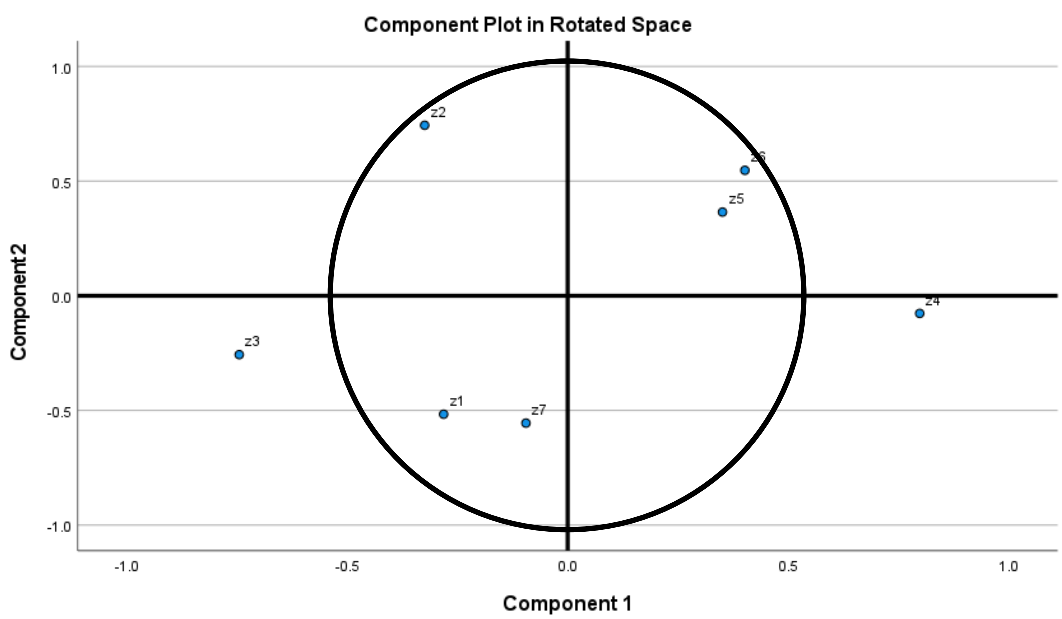
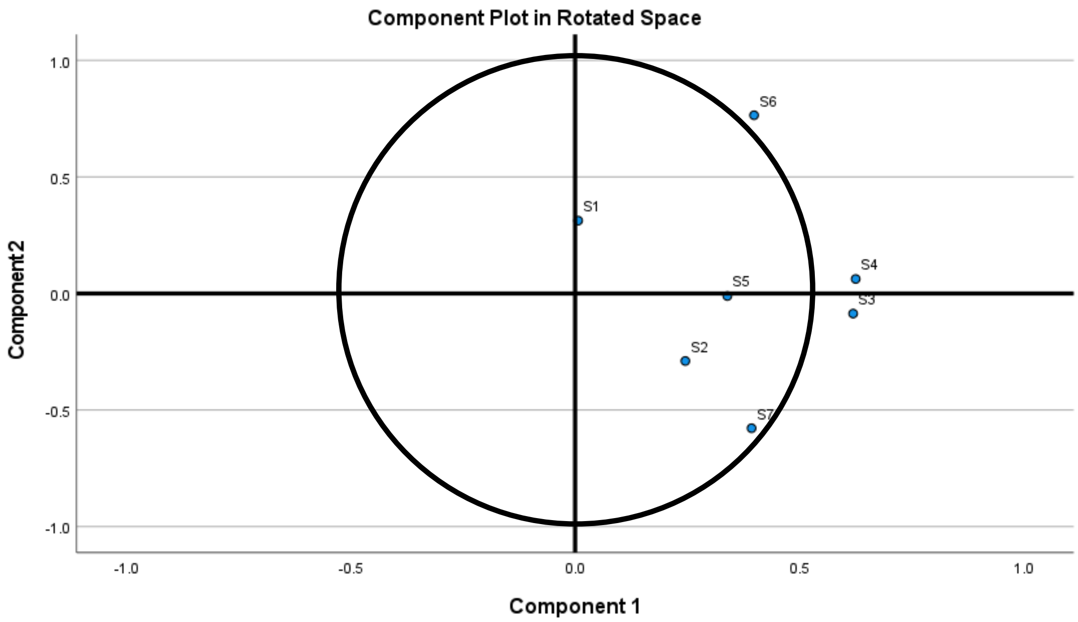


Figure 4.18

Component loading plot for residuals for randomly generated data



Final Variable Maps with Interpretation

As discussed in Chapter 1, one of the benefits of the RGS approach and scenario-based scales is the interpretability of the scores. Detailed descriptions of people at particular locations along the hierarchical continuum can be developed to provide useful information to those using the data from the survey instrument (Ludlow et al., 2020). Because items and people are placed on the same scale (as displayed in the person-item variable maps in this chapter), detailed descriptions can be developed for people with raw scores at particular locations on the continuum by using the scenarios at those particular locations. As noted in Chapter 1, in the context of this work, the detailed descriptions of teachers along the continuum of using data to inform classroom instruction can be utilized by school or district staff to help differentiate professional development for teachers and/or they can use this information when supporting teacher teams by identifying particular areas in which a team may need additional support.

To develop variable maps with detailed interpretations of scores, the person-item variable map is modified in Figure 4.19 to show raw scores in addition to logits on the left hand side of the map. The raw scores are equal to an individual's total score on the survey (i.e., the sum of their responses to all items on the survey) and range from 7 (selecting a 1, or "Teacher X is much more skilled than me" for all 7 scenarios) to 35 (selecting a 5, or "Teacher X is much less skilled than me" for all 7 scenarios). Appendix J displays the score conversion table that links logits and raw scores. In addition to converting logits into raw scores in the person-item variable map in Figure 4.19, lines have been drawn on this map to identify zones where average scores fall: for example, where a person scoring an average of 4 on all items falls on this map. The average score lines divide the person-item variable map into score zones with example scenarios that describe an individual in that zone. Table 4.21 displays similar information to the person-item variable map in Figure 4.19, but with a more detailed description of an individual in each score zone, based on the scenarios within that zone and above and below that zone.

Teachers scoring in the range of 7-13 are expected to identify their own skillset as much lower than or slightly lower than the teachers in all of the scenarios. There is no example scenario for this score range, and only one respondent fell into this score range on the full administration. Teachers scoring in this range have difficulty generating or narrowing down their thoughts on student learning to write clearly defined questions. They struggle to identify data to answer these questions about student learning and have trouble organizing and displaying data clearly. Teachers in this range cannot compare multiple data sources, struggle to gather data, and need support to critically examine data or analyze data to answer their questions about student learning. They also need support to interpret the results from their data analysis and have trouble writing clear and aligned goals for student learning based on hypotheses for improvement of student learning. They also struggle to consider potential consequences of their goals. Teachers in this range need help to both identify interventions to attain their goals for student learning and determine if they need any professional development before implementing the intervention. They also require support to identify any data required to document their progress towards their goals. Once they have identified an intervention, they often struggle to implement it as intended and need support to plan for and collect data to document their progress towards these goals, struggling to do this in a systematic way. They often do not share their progress with colleagues during the intervention. Teachers in this range are unable to independently examine student outcome data from the intervention to determine if the outcomes meet their stated goals or if the intervention was implemented as intended, and are not able to identify next steps without support.

Teachers scoring in the range of 14-16 are expected to identify their own skillset as slightly lower than the teachers in the scenarios in this range (D and F) and much lower than the scenarios above this range. Teachers in this range often need help narrowing down their thoughts on student learning to write clearly defined questions. They struggle to both understand what data can help answer these questions and to organize and display this data visually. They also need help to gather these data and then think critically about and analyze these data, as well as determine if they can compare multiple data sources. They need support to interpret the results from their data analysis and have trouble generating hypotheses about how to improve student learning. They also need support to write clear and aligned goals for student learning based on these hypotheses for improvement and struggle to consider potential consequences of their goals for student learning. They need support from others to identify potential interventions and to determine if they need professional development prior to implementing their intervention. They also require support to identify data to document their progress towards their student learning goals. They have trouble implementing their chosen intervention as intended, don't share progress with others during the implementation of the intervention, and need help to systematically plan for and monitor progress towards their goals. Teachers in this range are unable to examine student outcome data from the intervention on their own to determine if the outcomes meet their stated goals or if the intervention was implemented as intended, and are not able to identify next steps without support.

Teachers scoring in the range of 17-20 are expected to identify their own skillset as slightly lower than the teachers in the scenarios in this range (B and E), much lower than the scenarios above this range, and slightly higher than the scenarios below this range (D and F). Teachers scoring in this range often need support generating or narrowing down their thoughts on student learning to write clearly defined questions. They also often seek support to identify data to answer these questions about student learning and may often need support to organize and display data clearly. Teachers in this range benefit from support to critically examine data or analyze data to answer their questions about student learning, and also need some support to compare multiple data sources and gather data. They often need help to interpret the results from their data analysis and to generate hypotheses about how to improve student learning. They have trouble writing clear and aligned goals for student learning based on these hypotheses for improvement and struggle to consider potential consequences of their goals for student learning. Teachers in this range may require some support to both identify interventions to attain their goals for student learning and determine if they need any professional development before implementing the intervention. They often need help identifying any data to document their progress towards their goals. Once they have identified an intervention, they often struggle to implement it as intended and often need support to plan for and collect data to document their progress towards these goals, struggling to do this in a systematic way. They often do not share progress with colleagues during the intervention. Teachers in this range need help to examine student outcome data from the intervention to determine if the outcomes meet their stated goals or if the intervention was implemented as intended, and need support to identify next steps.

Teachers scoring in the range of 21-27 are expected to identify their own skillset as similar to the teacher in the scenario in this range (A), as slightly higher or much higher than the teachers in the scenarios below this score range, and as slightly lower or much lower than the teachers in the scenarios above this score range. Scenario A is the example scenario within this score range. Teachers scoring in this range are successful at generating or narrowing down their thoughts on student learning to write clearly defined questions. They can usually identify data to answer these questions about student learning, but may need some help organizing and displaying data clearly. Teachers in this range can independently gather data, but sometimes need the help of others to compare multiple data sources or critically examine or analyze data to answer their questions about student learning. They occasionally need support to interpret the results from their data analysis and to generate hypotheses about how to improve student learning. They find support helpful to write clear and aligned goals for student learning based on their hypotheses for improvement. They sometimes consider potential consequences of their goals for student learning. Teachers in this range are successful at identifying interventions to attain their goals for student learning and determining if they need any professional development before implementing the intervention. They are also able to plan to assess progress towards their goals. Once they have identified an intervention, they sometimes have trouble implementing it as intended and sometimes need support to plan for and collect data to document their progress towards these goals. They usually share progress with colleagues during the intervention. Teachers in this range find support helpful when examining student outcome data from the intervention to determine if the outcomes meet their stated goals or if the intervention was implemented as intended, and can usually determine their next steps without support.

Teachers scoring in the range of 28-29 are expected to identify their own skillset as slightly higher than the teacher in the scenario in this range (G), much higher than the teachers in the scenarios in the ranges below, and slightly lower than the teacher in the scenario above this range. Scenario G is the example scenario within this score range. Teachers scoring in this range can successfully generate or narrow down their thoughts on student learning to write clearly defined questions. They are usually able to identify data to answer these questions about student learning, but may occasionally need some support to organize and display data clearly. Teachers in this range can successfully compare multiple data sources and gather data, and are able to critically examine data and analyze data to answer their questions about student learning. They are able to independently interpret results from their data analysis and generate hypotheses about how to improve student learning, but may benefit from support to write clear and aligned goals for student learning based on their hypotheses for improvement. They consider potential consequences of their goals for student learning. Teachers in this range can independently identify interventions to attain their goals for student learning and determine if they need any professional development before implementing the intervention. They can successfully plan to document progress towards their goals. Once they have identified an intervention, they can implement it as intended while systematically collecting data to document their progress towards these goals. They share progress with colleagues during the intervention. Teachers in this range may benefit from some support to examine student outcome data from the intervention to determine if the outcomes meet their stated goals or if the intervention was implemented as intended, but can usually identify next steps without support.

Teachers scoring in the range of 30-34 are expected to identify their own skillset as slightly higher than the teacher in the scenario in this range (C) and much higher than the teachers in the scenarios in the ranges below. Scenario C is the example scenario within this score range. Teachers scoring in this range can successfully generate or narrow down their thoughts on student learning to write clearly defined questions. They can independently identify data to answer these questions about student learning and clearly organize and display their data. Teachers in this range are able to compare multiple data sources, gather data, and critically examine and analyze data to answer their questions about student learning. They are successful at interpreting the results from their data analysis and generating hypotheses about how to improve student learning, while writing clear and aligned goals for student learning based on these hypotheses for improvement. They consider potential consequences of their goals for student learning. Teachers in this range can both identify interventions to attain their goals for student learning and determine if they need any professional development before implementing an intervention. They can successfully plan to document progress towards these goals. Once they have identified an intervention, they can implement it as intended while also systematically collecting data to document their progress towards these goals. They also share progress with colleagues during the intervention. Teachers in this range successfully examine student outcome data from the intervention to determine if the outcomes meet their stated goals and if the intervention was implemented as intended, and are able to identify next steps.

Teachers scoring at the top of the range (35) are expected to identify their own skillset as much higher than the teachers in the scenarios for all levels of skills. There is no example scenario for a score of 35 and teachers scoring at 35 are highly successful at using data to inform classroom instruction.

Figure 4.19

Interpretative person-item variable map from the full administration

```
SCORE  MEASURE                                         PERSON - MAP - ITEM
                                                            <more>|<rare>
  33      6                                                       |      Average item score = 5
                                                                  +
                                                                  |
                                                                  |
  32                                                 6 7 1 1 1 1  |
          5                                                       +
                                                               2 T|
  31                                       3 1 2 5 5 9 1 1 1 1 1  |T
                                                                  |
          4                                                       +
  30                                   2 9 2 3 7 8 8 1 1 1 1 1    |
                                                                 S| C
  29             4 7 1 1 3 3 5 5 6 7 8 1 1 1 1 1 1 1 1 1 1 1     |
          3                                                       +
  28             6 2 2 3 4 7 8 9 9 9 9 1 1 1 1 1 1 1 1 1 1 1     | G    Average item score = 4
                                                             1 2  |
  27           8 1 3 4 4 5 5 8 8 9 9 9 9 1 1 1 1 1 1 1 1 1 2     |S
          2                                                  1 M+
  26                     3 5 6 6 6 6 7 1 1 1 1 1 1 1 1 1 1 1 2    |
  25         5 1 2 2 3 3 4 4 5 6 6 7 8 9 1 1 1 1 1 1 1 1 1 1      |
                                                                  |
  24      1            1 2 3 4 4 6 7 8 1 1 1 1 1 1 1 1 1 1 1 1    +
  23                                          2 2 3 4 7 7 1 1 1 1 |
                                                                 S| A
  22                                  1 1 1 5 1 1 1 1 1 1 1 1 1   |
  21      0                                              8 1 1 1  +M    Average item score = 3
  20                                                     2 4 1 1  |
                                                               1  |
  19                                                     1 5 7 8 T| E
         -1                                                       +
  18                                                              |
  17                                                           6  | B
                                                                  |
  16     -2                                                    1  +
                                                                  |S D
  15                                                           4  | F
                                                                  |
  14     -3                                                       +     Average item score = 2
                                                                  |
  13                                                              |
                                                                  |
  12     -4                                                       +
                                                                  |
  11                                                              |T
                                                                  |
  10     -5                                                       +
                                                                  |
                                                                  |
   9                                                              |
         -6                                                       +
                                                                  |
                                                                  |
   8                                                              |
         -7                                                    1  +
                                                                  +
                                                                  |
                                                                  |
                                                                  |
   7                                                              +     Average item score = 1
```

C: Teacher C can identify data to answer their questions about student learning and consistently organizes that data clearly. This teacher independently interprets results from their data analysis related to their questions about student learning, while writing clear and aligned goals for student learning based on hypotheses about how to improve student learning that come from their analysis. Once they have implemented a student learning intervention, they can successfully examine student outcome data collected from this intervention and are able to identify their next steps.

A: Teacher A can independently identify interventions to meet their student learning goals, while also planning to assess progress towards these goals. They sometimes have trouble implementing their chosen intervention as intended, and sometimes have difficulty collecting data about student progress towards their goals during the intervention. Teacher A finds support helpful when examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.

B: Teacher B occasionally has difficulty thinking critically about and analyzing data to help answer their questions about student learning. Once they complete their data analysis focused on these questions, they may need support to generate hypotheses about how to improve student learning from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher B has trouble planning for student progress towards their student learning goals and often holds back on sharing progress with colleagues during their implementation.

Table 4.21

Scale score interpretations

Score range	Description	Example Scenario	Skillsets in this range by facet
7-13	Teacher identifies their own skillset as much lower than or slightly lower than the teachers in all of the scenarios	N/A	**Q**: Difficulty generating/narrowing down thoughts into clearly defined questions. **D:** Struggles to identify data to answer questions; has trouble identifying ways to organize and display data clearly. **E:** Struggles to gather data; cannot compare multiple data sources; needs support to critically examine/analyze data. **G:** Needs support to interpret results from data analysis; has trouble generating hypotheses about how to improve student learning and struggles to write clear and aligned goals based on hypotheses for improvement. Struggles to consider potential consequences of goals. **I:** Needs help to identify interventions and determine if they need professional development. Requires support to identify data to document progress towards goals. **II:** Struggles to implement intervention as intended and needs support to collect data to document progress towards goals; often does not share progress with colleagues. **O:** Unable to independently examine

			student outcome data to determine if outcomes meet goals or if intervention was implemented as intended; unable to identify next steps without support.
14-16	Teacher identifies their own skillset as slightly lower than the teachers in scenarios describing low or medium/low levels of skills and much lower than scenarios describing medium/high or high levels of skills	D: Teacher D occasionally needs help narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they require the support of others to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D rarely shares progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.	**Q**: Often needs help generating/narrowing down thoughts into clearly defined questions. **D:** Struggles to identify data to answer questions; has trouble identifying ways to organize and display data clearly. **E:** Struggles to gather data; cannot compare multiple data sources; needs support to critically examine/analyze data. **G:** Needs support to interpret results from data analysis; has trouble generating hypotheses about how to improve student learning and struggles to write clear and aligned goals based on hypotheses for improvement. Struggles to consider potential consequences of goals. **I:** Needs help to identify interventions and determine if they need professional development. Requires support to identify data to document progress towards goals. **II:** Struggles to implement intervention as intended and needs support to collect data to document progress towards goals; often does not share progress with colleagues.

			O: Unable to independently examine student outcome data to determine if outcomes meet goals or if intervention was implemented as intended; unable to identify next steps without support.
17-20	Teacher identifies their own skillset as slightly lower than the teachers in scenarios describing medium/low or medium levels of skills, much lower than scenarios describing medium/high or high levels of skills, and slightly higher than the teachers in scenarios describing low levels of skills	B: Teacher B occasionally has difficulty thinking critically about and analyzing data to help answer their questions about student learning. Once they complete their data analysis focused on these questions, they may need support to generate hypotheses about how to improve student learning from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher B has trouble planning for student progress towards their student learning goals and often holds back on sharing progress with colleagues during their implementation.	**Q**: Often needs help generating/narrowing down thoughts into clearly defined questions. **D:** Often needs help to identify data to answer questions; benefits from support to identify ways to organize and display data clearly. **E:** Benefits from support to gather data; needs some help to compare multiple data sources; often needs support to critically examine/analyze data. **G:** Often needs support to interpret results from data analysis; benefits from support to generate hypotheses about how to improve student learning; struggles to write clear and aligned goals based on hypotheses for improvement. Struggles to consider potential consequences of goals. **I:** Needs help to identify interventions and benefits from help to determine if they need professional development. Often requires support to identify data to document progress towards goals. **II:** Struggles to implement intervention as intended and often

			needs support to collect data to document progress towards goals; often does not share progress with colleagues. **O:** Unable to independently examine student outcome data to determine if outcomes meet goals or if intervention was implemented as intended; unable to identify next steps without support.
21-27	Teacher identifies their own skillset as the same as teachers in scenarios describing medium/high levels of skills, higher than teachers in scenarios describing low or low/medium levels of skills, and lower than teachers in scenarios describing high levels of skills	A: Teacher A can independently identify interventions to meet their student learning goals, while also planning to assess progress towards these goals. They sometimes have trouble implementing their chosen intervention as intended, and sometimes have difficulty collecting data about student progress towards their goals during the intervention. Teacher A finds support helpful when examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.	**Q**: Can successfully generate/narrow down thoughts into clearly defined questions. **D:** Is usually able to identify data to answer questions; occasionally needs help to identify ways to organize and display data clearly. **E:** Can independently gather data; sometimes needs help to compare multiple data sources or to critically examine/analyze data. **G:** Occasionally needs support to interpret results from data analysis; benefits from support to generate hypotheses about how to improve student learning and benefits from some support to write clear and aligned goals based on hypotheses for improvement. Sometimes considers potential consequences of goals. **I:** Can independently identify interventions and can determine if they need professional development. Can

			successfully plan to document progress towards goals. **II:** Sometimes has trouble implementing intervention as intended and sometimes needs support to collect data to document progress towards goals; usually shares progress with colleagues. **O:** Finds support helpful when examining student outcome data to determine if outcomes meet goals or if intervention was implemented as intended; can usually identify next steps without support.
28-29	Teacher identifies their own skillset as slightly higher than teachers in scenarios describing medium/high levels of skills, much higher than teachers in scenarios describing low, low/medium, and medium levels of skills, and slightly lower than teachers in scenarios describing high levels of skills	G: Teacher G successfully compares multiple data sources while critically examining data to help answer their questions about student learning. Interpreting meaning from their data analysis to develop student learning goals is a strength of Teacher G's, and they consider all potential consequences of these goals before moving forward. Teacher G benefits from some support to both identify interventions to help attain these goals and to identify the data required to document progress towards these goals.	**Q**: Can successfully generate/narrow down thoughts into clearly defined questions. **D:** Is usually able to identify data to answer questions; occasionally needs help to identify ways to organize and display data clearly. **E:** Can independently gather data; successfully compares multiple data sources and critically examines/analyzes data. **G:** Can independently interpret results from data analysis and generate hypotheses about how to improve student learning; may benefit from some support to write clear and aligned goals based on hypotheses for improvement. Considers potential

			consequences of goals. **I:** Can independently identify interventions and can determine if they need professional development. Can successfully plan to document progress towards goals. **II:** Is able to implement intervention as intended while collecting data to document progress towards goals; shares progress with colleagues. **O:** Finds support helpful when examining student outcome data to determine if outcomes meet goals or if intervention was implemented as intended; can usually identify next steps without support.
30-34	Teacher identifies their own skillset as slightly higher than teachers in scenarios describing high levels of skills and much higher than teachers in scenarios describing low, low/medium, medium, and medium/high levels of skills	C: Teacher C can identify data to answer their questions about student learning and consistently organizes that data clearly. This teacher independently interprets results from their data analysis related to their questions about student learning, while writing clear and aligned goals for student learning based on hypotheses about how to improve student learning that come from their analysis. Once they have implemented a student learning intervention, they can successfully examine student outcome data collected from this intervention and are able to identify their next	**Q**: Can successfully generate/narrow down thoughts into clearly defined questions. **D:** Is able to identify data to answer questions; can independently organize and display data clearly. **E:** Can independently gather data; successfully compares multiple data sources and critically examines/analyzes data. **G:** Can independently interpret results from data analysis and generate hypotheses about how to improve student learning; writes clear and aligned goals based on hypotheses for improvement. Considers potential

		steps.	consequences of goals. **I:** Can independently identify interventions and can determine if they need professional development. Can successfully plan to document progress towards goals. **II:** Is able to implement intervention as intended while collecting data to document progress towards goals; shares progress with colleagues. **O:** Successful at examining student outcome data to determine if outcomes meet goals or if intervention was implemented as intended; identifies next steps without support.
35	Teacher identifies their own skillset as much higher than teachers in scenarios describing all levels of skills	N/A	Highly successful on all skills involved in using data to inform classroom instruction.

Note. The following abbreviations in the 'Skillsets' column refer to the specific facets in each scenario. Q: Ask questions. D: Identify data. E: Examine data. G: Interpret data to set goals. I: Identify intervention. II: Implement intervention. O: Examine outcomes.

Chapter Summary

This chapter detailed the results from the pre-pilot, the pilot, and the full administrations of the UDII scale. Results from the pre-pilot administration informed modifications to the scenario content for the pilot administration, and results from the pilot administration informed modifications to the scenario content for the full administration. These modifications were described, along with the reasons for these modifications. This chapter also included an interpretative variable map and table for results from the full administration to describe individuals scoring in particular raw score ranges along the continuum of using data to inform classroom instruction. This interpretative variable map and table can be utilized by schools and districts who use this instrument to support individual teachers and teams of teachers in using data to inform classroom instruction by identifying where teachers fall along this continuum.

Chapter 5 : Discussion and Implications

This chapter provides an overview of the results from Chapter 4, followed by a discussion of these results with a particular focus on how they can be used in practice. It then details limitations of the current study and provides suggestions for future directions for research. It closes with a discussion of implications and conclusions.

Overview of Results

The UDII scale was administered in three separate administrations (the pre-pilot, the pilot, and the full), with the results from the data analysis from the first two administrations used to revise the scenarios prior to the next administration. Based on the data analyses, the revisions made to the scenarios after the pre-pilot administration and after the pilot administration were successful, meaning that the revisions resulted in the intended changes.

The pre-pilot administration had responses from 22 current or retired teachers from a convenience sample. One of the major advantages of the RGS methodology in instrument design is the ability to identify and make specific, targeted changes to the scenarios to better capture the construct's hypothesized structure based on the results from the data analysis. Six of the eight scenarios were revised prior to the pilot administration based on the data analysis from the pre-pilot data, which is described in detail in Chapter 4. The Andrich thresholds from the pre-pilot showed that respondents were more likely to select the highest two response options (Teacher X is slightly less skilled than me (4) and Teacher X is much less skilled than me (5)) for all scenarios except scenario C, and the probability of responding with a response of 5 was relatively high as clearly displayed in the CCCs. This indicated that increasing the difficulty of the majority of the scenarios would be beneficial. Based on these results, scenarios A, B, E, G, and P were revised to make them harder to endorse (meaning it is more difficult for a respondent to rate themselves as a higher skill level than the teacher in the scenario). Scenario G also had high INFIT and OUTFIT values, indicating that it was problematic across all ability levels (INFIT) and, in addition, had at least one highly unexpected response from at least one respondent (OUTFIT). Based on the subsequent residual analysis, misfitting respondents were generally selecting a higher response category than they were expected to for scenario G, providing additional evidence that revisions to make G more difficult to respond to would be beneficial. Scenario D was revised to make it easier to endorse (meaning it is easier for a respondent to rate themselves as a higher skill level than the teacher in the scenario) in an effort to spread scenarios B and D apart in the person-item variable map, as these two scenarios were clustered together and did not capture the hypothesized continuum well enough.

The pilot administration had responses from 169 respondents from Amazon Mechanical Turk (MTurk) who were employed in the education industry. Of the 169 responses, one respondent abandoned the survey, resulting in a dataset of 168 respondents. For the pilot administration, the M-C Form C (Reynolds, 1982), a short form of the Marlowe-Crowne social desirability scale, was included to assess the extent of social desirability bias in the responses. The revisions made to the scenarios after the pre-pilot administration were generally successful and resulted in the intended changes. There was more variation in the response choices for the pilot than in the pre-pilot: the probability of responding with a 3 was lower and the CCCs fit the rating scale structure better than in the pre-pilot, although the probability of responding with a 5 was still relatively high. In addition, scenario D was located as hypothesized below scenario B in the pilot administration data. Finally, scenario G had values under 1.4 for INFIT and OUTFIT on the pilot, indicating that the misfit identified for scenario G on the pre-pilot was not present in the revised scenario on the pilot administration.

Although there were improvements to the scenarios from the pre-pilot to the pilot administration, the data analysis from the pilot identified some additional areas for revisions. Specifically, scenarios C and G (the scenarios hypothesized to be the most difficult to endorse) had correlations with the M-C Form C (Reynolds, 1982) that were statistically significantly different from zero ($p<.05$), indicating that there may have been some social desirability bias in the responses to these scenarios. These two scenarios were constructed as the most difficult scenarios for respondents to respond to, and it is possible that respondents could identify this design feature, prompting them to respond in the socially desirable way rather than in a way that reflected their lived experiences. When looking at misfitting individuals for scenarios C and G, the majority had higher than expected responses to scenarios C and G, indicating that it was possible that these respondents were providing the socially desirable response. Additionally, scenarios P and F did not follow their hypothesized order in the data from the pilot administration: scenario F was located alongside scenario E in the person-item variable map, although scenario F should have been the easiest to respond to and should have been located at the bottom of the person-item variable map. Misfitting individuals for scenario F generally provided lower than expected responses. Scenario P was located with scenario B in the person-item variable map, although scenario P was hypothesized to be more difficult for respondents to endorse than B, which would place it above B in the person-item variable map. Finally, scenarios B, D, E, and F remained clustered together on the person-item variable map and could benefit from some greater separation in their difficult levels. In contrast to these remaining item development challenges, the principal components analysis on the residuals from the pilot administration provided some positive evidence of unidimensionality in the residual data, meeting one of the critical statistical assumptions of the Rasch model.

Based on the data analysis from the pilot administration, the following revisions to the scenarios were made. Revisions were made to six of the eight scenarios. The facet compositions for scenarios C and F were changed by swapping the facets in scenario C with the facets in scenario F. This was done to make it less obvious that the teacher in scenario C has the highest skillset by including facets in scenario C that were not the first three in the inquiry cycle. Scenario G was revised to make it slightly more difficult to score high on by changing the language so it sounds less positive. Scenarios B and D were revised to be slightly easier to score high on in order to separate their scenario locations from the others. Scenario P was revised to be slightly harder by altering its language. Finally, language containing the word "hypotheses" was added to the scenarios that contain facet G (specifically, scenarios C, G, and B) based on the language in the inquiry cycle described by Hamilton et al. (2009).

The full administration had responses from 287 teachers in a public school district in a medium-sized city in Massachusetts. The final dataset included responses from 203 individuals after removing responses from individuals who did not provide consent, did not respond to any survey items, answered only the practice item, or answered only the practice item and one or two scenarios. The revisions made to the scenarios after the pilot administration were generally successful and resulted in the intended changes. Although the social desirability scale was not included on the full administration, the revisions to scenarios C and G appear to have potentially reduced the social desirability bias in the responses that was identified in the pilot administration. The examination of misfitting individuals on the full administration shows few with misfitting responses to scenario C. For misfitting individuals on scenario G, the responses show individuals with either lower than expected or higher than expected responses, indicating that there may be some social desirability bias remaining in the responses to scenario G, but the misfit in responses may also be due to survey fatigue as scenario G is the final scenario on the survey. Scenarios F and P are now in their hypothesized order on the person-item variable map and scenarios E and B have been separated further apart than they were in the pilot data. Scenarios D and F are still closely clustered together at the bottom of the scale, but they are now in their hypothesized order. The person-item variable map represents a ladder-like continuum as intended as part of the Rasch model, although there is a larger gap between scenarios A and G (or raw scores 21-27) than between other scenarios.

The data analysis from the full administration also indicated that the probability of responding with a 5 (Teacher X is much less skilled than me) is lower than on the pre-pilot and pilot administrations, which provides some evidence of successful revisions to make the scenarios more difficult to endorse. Finally, the principal components analysis on the residuals from the full administration again provides positive evidence that there is no unidentified and unintended construct in the scenarios; the primary influence underlying the responses to the scenarios is an individual's experience using data to inform classroom instruction.

Discussion of Results

The results from the full administration of the UDII scale provide empirical evidence that the scenarios fall in the intended order on the hypothesized hierarchical continuum of using data to inform classroom instruction. This means that the person-item variable map with interpretative scores presented in Chapter 4 supports content and construct valid detailed descriptions of individuals at particular locations along the construct's continuum. The inclusion of raw scores on this interpretative map allows users of the UDII scale to identify where individual respondents fall on the scale's continuum and apply a description of an individual at that raw score location. This description allows the users of this instrument to understand an individual's current skillset with using data to inform classroom instruction in an effort to provide additional support or professional development targeted for that individual. Additionally, because teams of teachers often engage in this work as a group, the interpretative person-item variable map can also be used with teams of teachers to identify the skillsets of individual team members and identify any potential areas where that team may need additional support or professional development. In addition, this map can be used to identify change in teachers' skillsets after professional development or support if administered both before and after the professional development/additional support is provided.

Importantly, the results from the UDII scale should not be used to evaluate teachers for professional status or as part of their job evaluation, as this instrument is not designed for these evaluative purposes. It is meant to be used to support individual teachers and teams of teachers by providing differentiated support and professional development opportunities for teachers based on their current skillset with using data to inform classroom instruction, based on their responses to this survey instrument. While the UDII scale should not be used to evaluate individual teachers, it could be used to evaluate specific programs or professional development sessions that are put in place at a school or district that are focused on supporting and training teachers to use data to inform their classroom instruction. The instrument could be used to measure any potential change from a specific program or professional development session designed for this purpose.

Limitations

There are some specific limitations to the UDII scale in its current form. First, only one public school district was included in the full administration due to the ongoing COVID-19 pandemic. The inclusion of only one public school district limits the generalizability of the results to other public school districts with different characteristics than the one included in this research. Future research could include the administration of this survey in other types of districts to enhance generalizability. In addition, there is no example scenario for respondents with an average score of one in the interpretative person-item variable map, as this response was not selected very often in the full administration. This limits the usability of the interpretative person-item variable map for schools and districts when they have respondents who have an average score of one. However, the fact that an average score of one is located at the bottom of the scale may help with interpretation of the skillset of an individual with an average score of one. This individual would benefit from support with all of the skills identified as part of the construct of using data to inform classroom instruction and described in the example scenarios for individuals with an average score of 2, 3, and 4, as individuals with an average score of one rated their skillset as much lower than or lower than the teachers in all of the scenarios. There is also no example scenario for respondents with an average score of five in the interpretative person-item variable map, but as an average score of five is at the top of the scale, individuals with an average score of five are likely to have strong skills for using data to inform classroom instruction.

Additionally, the response rate for the full administration was 17% (203 responses out of 1,196 emails sent). This response rate is likely artificially low, as the list of teacher email addresses that was provided by the district included teachers out on leave, many of whom were not checking email, as well as some individuals who did not identify as teachers who would use data for the purposes of this research. Without the inclusion of these individuals in the total, the response rate would likely be slightly higher. Regardless, it is possible that those who responded to the survey are different in some way from those who did not respond. To check the representativeness of the responses, the demographics of respondents were compared to the demographics of teachers in the district. A slightly higher percentage of females responded to the survey than employed by the district (81% of respondents compared to 73% of all teachers). A slightly higher percentage of teachers who identify as Asian (8% of respondents compared to 5% of all teachers) and teachers who identify as more than one race (4% of respondents compared to 1% of all teachers) responded to the survey than employed by the district, while a slightly lower percentage of teachers who identify as African American/Black (0.5% of respondents compared to 3% of all teachers) and Hispanic/Latinx (0.5% of respondents compared to 4% of all teachers) responded to the survey than employed by the district. It is difficult to estimate the representativeness of the survey respondents based on their teaching role, as many teachers self-identified as an "other" category on the survey (28%), but are not classified in that way in publicly available data on teacher roles for the district. The data focused on the representativeness of the survey respondents indicate that the results may not reflect the district as a whole in terms of racial/ethnic identity and primary teaching role. Future research focused on expanding this instrument into other types of districts could also consider increasing the representativeness of the sample of respondents.

Finally, this study did not employ differential item functioning analysis (or DIF) to investigate whether any differences to responses exist based on teacher role. This analysis should be considered in future work.

Future Directions for Research

There are multiple possibilities for future directions for research. As described in the limitations section, there are no example scenarios for an individual with an average score of one or five. Future research could focus on adapting the current scenarios or adding new scenarios to the bottom and top of the scale to provide example scenarios for these average scores. Another line of research could focus on different combinations of facets and levels within scenarios. As mentioned earlier, the facets and levels within each scenario were chosen based on professional experience; however, other combinations of facets and levels within scenarios are plausible. Research could focus on creating scenarios with different facet level combinations, which could provide a more nuanced description of individuals within each score range. Future research could also investigate the use of more facets within each scenario (such as the inclusion of four or five facets per scenario) to determine if the cognitive load of these longer scenarios is too much for respondents or is acceptable; as a reminder, only three facets were included in the current research based on expert feedback that four facets within a scenario made the scenarios too lengthy and difficult to read and interpret in the context of comparison to an individual's own experience. This feedback, however, could be investigated through future research. Additionally, future research could implement this survey in other types of public school districts to increase generalizability and evaluate the performance of the instrument in different district types, such as a small suburban district, a large urban district, or a rural district, and with a larger sample of teacher roles and racial/ethnic identities.

Future research could also focus on the use of the UDII scale for the evaluation of professional development programs in schools. The current research focused solely on the development of the survey, while the application of the survey was outside of the scope of the current research. Future research could utilize this scale to evaluate professional development programs or supports put in place for individual teachers or teams of teachers. Future research could also focus on how to use the results of the survey within schools to support individual teachers or teams of teachers. This line of research could focus on questions such as how to support teachers in a particular score range on the scale to help improve their skills for using data to inform classroom instruction, or focus on what specific structures or supports are successful in supporting individuals or teams of teachers within specific score ranges on the scale. Additionally, future research could focus on how individual schools or districts use the results of the survey to support individual teachers or teams of teachers and the effectiveness of the strategies that schools or districts use for this support.

Another line of future research could target the specific types of data that are the focus when teachers respond to this survey. In the current research, "data" in the context of the scenario items are described to respondents as including all types of data; respondents are instructed to think about all types of data when responding to the scenario items. As mentioned in the description of the pre-dissertation instrument, some teachers felt unable to respond to the pre-dissertation instrument with such a broad definition of data. Although no responses were received for the current instrument that indicated that the lack of a specific type of data was an issue in responding, it is possible that some teachers did not respond to the current instrument given this lack of a specific description of data. Future research could focus on whether there are differences in responses from teachers when they are prompted to think about specific types of data when responding to the survey instrument. This line of research would align with the Teacher Data Use Survey (Wayman et al., 2016), where the district or school using that survey can pick their own type of data as the focus for teachers when completing the survey.

Future research could also explore the use of the UDII scale with students after teaching students how to use their own data in an inquiry cycle or with teachers after having teachers teach students how to use their own data. Teaching students how to use their own data to set learning goals is included as one of Hamilton et al.'s (2009) recommendations in their practice guide on using student data for decision making. Jimerson et al. (2016) describe this as ‘student-involved data use’ (SIDU) and note that although this is a relatively new area of research and additional research should be undertaken to identify how this can be constructively utilized in schools, teachers who engage in this type of work do so because they believe it is beneficial for student learning and reflection (Jimerson et al., 2019). Schildkamp (2019) also notes that students can use their own data to improve their own learning.

Finally, future research could focus on how central office administrators use data, given that this is a focus of federal policies such as ESSA, and that Honig & Venkateswaran (2012) indicate that there are relationships and possible dependencies between central offices and schools related to data use; specifically, central office staff play a role in the use of data in schools and there is some evidence of schools playing a role in the use of data in central offices (Honig & Venkateswaran, 2012). Coburn et al. (2020) also found varied organizational routines among district leaders related to using data and research to make decisions, and noted that the type of these routines determine the type of data or information that these leaders used to make decisions.

Implications and Conclusion

The results from the full administration of the UDII scale provide empirical evidence of the hierarchical construct of using data to inform classroom instruction, while also demonstrating the utility of the Rasch/Guttman Scenario (RGS) approach for the development of this type of instrument. While the results of this research culminate in a survey instrument that can be utilized to measure teachers’ use of data to inform classroom instruction, there are additional avenues for future research, as described above, as well as additional areas that schools and districts may want to consider when using survey results, as the use of data does not occur in a vacuum and external factors can interact with this process (Coburn & Turner, 2011). These considerations include the impact of individual and collective biases on the use of data to inform classroom instruction and the ways in which teachers collect and interpret data, as well as data use for equity. Teacher beliefs about data use and efforts to build a culture of data use within schools can also be considered when using the survey instrument. Finally, critical data-driven decision making can be explored when using the results of this survey instrument. These considerations are briefly discussed next, although they are outside the scope of the current work.

The impact of individual and collective biases on the use of data to inform classroom instruction should be considered by schools and districts that use this survey instrument. These biases are often unconscious and many districts and schools are currently engaged in professional development work focused on identifying and naming these biases. In addition, confirmation bias (meaning that data is interpreted to confirm teachers' hypotheses about student competency rather than challenge them) frequently exists (Vanlommel & Schildkamp, 2019) and teachers may use heuristics (mental shortcuts informed by beliefs and prior experience) to come to quicker conclusions, which may not be valid if the heuristics are biased (Vanlommel & Schildkamp, 2019). Given that sense-making with data inherently involves filtering data through an individual's own lens and experience, which may result in different decisions for different people (Schildkamp, 2019; Vanlommel & Schildkamp, 2019), and that people often choose particular pieces of data "to negotiate arguments about the nature of problems as well as potential solutions" (Spillane, 2012, p. 114) when using data to inform instruction, identifying, naming, and thinking about how these individual and collective biases impact this work is important.

Engaging in this work as part of a team has the benefit of bringing multiple lenses and experiences that can be utilized to select and analyze data; however, even within a team, these lenses and experiences can affect interpretation, especially if individuals on the team have similar lenses and experiences, or if there are power inequalities within that team (Coburn & Turner, 2011). Coburn and Turner (2011) note that individual characteristics, along with social interaction dynamics present in teams and schools, influence both the process and the decisions that are made as part of this work.

Given that data use is an interpretative process (Coburn & Turner, 2011), the ways in which teachers collect and interpret data can influence the process of using data to inform classroom instruction. Vanlommel and Schildkamp (2019) note that teachers collect and interpret data using both data-based decision making processes and intuitive processes, which differ in how data are collected, analyzed, and interpreted; specifically, as described earlier in this dissertation, data-based decision making focuses on systematic procedures (such as an inquiry cycle), while intuitive processes focus on data that comes from spontaneous, non-systematic collection (such as student observations and group discussions) and emphasizes the personal knowledge of experts (such as teachers) in this process. Although these two processes are juxtaposed in this description, Vanlommel & Schildkamp (2019) note that these two processes can be and often are complementary and can work together.

The use of data triangulation (using multiple data sources), identifying and testing alternative hypotheses, and identifying pre-set criteria for a decision can help ameliorate some of the issues that arise from intuitive processes in the data interpretation stage of the process of using data, especially in a high-stakes decision process (Vanlommel & Schildkamp, 2019). Vanlommel and Schildkamp (2019) state that a focus solely on using data to improve decision making is not enough; focus should also be placed on changing beliefs about student abilities (Vanlommel & Schildkamp, 2019), which can be related to unconscious biases that teachers may hold. This also provides some evidence for the value of engaging in this work as a team, as it provides the opportunity to share conclusions and vet them as a group, and helps to make this process more transparent and public. It also provides some evidence for the value of professional development and training to provide teachers with the tools to engage in the systematic process of using data to inform classroom instruction.

The use of data for equity is currently an important topic in literature and in practice, and is also important for schools and districts that are engaging in this work to consider. Datnow and Park (2018) note that the following areas are important to consider from an equity lens: data meetings that focus on instructional improvement and administrative compliance, a focus on large scale assessment data and multiple types of data, a focus on small groups of students versus examining data for all students, the use of data to confirm assumptions versus using data to challenge beliefs, and the use of data for tracking or flexible grouping (Datnow & Park, 2018). Although outside the scope of the current work, schools and districts who plan to use the UDII scale should consider their work with using data to inform classroom instruction from an equity lens when using the results of the survey to support individual teachers and teams of teachers.

Additional factors that may be important for schools and districts to consider when thinking about using data to inform classroom instruction, but also outside the scope of this dissertation, include teachers' individual beliefs about data use and building a culture of data use within buildings and districts. Prenger and Schildkamp (2018) note that psychological characteristics of teachers, including perceived control, attitude (which can also be conceptualized as beliefs about a specific behavior), and intention regarding data use can influence this process. Specifically, Prenger and Schildkamp (2018) note that teachers' feelings and emotions about using data, as well as their beliefs on whether data use can be effective, can play a role in teacher data use. In addition, teacher beliefs about others' expectations regarding data use can influence teacher intentions to use data, as can perceived control (i.e., the extent to which teachers feel they have autonomy in this practice) (Prenger & Schildkamp, 2018). Coburn and Turner (2011) note that teacher beliefs play a role in both the data that teachers utilize and the decisions that they make based on that data. Future research could examine how these psychological characteristics can impact teachers' responses to this survey instrument, especially given that psychological characteristics can be changed with specific interventions (Prenger & Schildkamp, 2018).

When building a culture of data use within buildings and districts, Schildkamp et al. (2019) suggest five key building blocks that they describe as transformational leadership components that school leaders can use to help enable data use in data teams within their schools, which can be used in conjunction with the results from the UDII scale. They note that all five of these key building blocks are necessary for sustainable data use within buildings (Schildkamp et al., 2019). These five key building blocks include establishing visions, norms, and goals; providing individualized support (an example given is emotional support); intellectual stimulation (such as knowledge sharing and ensuring autonomy); creating a climate for data use (focused on improvement, not accountability); and networking to connect the school organization (Schildkamp et al., 2019). As previously mentioned, the UDII scale can be utilized to provide individualized support to teachers and group of teachers engaged in this type of work, and can be a tool for this type of building block for leaders building sustainable data use within their buildings. Although outside the scope of this dissertation, the five key areas identified by Schildkamp et al. (2019), and especially the use of the UDII scale to provide individualized support to teachers and teams of teachers, can be explored and implemented by leaders who want to engage in this type of work in their school buildings.

Finally, critical data-driven decision making is emerging as a way to focus on equity when utilizing data-driven decision making processes within schools (Dodman et al., 2021). Dodman et al. (2021) state that although many believe that the steps in the DDDM process are objective and neutral, in reality, they are couched in historical and current forces. If these historical and current forces are acknowledged and confronted, Dodman et al. (2021) note that the process of using data can be used to identify school practices and policies that perpetuate inequities. These processes of critical data-driven decision making can help teachers "identify systemic inequity in their schools, use data to reflect on policies and practices, and take action against counter-productive school reforms" (Dodman et al., 2021, p. 2). A key point is that data-driven decision making can often involve deficit logic, where roots of educational inequities are ascribed to the communities that are experiencing them, rather than historical and sociocultural factors (Dodman et al., 2021). This can keep teachers from interrogating the structural disadvantages that their students may face or keep them from evaluating whether their goals for student learning may perpetuate inequities, leading to increases in opportunity gaps for students (Dodman et al., 2021).

Dodman et al. (2021) note that critical data-driven decision making is an expansion of DDDM, focusing on the immediate instructional needs of students that specifically considers the systemic influences that affect the learning of students, with a focus on identifying ways to increase equity within schools. A key point is that this process focuses on the systems and structures that affect students, rather than focusing solely on outcomes (Dodman et al., 2021). Dodman et al. (2021) write, "Critical data-driven decision making thus shifts the question driving data use away from "What will close these gaps" to "What will increase and deepen equity in our school?" (p. 5). Focusing on critical data-driven decision making in professional development or teacher support can be done in conjunction with the use of the UDII scale developed in this dissertation.

There are many possibilities for future research and considerations to take into account when measuring teachers' use of data to inform classroom instruction. The current research adds to the body of literature focused on teachers' use of data to inform classroom instruction/data-driven decision making, while also contributing an additional instrument designed to measure this construct. In addition, this work adds to the growing body of research utilizing the Rasch/Guttman Scenario (RGS) approach, and provides additional support for the approach's utility in measuring a variety of constructs. The UDII scale can be utilized by schools and districts interested in measuring their teachers' use of data to inform classroom instruction with the intention of providing differentiated support and professional support to individuals and teams of teachers to aid them in this important work.

References

Andrich, D. (1996). Measurement criteria for choosing among models with graded responses. In Von Eye & C. C. Clogg (Eds.), *Categorical variables in developmental research: Methods of analysis* (3-35). San Diego: Academic Press, Inc.

Anfara, V.A., & Donhost, M.J. (2010). What research says: Data-driven decision making. *Middle School Journal*, *42*(2), 56-63.

Antipkina, I., & Ludlow, L. (2020). Measuring parental involvement as a holistic concept using Rasch/Guttman scenario scales. *Journal of Psychoeducational Assessment, 38*(7), 846-865. https://doi.org/10.1177/0734282920903164

Báez Cruz, M.E. (2021). *Measuring teachers' promotion of sociocultural integration in K-12 schools in the United States: A scale development using Rasch/Guttman scenario methodology* (Publication No. 2541847754) [Doctoral dissertation, Boston College]. ProQuest Dissertations Publishing.

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

Bernhardt, V.L. (2004). *Data analysis for continuous school improvement* (2nd ed.). Eye on Education.

Bocala, C., Henry, S.F., Mundry, S., & Morgan, C. (2014). *Practitioner data use in schools: Workshop toolkit (REL 2015-043).* Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory, Northeast & Islands. http://ies.ed.gov/ncee/edlabs.

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

Boudett, K.P., & Steele, J.L. (Eds). (2007). *Data wise in action: Stories of schools using data to improve teaching and learning.* Harvard Education Press.

Brown, C., Schildkamp, K., & Hubers, M. (2017). Combining the best of two worlds: A conceptual proposal for evidence-informed school improvement. *Educational Research, 59*(2), 154-172.

Champion, R. (2017). Let's really talk about data: How to infuse meaningful data into daily decisions and conversations. *The Learning Professional, 38*(3), 56-60.

Chang, W-C.C. (2017). *Measuring the complexity of teachers' enactment of practice for equity: A Rasch model and facet theory-based approach* (Publication No. 10269093) [Doctoral dissertation, Boston College]. ProQuest Dissertations Publishing.

Chang, W-C., Ludlow, L.H., Grudnoff, A., Ell, F., Haigh, M., Hill, M., & Cochran-Smith, M. (2019). Measuring the complexity of teaching practice for equity: Development of a scenario-format scale. *Teaching and Teacher Education, 82*, 69-85.

Chen, W-H., Lenderking, W., Jin, Y., Wyrwich, K.W., Gelhorn, H., & Revicki, D.A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research, 23*(2), 485-493.

Coburn, C.E., Spillane, J.P., Bohannon, A.X., Allen, A-R., Ceperich, R., Beneke, A., Wong, L-S. (2020). *The role of organizational routines in research use in four large urban school districts* (Technical report No. 5). National Center for Research in Policy and Practice. https://files.eric.ed.gov/fulltext/ED612257.pdf

Coburn, C.E. & Turner, E.O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research and Perspectives, 9*(4), 173-206.

Coburn, C.E. & Turner, E.O. (2012). The practice of data use: An introduction. *American Journal of Education, 118*(2), 99-111.

Crowne, D.P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*(4), 349-354.

Dana, N.F. & Yendol-Silva, D. (2003). *The reflective educator's guide to classroom research: Learning to teach and teaching to learn through practitioner inquiry*. Corwin Press, Inc.

Datnow, A. (2011). Collaboration and contrived collegiality: Revisting Hargreaves in the age of accountability. *Journal of Educational Change, 12*(2), 147-158.

Datnow, A., & Park, V. (2018). Opening or closing doors for students? Equity and data use in schools. *Journal of Educational Change, 19*, 131-152.

Deming, W.E. (2018). *The new economics: For industry, government, education* (3rd ed.). The MIT Press.

Dodman, S.L., Swalwell, K., DeMulder, E.K., View, J.L., & Stribling, S.M. (2021). Critical data-driven decision making: A conceptual model of data use for equity. *Teaching and Teacher Education, 99*, 1-12. https://doi.org/10.1016/j.tate.2020.103272

Doyle, D.P. (2003). Data-driven decision making: Is it the mantra of the month or does it have staying power? *T.H.E. Journal*. Accessed online at http://thejournal.com/Articles/2003/05/01/DataDriven-DecisionMaking.aspx

Dunn, K.E., Airola, D.T., Lo, W.J., & Garrison, M. (2013). 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*, 87-98.

Guttmann, R., & Greenbaum, C.W. (1998). Facet theory: Its development and current status. *European Psychologist, 3*(1), 13-36.

Hackett, P.M.W. (2014). *Facet theory and the mapping sentence: Evolving philosophy, use and application.* Basingstoke: Palgrave Macmillan.

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

Herman, J.L., & Haertel, E.H. (Eds). (2005). *Uses and misuses of data for educational accountability and improvement: The 104th yearbook of the National Society for the Study of Education part 2*. Malden, MA: Blackwell Publishing.

Hirsh, S., & Crow, T. (2018). *Becoming a learning team: A guide to a teacher-led cycle of continuous improvement* (2nd ed.). Learning Forward.

Honig, M.I., & Venkateswaran, N. (2012). School-central office relationships in evidence use: Understanding evidence use as a systems problem. *American Journal of Education, 118*, 199-222.

Horn, I.S., Kane, B.D., & 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.

Jimerson, J.B. (2013). Thinking about data: Exploring the development of mental models for "data use" among teachers and school leaders. *Studies in Educational Evaluation*. http://dx.doi.org/10.1016/j.stueduc.2013.10.010

Jimerson, J.B., Cho, V., Scroggins, K.A., Balial, R., & Robinson, R.R. (2019). How and why teachers engage students with data. *Educational Studies, 45*(6), 667-691.

Jimerson, J.B., Cho, V., & Wayman, J.C. (2016). Student-involved data use: Teacher practices and considerations for professional learning. *Teaching and Teacher Education, 60*, 413-424.

Kekahio, W., & Baker, M. (2013). *Five steps for structuring data-informed conversations and action in education* (REL 2013-001). Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory, Pacific. http://eric.ed.gov/?id=ED544201.

Linacre, J.M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions, 7*(4), 328. https://www.rasch.org/rmt/rmt74m.htm

Linacre, J.M. (2002a). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J.M. (2002b). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878. https://www.rasch.org/rmt/rmt162f.htm

Little, J.W. (2012). Understanding data use practice among teachers: The contribution of micro-process studies. *American Journal of Education, 118*(2), 143-166.

Lord, F.M. (1980). *Applications of item response theory to practical testing problems.* Lawrence Erlbaum Associates, Inc., Publishers.

Ludlow, L.H. (1983). *The analysis of Rasch model residuals* [Unpublished doctoral dissertation]. University of Chicago.

Ludlow, L.H., Anghel, E., Szendey, O., O'Keefe, T., Howell, B., Matz-Costa, C., & Braun, H. (2020). The Boston College living a life of meaning and purpose (BC-LAMP) portfolio: An application of Rasch/Guttman scenario methodology. *Journal of Applied Measurement, 21*(2), 134-153.

Ludlow, L.H., Baez-Cruz, M., Chang, W-C., & Reynolds, K. (2020). Rasch/Guttman scenario scales: a methodological framework. *Journal of Applied Measurement, 21*(4), 361-378.

Ludlow, L.H., Enterline, S.E., & Cochran-Smith, M. (2008). Learning to teach for social justice-beliefs scale: An application of Rasch measurement principles. *Measurement and Evaluation in Counseling and Development, 40*(4), 194-214.

Ludlow, L.H., Matz-Costa, C., Johnson, C., Brown, M., Besen, E., & James, J.B. (2014). Measuring engagement in later life activities: Rasch-based scenario scales for work, caregiving, informal helping, and volunteering. *Measurement and Evaluation in Counseling and Development*, *47*(2), 127-149.

Ludlow, L.H., Matz-Costa, C., & Klein, K. (2019). Enhancement and validation of the productive engagement portfolio-scenario (PEP-S8) scales. *Measurement and Evaluation in Counseling and Development, 52*(1), 15-37. https://doi.org/10.1080/07481756.2018.1497430

Ludlow, L.H., Reynolds, K., Baez-Cruz, M., & Chang, W-C. (2021). Enhancing the interpretation of scores through Rasch-based scenario-style items. In A.G. Harbaugh (Ed.), *Basic elements of survey research in education: Addressing the problems your advisor never told you about* (pp. 675-720). Information Age Publishing.

Mandinach, E.B. (2012). A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist, 47*(2), 71-85.

Mandinach, E.B., & Gummer, E.S. (2016a). *Data literacy for educators: Making it count in teacher preparation and practice.* Teachers College Press and WestEd.

Mandinach, E.B., & Gummer, E.S. (2016b). Every teacher should succeed with data literacy. *Phi Delta Kappan, 97*(8), 43-46.

Marsh, J.A., Pane, J.F., & Hamilton, L.S. (2006). *Making sense of data-driven decision making in education*. Santa Monica, CA: RAND Corporation.

Mason, S. (2002). *Turning data into knowledge: Lessons from six Milwaukee public schools*. Madison, WI: Wisconsin Center for Education Research.

National Council of Teachers of Mathematics. (2010). *How can teachers and schools use data effectively?* Reston, VA: M. Schleppenbach.

National Forum on Education Statistics. (2012). *Forum guide to taking action with education data* (NFES 2013-801). https://nces.ed.gov/pubs2013/2013801.pdf

Neuman, S.B. (2016). Code red: The danger of data-driven instruction. *Educational Leadership, 74*(3), 24-29.

Pella, S. (2012). What should count as data for data-driven instruction? Toward contextualized data-inquiry models for teacher education and professional development. *Middle Grades Research Journal, 7*(1), 57-75.

Peoples, S.M., O'Dwyer, L.M., Wang, Y., Brown, J.J., & Rosca, C.V. (2014). Development and application of the elementary school science classroom environment scale (ESSCES): Measuring student perceptions of constructivism within the science classroom. *Learning Environments Research, 17*(1)*,* 49-73.

Prenger, R., & Schildkamp, K. (2018). Data-based decision making for teacher and student learning: A psychological perspective on the role of the teacher. *Educational Psychology, 38*(6), 734-752.

Schildkamp, K. (2019). Data-based decision making for school improvement: Research insights and gaps. *Educational Research, 61*(3), 257-273.

Schildkamp, K., Poortman, C.L., Ebbeler, J., & Pieters, J.M. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making. *Journal of Educational Change, 20*, 283-325.

Spillane, J.P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education, 118*(2), 113-141.

Rasch, G. (1966). An individualistic approach to item analysis. In P.F. Lazarsfeld & N.W. Henry (Eds.), *Readings in Mathematical Social Science* (84-108). Chicago: Science Research Associates.

Reynolds, K.A. (2020). *Measuring students' perceptions of faculty availability outside of class using Rasch/Guttman scenario scales.* [Unpublished doctoral dissertation dissertation]. Boston College.

Reynolds, W.M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology, 38*(1), 119-125.

Tichnor-Wagner, A., Wachen, J., Cannata, M., & Cohen-Vogel, L. (2017). Continuous improvement in the public school context: Understanding how educators respond to plan-do-study-act cycles. *Journal of Educational Change, 18*, 465-494.

Vanlommel, K., & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision making? *American Educational Research Journal, 56*(3), 792-821.

Walker, D.A., Reeves, T.D., & Smith, T.J. Confirmation of the data-driven decision-making efficacy and anxiety inventory's score factor structure among teachers. *Journal of Psychoeducational Assessment*. Prepublished December, 1, 2016, DOI: 10.1177/0734282916682905

Wayman, J.C., Wilkerson, S.B., Cho, V., Mandinach, E.B., & Supovitz, J.A. (2016). *Guide to using the Teacher Data Use Survey* (REL 2017-166). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia.

Wright, B.D. (1977). Solving measurement problems with the Rasch model. *Journal of Education Measurement, 14*(2), 97-116.

Wright, B.D., & Masters, G.N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.

Appendix A: Expert Feedback Recruitment Email

Dear MESA students,

I hope you are having a wonderful summer. I am hoping you have 15-20 minutes to help provide some expert feedback on the first draft of my survey instrument for my dissertation. It focuses on the construct of using data to inform classroom instruction and was developed using Rasch/Guttman Scenario (RGS) methodology). The intended audience is teachers.

The attachment includes a brief description of the construct, descriptions of teachers at high, medium, and low levels of the construct, and the items themselves. I would appreciate your feedback on the following:

1. Are the instructions and practice item clear?
2. The wording of the scenarios: is any wording confusing? Are any scenarios difficult to answer because it is unclear what is being described in the scenario?
3. The length of the scenarios: Are they too lengthy to respond to?
4. The response options: Are they clear? Do they make sense in conjunction with the scenarios?
5. Coverage of the construct in the scenarios: Is anything missing from the group of scenarios as a whole (based on my definition of the construct)?
6. Any other feedback you may have.

I really appreciate any feedback that you can provide! Feedback can be sent by email. I am hoping to receive feedback by next Wednesday, August 18th.

Best,

Katy

Appendix B: Pre-Pilot (convenience sample) Recruitment Email

Dear [name],

I'm writing to ask for your help in a research project that I am conducting for my doctoral dissertation at Boston College. I am interested in measuring teachers' skills using data to inform their classroom instruction and I have developed a survey to measure this.

I'm hoping you can help me gather feedback by taking this survey. Your responses will help me to make any necessary changes to the survey before the full administration to a large sample of public school teachers. If you have any comments or feedback on the instructions or survey itself, please send me an email with that feedback.

I estimate that the survey should take 15-20 minutes to complete.

To access the survey, please click on this link: [link}

You can also copy and paste the link into your internet browser.

You are free to withdraw from the survey at any time or skip questions for any reason. There are no penalties for withdrawing or skipping questions.

Thank you in advance for your help!

Best,

Katy

Appendix C: Pre-Pilot Survey

Instructions: These scenarios describe different teachers' skillsets when using data to inform their classroom instruction. Data refers to a wide variety of data sources in these scenarios, not solely assessment data.

- Reflect on your own skills using data to inform your classroom instruction while reading each scenario.
- Compare your own skillset to the teacher's experience in the scenario.
- Choose the response that most accurately reflects your comparison.

Practice Item: As the scenario-type survey may be a new experience, this practice item allows you to practice engaging with a scenario. Please read the following practice scenario and compare your own experience to Teacher P.

Teacher P sometimes has difficulty pinpointing and writing clearly defined questions that describe their thoughts about student learning. They may need some support systematically gathering data related to these questions and comparing multiple data sources. This teacher sometimes has trouble identifying interventions to meet their goals related to student learning or planning to assess progress towards these goals on their own and looks to others for support.

How do your skills using data to inform classroom instruction compare to Teacher P?

- Teacher P is much less skilled than me
- Teacher P is slightly less skilled than me
- Teacher P is just like me
- Teacher P is slightly more skilled than me
- Teacher P is much more skilled than me

1. Teacher A can successfully identify interventions to meet their student learning goals and plan to assess progress towards these goals. However, they sometimes struggle to implement their chosen intervention with fidelity and to collect data about student progress towards their goals during the intervention. Teacher A benefits from support with examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.

How do your skills using data to inform classroom instruction compare to Teacher A?

- Teacher A is much less skilled than me
- Teacher A is slightly less skilled than me
- Teacher A is just like me
- Teacher A is slightly more skilled than me

- Teacher A is much more skilled than me

2. Teacher B may need some support thinking critically about and analyzing data that helps answer their questions about student learning. Once they complete their data analysis focused on these questions, they sometimes have trouble extracting meaning from this analysis and often do not seek the support of others to validate their interpretations. When they choose an intervention for student learning, Teacher B struggles to both plan for student progress towards their student learning goals and to share progress with colleagues during their implementation.

How do your skills using data to inform classroom instruction compare to Teacher B?

- Teacher B is much less skilled than me
- Teacher B is slightly less skilled than me
- Teacher B is just like me
- Teacher B is slightly more skilled than me
- Teacher B is much more skilled than me

3. Teacher C is successful generating and writing clearly defined questions that clarify their thoughts on student learning. They do not need support understanding what data can help answer their questions or identifying different techniques to display their data. This teacher can independently gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.

How do your skills using data to inform classroom instruction compare to Teacher C?

- Teacher C is much less skilled than me
- Teacher C is slightly less skilled than me
- Teacher C is just like me
- Teacher C is slightly more skilled than me
- Teacher C is much more skilled than me

4. Teacher D sometimes has difficulty narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they require support to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D often holds back on sharing progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.

How do your skills using data to inform classroom instruction compare to Teacher D?

- Teacher D is much less skilled than me
- Teacher D is slightly less skilled than me
- Teacher D is just like me

- Teacher D is slightly more skilled than me
- Teacher D is much more skilled than me

5. Teacher E occasionally needs help distilling their thoughts and wonderings about student learning into clearly defined questions. They often need support understanding what data can help answer these questions and organizing their data clearly once they have identified it. This teacher occasionally struggles with analyzing the implementation of the intervention they have chosen for fidelity and identifying their next steps once the intervention is complete.

How do your skills using data to inform classroom instruction compare to Teacher E?

- Teacher E is much less skilled than me
- Teacher E is slightly less skilled than me
- Teacher E is just like me
- Teacher E is slightly more skilled than me
- Teacher E is much more skilled than me

6. Teacher F struggles to understand what data can answer their questions about student learning and to organize that data clearly. This teacher has trouble interpreting results from their data analysis related to their questions about student learning and struggles to write clear and aligned goals for student learning based on their analysis. Once they have implemented a student learning intervention, they need support to examine student outcome data collected from this intervention and to identify their next steps.

How do your skills using data to inform classroom instruction compare to Teacher F?

- Teacher F is much less skilled than me
- Teacher F is slightly less skilled than me
- Teacher F is just like me
- Teacher F is slightly more skilled than me
- Teacher F is much more skilled than me

7. Teacher G is able to think critically about data that can help answer their questions about student learning and successfully make observations about their data while comparing multiple data sources. They can also independently interpret meaning from their data analysis to develop student learning goals and consider all potential consequences of these goals. Teacher G may require some support to both identify interventions to help attain these goals and identify the data required to document progress towards these goals.

How do your skills using data to inform classroom instruction compare to Teacher G?

- Teacher G is much less skilled than me

- Teacher G is slightly less skilled than me
- Teacher G is just like me
- Teacher G is slightly more skilled than me
- Teacher G is much more skilled than me

Demographic questions:

1. Please select the gender with which you most identify.

- Female
- Male
- Nonbinary (neither female nor male)
- Prefer not to answer

2. Please select the race/ethnicity with which you most identify (you may select more than one).

- African American/Black
- Asian
- Hispanic/Latinx
- Native American/Alaskan Native
- Native Hawaiian/Pacific Islander
- White
- Other (please specify)
- Prefer not to answer

3. How many years of teaching experience do you have?

- 0-3 years
- 3-5 years
- 5-10 years
- 10-15 years
- More than 15 years

4. Please select your primary teaching role.

- Administrator
- Elementary homeroom teacher
- Elementary specialist teacher
- Secondary English/ELA teacher
- Secondary History/Social Studies teacher
- Secondary Math teacher
- Secondary Science teacher
- Secondary World Languages teacher
- Secondary specialist teacher

- Other (please specify)

Appendix D: Pilot Survey

Instructions: These scenarios describe different teachers' skillsets when using data to inform their classroom instruction. Data refers to a wide variety of data sources in these scenarios, not solely assessment data.

- Reflect on your own skills using data to inform your classroom instruction while reading each scenario.
- Compare your own skillset to the teacher's experience in the scenario.
- Choose the response that most accurately reflects your comparison.

Practice Item: As the scenario-type survey may be a new experience, this practice item allows you to practice engaging with a scenario. Please read the following practice scenario and compare your own experience to Teacher P.

Teacher P occasionally has difficulty pinpointing and writing clearly defined questions that describe their thoughts about student learning. Once they have defined their questions, they may need some support to systematically gather data related to these questions and compare multiple data sources. This teacher can sometimes benefit from help to identify interventions to meet their goals related to student learning and to plan to assess progress towards their goals.

How do your skills using data to inform classroom instruction compare to Teacher P?

- Teacher P is much less skilled than me
- Teacher P is slightly less skilled than me
- Teacher P is just like me
- Teacher P is slightly more skilled than me
- Teacher P is much more skilled than me

1. Teacher A can independently identify interventions to meet their student learning goals, while also planning to assess progress towards these goals. They sometimes have trouble implementing their chosen intervention as intended, and sometimes have difficulty collecting data about student progress towards their goals during the intervention. Teacher A finds support helpful when examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.

How do your skills using data to inform classroom instruction compare to Teacher A?

- Teacher A is much less skilled than me
- Teacher A is slightly less skilled than me
- Teacher A is just like me
- Teacher A is slightly more skilled than me
- Teacher A is much more skilled than me

2. Teacher B sometimes struggles to think critically about and analyze data to help answer their questions about student learning. Once they complete their data analysis focused on these questions, they may look for support to extract meaning from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher B has difficulty planning for student progress towards their student learning goals and often holds back on sharing progress with colleagues during their implementation.

How do your skills using data to inform classroom instruction compare to Teacher B?

- Teacher B is much less skilled than me
- Teacher B is slightly less skilled than me
- Teacher B is just like me
- Teacher B is slightly more skilled than me
- Teacher B is much more skilled than me

3. Teacher C is successful generating and writing clearly defined questions that clarify their thoughts on student learning. They do not need support understanding what data can help answer their questions or identifying different techniques to display their data. This teacher can independently gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.

How do your skills using data to inform classroom instruction compare to Teacher C?

- Teacher C is much less skilled than me
- Teacher C is slightly less skilled than me
- Teacher C is just like me
- Teacher C is slightly more skilled than me
- Teacher C is much more skilled than me

4. Teacher D sometimes needs help narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they need the support of others to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D rarely shares progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.

How do your skills using data to inform classroom instruction compare to Teacher D?

- Teacher D is much less skilled than me
- Teacher D is slightly less skilled than me
- Teacher D is just like me

- Teacher D is slightly more skilled than me
- Teacher D is much more skilled than me

5. Teacher E sometimes benefits from help to distill their thoughts and wonderings about student learning into clearly defined questions. Once they have defined their questions, this teacher occasionally seeks support to understand what data can help answer these questions and to organize this data clearly. This teacher may struggle to analyze how well the implementation of their intervention has gone and is unsure of their next steps once the intervention is complete.

How do your skills using data to inform classroom instruction compare to Teacher E?

- Teacher E is much less skilled than me
- Teacher E is slightly less skilled than me
- Teacher E is just like me
- Teacher E is slightly more skilled than me
- Teacher E is much more skilled than me

6. Teacher F struggles to understand what data can answer their questions about student learning and to organize that data clearly. This teacher has trouble interpreting results from their data analysis related to their questions about student learning and struggles to write clear and aligned goals for student learning based on their analysis. Once they have implemented a student learning intervention, they need support to examine student outcome data collected from this intervention and to identify their next steps.

How do your skills using data to inform classroom instruction compare to Teacher F?

- Teacher F is much less skilled than me
- Teacher F is slightly less skilled than me
- Teacher F is just like me
- Teacher F is slightly more skilled than me
- Teacher F is much more skilled than me

7. Teacher G can critically examine data to help answer their questions about student learning, while independently comparing multiple data sources to make observations about their data. Interpreting meaning from their data analysis to develop student learning goals is a strength of Teacher G's and they consider all potential consequences of these goals. Teacher G may require some support to both identify interventions to help attain these goals and identify the data required to document progress towards these goals.

How do your skills using data to inform classroom instruction compare to Teacher G?

- Teacher G is much less skilled than me
- Teacher G is slightly less skilled than me

- Teacher G is just like me
- Teacher G is slightly more skilled than me
- Teacher G is much more skilled than me

Social Desirability scale: M-C Form C (Reynolds, 1982) [response options are True/False; this title was not presented to respondents]

8. It is sometimes hard for me to go on with my work if I am not encouraged.

9. I sometimes feel resentful when I don't get my way.

10. On a few occasions, I have given up doing something because I thought too little of my ability.

11. There have been times when I felt like rebelling against people in authority even though I knew they were right.

12. No matter who I'm talking to, I'm always a good listener.

13. There have been occasions when I took advantage of someone.

14. I'm always willing to admit it when I make a mistake.

15. I sometimes try to get even rather than forgive and forget.

16. I am always courteous, even to people who are disagreeable.

17. I have never been irked when people expressed ideas very different from my own.

18. There have been times when I was quite jealous of the good fortune of others.

19. I am sometimes irritated by people who ask favors of me.

20. I have never deliberately said something that hurt someone's feelings.

Demographic questions:

1. Please select the gender with which you most identify.

- Female
- Male
- Nonbinary (neither female nor male)
- Prefer not to answer

2. Please select the race/ethnicity with which you most identify (you may select more than one).

- African American/Black

- Asian
- Hispanic/Latinx
- Native American/Alaskan Native
- Native Hawaiian/Pacific Islander
- White
- Other (please specify)
- Prefer not to answer

3. How many years of teaching experience do you have?

- 0-3 years
- 3-5 years
- 5-10 years
- 10-15 years
- More than 15 years

4. Please select your primary teaching role.

- Administrator
- Elementary homeroom teacher
- Elementary specialist teacher
- Secondary English/ELA teacher
- Secondary History/Social Studies teacher
- Secondary Math teacher
- Secondary Science teacher
- Secondary World Languages teacher
- Secondary specialist teacher
- Other (please specify)

Appendix E: Person-Response Table from the Pilot

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

    127  127                         .77     3.9   A    3.8
  OBSERVED: 1:  1    1    5    5    4     3    5    5
  EXPECTED:     3.9  3.5  3.9  2.8  4.0   3.8  3.9  3.2
Z-RESIDUAL:    -3   -2         2

    105  105                        1.08     3.0   B    3.0
  OBSERVED: 1:  3    5    2    5    4     5    2    5
  EXPECTED:     4.1  3.8  4.1  3.0  4.2   4.1  4.1  3.5
Z-RESIDUAL:              -2                   -2

     35  35                          .63     3.0   C    3.0
  OBSERVED: 1:  5    3    4    5    2     2    2    5
  EXPECTED:     3.8  3.4  3.8  2.6  3.9   3.7  3.7  3.0
Z-RESIDUAL:                    2   -2               2

      7  7                          -.83     2.8   D    2.9
  OBSERVED: 1:  1    5    2    2    2     4    1    1
  EXPECTED:     2.5  2.1  2.4  1.7  2.6   2.4  2.4  1.9
Z-RESIDUAL:          3

     29  29                         1.87     2.4   E    2.7
  OBSERVED: 1:  5    4    2    5    5     4    5    5
  EXPECTED:     4.6  4.4  4.5  3.8  4.6   4.5  4.5  4.1
Z-RESIDUAL:              -4

     80  80                         -.50     2.4   F    2.6
  OBSERVED: 1:  3    2    1    4    5     2    1    2
  EXPECTED:     2.7  2.4  2.7  1.8  2.9   2.7  2.7  2.1
Z-RESIDUAL:                    2    2

    141  141                         .49     2.4   G    2.3
  OBSERVED: 1:  3    5    3    2    5     1    3    5
  EXPECTED:     3.7  3.3  3.6  2.5  3.8   3.6  3.6  2.9
Z-RESIDUAL:                              -2         2

     36  36                          .63     2.3   H    2.3
  OBSERVED: 1:  5    4    3    2    5     3    1    5
  EXPECTED:     3.8  3.4  3.8  2.6  3.9   3.7  3.7  3.0
Z-RESIDUAL:                                   -2    2

     67  67                         -.83     1.9   I    2.2
  OBSERVED: 1:  1    2    2    3    2     1    3    4
  EXPECTED:     2.5  2.1  2.4  1.7  2.6   2.4  2.4  1.9
Z-RESIDUAL:                                         2

     89  89                          .77     2.1   J    2.0
  OBSERVED: 1:  5    3    3    5    4     3    5    1
  EXPECTED:     3.9  3.5  3.9  2.8  4.0   3.8  3.9  3.2
Z-RESIDUAL:                    2                   -2

    120  120                        -.50     2.0   K    2.1
  OBSERVED: 1:  3    5    2    1    1     3    2    3
  EXPECTED:     2.7  2.4  2.7  1.8  2.9   2.7  2.7  2.1
Z-RESIDUAL:          2

    121  121                        1.87     2.0   L    2.1
  OBSERVED: 1:  5    5    4    5    5     3    3    5
  EXPECTED:     4.6  4.4  4.5  3.8  4.6   4.5  4.5  4.1
Z-RESIDUAL:                              -2   -2

    106  106                         .21     1.9   M    2.0
  OBSERVED: 1:  4    4    3    4    2     1    3    4
  EXPECTED:     3.4  3.0  3.4  2.3  3.6   3.3  3.4  2.7
```

```
Z-RESIDUAL:                              -2

    132  132                      .35     1.9   N    1.9
  OBSERVED: 1:  3    4    2    4    4     4    1    4
  EXPECTED:     3.5  3.1  3.5  2.4  3.7   3.5  3.5  2.8
Z-RESIDUAL:                                   -2

     11  11                       .49     1.9   O    1.9
  OBSERVED: 1:  3    3    4    4    2     4    2    5
  EXPECTED:     3.7  3.3  3.6  2.5  3.8   3.6  3.6  2.9
Z-RESIDUAL:                                        2

     17  17                       .63     1.9   P    1.8
  OBSERVED: 1:  5    4    2    4    4     3    5    1
  EXPECTED:     3.8  3.4  3.8  2.6  3.9   3.7  3.7  3.0
Z-RESIDUAL:                                       -2

     39  39                      -.66     1.8   Q    1.8
  OBSERVED: 1:  2    4    1    2    4     1    2    3
  EXPECTED:     2.6  2.3  2.6  1.8  2.7   2.5  2.5  2.0
Z-RESIDUAL:          2

    167  167                      .35     1.6   R    1.8
  OBSERVED: 1:  4    2    4    5    3     2    4    2
  EXPECTED:     3.5  3.1  3.5  2.4  3.7   3.5  3.5  2.8
Z-RESIDUAL:                    2

     28  28                      2.14     1.6   S    1.8
  OBSERVED: 1:  4    5    4    5    5     5    3    5
  EXPECTED:     4.6  4.5  4.6  4.0  4.7   4.6  4.6  4.3
Z-RESIDUAL:                                   -2

      8  8                       -.35     1.5   T    1.7
  OBSERVED: 1:  3    2    3    4    1     3    2    3
  EXPECTED:     2.9  2.5  2.9  1.9  3.0   2.8  2.8  2.2
Z-RESIDUAL:                    2   -2

     87  87                      -.21     1.5   U    1.7
  OBSERVED: 1:  3    2    4    4    3     1    2    3
  EXPECTED:     3.0  2.6  3.0  2.0  3.2   2.9  3.0  2.3
Z-RESIDUAL:                    2

     13  13                       .35     1.7   V    1.6
  OBSERVED: 1:  4    2    5    2    5     4    1    3
  EXPECTED:     3.5  3.1  3.5  2.4  3.7   3.5  3.5  2.8
Z-RESIDUAL:                                   -2

     14  14                       .49     1.6   W    1.7
  OBSERVED: 1:  4    5    3    4    2     4    2    3
  EXPECTED:     3.7  3.3  3.6  2.5  3.8   3.6  3.6  2.9
Z-RESIDUAL:

     22  22                       .49     1.6   X    1.6
  OBSERVED: 1:  4    3    4    4    3     4    1    4
  EXPECTED:     3.7  3.3  3.6  2.5  3.8   3.6  3.6  2.9
Z-RESIDUAL:                                   -2

    136  136                      .35     1.6   Y    1.6
  OBSERVED: 1:  5    5    2    3    2     3    3    3
  EXPECTED:     3.5  3.1  3.5  2.4  3.7   3.5  3.5  2.8
Z-RESIDUAL:

    110  110                     -.07     1.4   Z    1.6
  OBSERVED: 1:  2    3    2    4    3     2    3    4
  EXPECTED:     3.2  2.7  3.1  2.1  3.3   3.1  3.1  2.4
Z-RESIDUAL:                    2

    128  128                      .21     1.6        1.5
  OBSERVED: 1:  5    4    3    2    3     1    3    4
```

```
  EXPECTED:     3.4  3.0  3.4  2.3  3.6   3.3  3.4  2.7
Z-RESIDUAL:                              -2

     18  18                          .63     1.5         1.6
  OBSERVED: 1:  4    3    2    4    4     5    2    4
  EXPECTED:     3.8  3.4  3.8  2.6  3.9   3.7  3.7  3.0
Z-RESIDUAL:

    123  123                         .21     1.5         1.5
  OBSERVED: 1:  2    4    5    1    4     4    4    1
  EXPECTED:     3.4  3.0  3.4  2.3  3.6   3.3  3.4  2.7
Z-RESIDUAL:

     79  79                         -.21     1.5         1.4
  OBSERVED: 1:  5    4    2    2    3     3    1    2
  EXPECTED:     3.0  2.6  3.0  2.0  3.2   2.9  3.0  2.3
Z-RESIDUAL:     2                              -2

    125  125                         .21     1.5         1.5
  OBSERVED: 1:  3    5    3    1    4     2    5    2
  EXPECTED:     3.4  3.0  3.4  2.3  3.6   3.3  3.4  2.7
Z-RESIDUAL:          2

     58  58                          .63     1.5         1.5
  OBSERVED: 1:  4    3    5    1    5     5    2    3
  EXPECTED:     3.8  3.4  3.8  2.6  3.9   3.7  3.7  3.0
Z-RESIDUAL:

    130  130                         .35     1.5         1.5
  OBSERVED: 1:  4    3    5    3    2     2    5    2
  EXPECTED:     3.5  3.1  3.5  2.4  3.7   3.5  3.5  2.8
Z-RESIDUAL:

    139  139                         .77     1.4         1.4
  OBSERVED: 1:  5    4    3    4    3     2    5    3
  EXPECTED:     3.9  3.5  3.9  2.8  4.0   3.8  3.9  3.2
Z-RESIDUAL:                              -2

     83  83                        -1.21     1.4         1.4
  OBSERVED: 1:  2    3    1    1    2     3    1    3
  EXPECTED:     2.2  1.9  2.2  1.5  2.3   2.1  2.1  1.7
Z-RESIDUAL:

     59  59                        1.43      1.4         1.4
  OBSERVED: 1:  5    5    5    2    3     5    5    3
  EXPECTED:     4.4  4.1  4.3  3.4  4.4   4.3  4.3  3.8
Z-RESIDUAL:                        -2

     31  31                          .21     1.4         1.4
  OBSERVED: 1:  5    3    2    4    4     2    3    2
  EXPECTED:     3.4  3.0  3.4  2.3  3.6   3.3  3.4  2.7
Z-RESIDUAL:

     49  49                         -.66     1.3         1.4
  OBSERVED: 1:  2    1    2    3    2     4    2    3
  EXPECTED:     2.6  2.3  2.6  1.8  2.7   2.5  2.5  2.0
Z-RESIDUAL:

    109  109                         .07     1.4         1.4
  OBSERVED: 1:  3    4    4    1    2     2    5    3
  EXPECTED:     3.3  2.9  3.3  2.2  3.4   3.2  3.2  2.5
Z-RESIDUAL:

     32  32                        2.93      1.1         1.4
  OBSERVED: 1:  5    5    4    5    4     5    5    5
  EXPECTED:     4.8  4.7  4.8  4.5  4.9   4.8  4.8  4.6
Z-RESIDUAL:              -2         -2

     43  43                          .07     1.4         1.4
```

```
  OBSERVED: 1:  4    3    2    2    4     4    1    4
  EXPECTED:     3.3  2.9  3.3  2.2  3.4   3.2  3.2  2.5
Z-RESIDUAL:                                    -2

     76  76                         2.93     1.1         1.4
  OBSERVED: 1:  5    5    4    5    4     5    5    5
  EXPECTED:     4.8  4.7  4.8  4.5  4.9   4.8  4.8  4.6
Z-RESIDUAL:              -2        -2

    107  107                        -.07     1.4         1.3
  OBSERVED: 1:  3    2    2    2    5     5    2    2
  EXPECTED:     3.2  2.7  3.1  2.1  3.3   3.1  3.1  2.4
Z-RESIDUAL:

     63  63                          .21     1.3         1.4
  OBSERVED: 1:  2    3    4    4    3     2    3    4
  EXPECTED:     3.4  3.0  3.4  2.3  3.6   3.3  3.4  2.7
Z-RESIDUAL:

     42  42                         -.66     1.2         1.3
  OBSERVED: 1:  2    1    3    2    2     3    2    4
  EXPECTED:     2.6  2.3  2.6  1.8  2.7   2.5  2.5  2.0
Z-RESIDUAL:                                         2

     38  38                         3.67     1.0         1.3
  OBSERVED: 1:  5    5    5    5    5     5    4    5
  EXPECTED:     4.9  4.9  4.9  4.7  4.9   4.9  4.9  4.8
Z-RESIDUAL:                                    -3

    140  140                        -.21     1.3         1.3
  OBSERVED: 1:  4    2    3    1    2     2    5    3
  EXPECTED:     3.0  2.6  3.0  2.0  3.2   2.9  3.0  2.3
Z-RESIDUAL:                                     2

     54  54                         -.21     1.1         1.3
  OBSERVED: 1:  2    3    2    4    4     3    2    2
  EXPECTED:     3.0  2.6  3.0  2.0  3.2   2.9  3.0  2.3
Z-RESIDUAL:                    2

    152  152                        1.43     1.2         1.1
  OBSERVED: 1:  4    5    4    5    5     4    3    3
  EXPECTED:     4.4  4.1  4.3  3.4  4.4   4.3  4.3  3.8
Z-RESIDUAL:

    158  158                        -.07     1.2         1.2
  OBSERVED: 1:  3    3    4    3    2     1    4    3
  EXPECTED:     3.2  2.7  3.1  2.1  3.3   3.1  3.1  2.4
Z-RESIDUAL:                               -2
```

Appendix F: Full Administration Initial Recruitment Email

Dear [teacher name],

My name is Katy Hogue and I'm writing to ask for your help in a research project that I am conducting for my doctoral dissertation at Boston College. I am interested in measuring teachers' skills using data to inform their classroom instruction and I have developed a survey to measure this. *[School district name]* has agreed to administer this survey and I am hoping you are willing to participate!

The survey should take 15-20 minutes to complete. I will be sending you a link to the survey on *[insert date which will be 2 days after this email is sent]* and your responses will be anonymous. If you decide to complete this survey, you can enter to win one of four $50 Amazon.com gift cards!

I hope you are willing to participate! Please keep an eye out for the survey link, which should arrive on *[date]*. Please let me know if you have any questions.

Best,

Katy

[*phone number*]

Appendix G: Full administration survey recruitment email

Dear [teacher name],

I hope you are well. As I mentioned in my email sent on *[date]*, I'm writing to ask for your help in a research project that I am conducting for my doctoral dissertation at Boston College. I am interested in measuring teachers' skills using data to inform their classroom instruction and I have developed a survey to measure this. [*School district name]* has agreed to administer this survey and I hope you are willing to participate!

To access the survey, please click on this link: [*link*]

You can also copy and paste the link into a web browser. Although this survey link is unique to you, the data collected by this survey will not be linked to your email address or your name, and your responses will be anonymous.

The survey should take 15-20 minutes to complete. If you decide to complete this survey, you can enter to win one of four $50 Amazon.com gift cards! You will be able to enter your email address at the end of the survey to be included in this raffle. Your email address will not be associated with your responses to the survey. You are free to withdraw from the survey at any time or skip questions for any reason. There are no penalties for withdrawing or skipping questions.

I hope you are willing to participate! Please let me know if you have any questions.

The survey will close on *[date]*.

Best,

Katy

[*phone number*]

Appendix H: Full administration survey

Instructions: These scenarios describe different teachers' skillsets when using data to inform their classroom instruction. Data refers to a wide variety of data sources in these scenarios, not solely assessment data.

- Reflect on your own skills using data to inform your classroom instruction while reading each scenario.
- Compare your own skillset to the teacher's experience in the scenario.
- Choose the response that most accurately reflects your comparison.

Practice Item: As the scenario-type survey may be a new experience, this practice item allows you to practice engaging with a scenario. Please read the following practice scenario and compare your own experience to Teacher P.

Teacher P sometimes has trouble clarifying and writing clearly defined questions that describe their thoughts about student learning. Once they have defined their questions, they may benefit from some support to systematically gather data related to these questions and compare multiple data sources. When identifying interventions to meet their student learning goals, the support of others is helpful, and this teacher finds that working with others to plan to assess progress towards their goals is useful.

How do your skills using data to inform classroom instruction compare to Teacher P?

- Teacher P is much less skilled than me
- Teacher P is slightly less skilled than me
- Teacher P is just like me
- Teacher P is slightly more skilled than me
- Teacher P is much more skilled than me

1. Teacher A can independently identify interventions to meet their student learning goals, while also planning to assess progress towards these goals. They sometimes have trouble implementing their chosen intervention as intended, and sometimes have difficulty collecting data about student progress towards their goals during the intervention. Teacher A finds support helpful when examining student outcome data collected from their intervention and identifying if these outcomes meet their stated goals.

How do your skills using data to inform classroom instruction compare to Teacher A?

- Teacher A is much less skilled than me
- Teacher A is slightly less skilled than me
- Teacher A is just like me

- Teacher A is slightly more skilled than me
- Teacher A is much more skilled than me

2. Teacher B occasionally has difficulty thinking critically about and analyzing data to help answer their questions about student learning. Once they complete their data analysis focused on these questions, they may need support to generate hypotheses about how to improve student learning from this analysis and to validate their interpretations. When they choose an intervention for student learning, Teacher B has trouble planning for student progress towards their student learning goals and often holds back on sharing progress with colleagues during their implementation.

How do your skills using data to inform classroom instruction compare to Teacher B?

- Teacher B is much less skilled than me
- Teacher B is slightly less skilled than me
- Teacher B is just like me
- Teacher B is slightly more skilled than me
- Teacher B is much more skilled than me

3. Teacher C can identify data to answer their questions about student learning and consistently organizes that data clearly. This teacher independently interprets results from their data analysis related to their questions about student learning, while writing clear and aligned goals for student learning based on hypotheses about how to improve student learning that come from their analysis. Once they have implemented a student learning intervention, they can successfully examine student outcome data collected from this intervention and are able to identify their next steps.

How do your skills using data to inform classroom instruction compare to Teacher C?

- Teacher C is much less skilled than me
- Teacher C is slightly less skilled than me
- Teacher C is just like me
- Teacher C is slightly more skilled than me
- Teacher C is much more skilled than me

4. Teacher D occasionally needs help narrowing down their thoughts about student learning to clearly defined questions. Once they have identified a student learning intervention, they require the support of others to both identify any professional development that they will need before implementing the intervention and to identify the data required to document progress towards their student learning goals. Teacher D rarely shares progress with others during their intervention's implementation and has trouble monitoring student progress towards their identified goals in a systematic way.

How do your skills using data to inform classroom instruction compare to Teacher D?

- Teacher D is much less skilled than me
- Teacher D is slightly less skilled than me
- Teacher D is just like me
- Teacher D is slightly more skilled than me
- Teacher D is much more skilled than me

5. Teacher E sometimes benefits from help to distill their thoughts and wonderings about student learning into clearly defined questions. Once they have defined their questions, this teacher occasionally seeks support to understand what data can help answer these questions and to organize this data clearly. This teacher may struggle to analyze how well the implementation of their intervention has gone and is unsure of their next steps once the intervention is complete.

How do your skills using data to inform classroom instruction compare to Teacher E?

- Teacher E is much less skilled than me
- Teacher E is slightly less skilled than me
- Teacher E is just like me
- Teacher E is slightly more skilled than me
- Teacher E is much more skilled than me

6. Teacher F struggles to generate and write clearly defined questions that clarify their thoughts on student learning. They have trouble understanding what data can help answer their questions or identifying different techniques to display their data. This teacher needs the support of others to gather data to help answer their questions, think critically about and analyze this data, and compare multiple data sources.

How do your skills using data to inform classroom instruction compare to Teacher F?

- Teacher F is much less skilled than me
- Teacher F is slightly less skilled than me
- Teacher F is just like me
- Teacher F is slightly more skilled than me
- Teacher F is much more skilled than me

7. Teacher G successfully compares multiple data sources while critically examining data to help answer their questions about student learning. Interpreting meaning from their data analysis to develop student learning goals is a strength of Teacher G’s, and they consider all potential consequences of these goals before moving forward. Teacher G benefits from some support to both identify interventions to help attain these goals and to identify the data required to document progress towards these goals.

How do your skills using data to inform classroom instruction compare to Teacher G?

- Teacher G is much less skilled than me
- Teacher G is slightly less skilled than me
- Teacher G is just like me
- Teacher G is slightly more skilled than me
- Teacher G is much more skilled than me

Demographic questions:

1. Please select the gender with which you most identify.

- Female
- Male
- Nonbinary (neither female nor male)
- Other
- Prefer not to answer

2. Please select the race/ethnicity with which you most identify (you may select more than one).

- African American/Black
- Asian
- Hispanic/Latinx
- Native American/Alaskan Native
- Native Hawaiian/Pacific Islander
- White
- Other (please specify)
- Prefer not to answer

3. How many years of teaching experience do you have?

- 0-3 years
- 4-6 years
- 7-10 years
- 11-15 years
- More than 15 years

4. Please select your primary teaching role.

- Administrator
- Elementary homeroom teacher
- Elementary specialist teacher
- Secondary English/ELA teacher
- Secondary History/Social Studies teacher
- Secondary Math teacher
- Secondary Science teacher

- Secondary World Languages teacher
- Secondary specialist teacher
- Other (please specify)

Appendix I: Person-Response Table from the full administration

```
TABLE OF POORLY FITTING PERSON   (ITEM IN ENTRY ORDER)
NUMBER - NAME -- --------------- MEASURE - INFIT (MNSQ) OUTFIT

     20   20                        1.09     9.5   A    9.6
  OBSERVED: 1:  5    5    5    1    3     3    2
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:     3         4   -5         -2

    119  119                        1.09     4.9   B    5.1
  OBSERVED: 1:  3    5    3    1    4     5    3
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:                   -5

     53   53                        2.25     3.4   C    3.8
  OBSERVED: 1:  5    4    3    5    3     3    4
  EXPECTED:     3.6  4.4  2.6  4.7  4.2   4.7  2.8
Z-RESIDUAL:     2                        -3

    201  201                        1.71     3.5   D    3.6
  OBSERVED: 1:  5    3    2    M    M     M    M
  EXPECTED:     3.4  4.2  2.3
Z-RESIDUAL:     2

     50   50                         .37     3.1   E    2.9
  OBSERVED: 1:  3    4    2    4    3     2    4
  EXPECTED:     2.9  3.7  1.8  4.0  3.4   4.1  2.1
Z-RESIDUAL:                              -3    2

    139  139                         .37     3.0   F    3.0
  OBSERVED: 1:  4    4    3    2    4     3    2
  EXPECTED:     2.9  3.7  1.8  4.0  3.4   4.1  2.1
Z-RESIDUAL:                   -3

    104  104                        3.75     2.2   G    2.9
  OBSERVED: 1:  3    5    4    5    5     4    4
  EXPECTED:     4.2  4.8  3.1  4.9  4.7   4.9  3.4
Z-RESIDUAL:                              -3

     25   25                         .73     2.8   H    2.7
  OBSERVED: 1:  3    2    1    5    5     4    3
  EXPECTED:     3.1  3.8  1.9  4.1  3.6   4.2  2.2
Z-RESIDUAL:         -2              2

     28   28                        4.39     1.0   I    2.8
  OBSERVED: 1:  5    5    3    4    5     5    4
  EXPECTED:     4.5  4.9  3.3  4.9  4.8   5.0  3.6
Z-RESIDUAL:                   -4

    138  138                        3.19     2.7   J    2.7
  OBSERVED: 1:  4    3    4    5    4     5    4
  EXPECTED:     4.0  4.7  2.9  4.8  4.5   4.9  3.2
Z-RESIDUAL:         -3

     91   91                        2.69     2.6   K    2.1
  OBSERVED: 1:  2    5    2    5    5     5    4
  EXPECTED:     3.8  4.6  2.7  4.8  4.3   4.8  3.0
Z-RESIDUAL:    -2

    164  164                        1.46     2.6   L    2.5
  OBSERVED: 1:  2    5    2    5    5     5    1
  EXPECTED:     3.3  4.1  2.2  4.4  3.8   4.5  2.5
Z-RESIDUAL:    -2                             -2

    126  126                       -2.02     2.1   M    2.6
  OBSERVED: 1:  3    2    2    2    2     3    2
  EXPECTED:     2.0  2.8  1.1  3.1  2.6   3.2  1.2
```

```
Z-RESIDUAL:               2

    175  175                        1.09     2.4   N    2.4
  OBSERVED: 1:  3    5    2    3    5     5    1
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:                   -2    2         -2

    118  118                        1.84     2.0   O    2.3
  OBSERVED: 1:  4    4    3    5    5     3    2
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:                              -3

    134  134                        2.69     1.9   P    2.1
  OBSERVED: 1:  5    4    3    4    4     4    4
  EXPECTED:     3.8  4.6  2.7  4.8  4.3   4.8  3.0
Z-RESIDUAL:

    156  156                        1.09     2.1   Q    2.0
  OBSERVED: 1:  3    3    4    4    3     4    3
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:               2

      8    8                        2.25     2.1   R    2.1
  OBSERVED: 1:  3    4    4    4    4     4    4
  EXPECTED:     3.6  4.4  2.6  4.7  4.2   4.7  2.8
Z-RESIDUAL:               2

     54   54                        1.84     2.0   S    2.0
  OBSERVED: 1:  3    5    2    5    5     5    1
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:                                   -2

     74   74                        1.84     2.0   T    2.0
  OBSERVED: 1:  3    5    2    5    5     5    1
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:                                   -2

    199  199                        1.71     1.9   U    2.0
  OBSERVED: 1:  4    3    3    M    M     M    M
  EXPECTED:     3.4  4.2  2.3
Z-RESIDUAL:

     26   26                        -.37     1.9   V    1.9
  OBSERVED: 1:  3    2    1    4    4     5    1
  EXPECTED:     2.7  3.4  1.5  3.7  3.1   3.8  1.8
Z-RESIDUAL:         -2

    161  161                        1.84     1.8   W    1.8
  OBSERVED: 1:  4    3    3    4    4     4    4
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:         -2                         2

    129  129                        1.46     1.7   X    1.6
  OBSERVED: 1:  3    4    4    4    3     4    3
  EXPECTED:     3.3  4.1  2.2  4.4  3.8   4.5  2.5
Z-RESIDUAL:               2

    137  137                        3.19     1.2   Y    1.7
  OBSERVED: 1:  4    4    3    5    5     4    4
  EXPECTED:     4.0  4.7  2.9  4.8  4.5   4.9  3.2
Z-RESIDUAL:                              -2

     59   59                        -.75     1.7   Z    1.6
  OBSERVED: 1:  2    4    1    3    2     5    2
  EXPECTED:     2.5  3.3  1.4  3.5  3.0   3.7  1.6
Z-RESIDUAL:                               2

     29   29                        3.75     1.6        1.1
  OBSERVED: 1:  5    5    3    5    5     5    2
```

```
  EXPECTED:     4.2  4.8  3.1  4.9  4.7   4.9  3.4
Z-RESIDUAL:                                   -2

     62   62                        1.09     1.6        1.5
  OBSERVED: 1:  3    4    2    5    5     4    1
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:                         2         -2

    187  187                        3.75     1.6        1.1
  OBSERVED: 1:  5    5    3    5    5     5    2
  EXPECTED:     4.2  4.8  3.1  4.9  4.7   4.9  3.4
Z-RESIDUAL:                                   -2

    141  141                        1.09     1.5        1.6
  OBSERVED: 1:  3    4    3    5    3     3    3
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:                              -2

      4    4                        3.19     1.5        1.2
  OBSERVED: 1:  3    5    2    5    5     5    4
  EXPECTED:     4.0  4.7  2.9  4.8  4.5   4.9  3.2
Z-RESIDUAL:

     44   44                        1.09     1.4        1.5
  OBSERVED: 1:  2    4    1    5    4     5    3
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:    -2

     46   46                        1.46     1.5        1.4
  OBSERVED: 1:  3    5    2    5    4     5    1
  EXPECTED:     3.3  4.1  2.2  4.4  3.8   4.5  2.5
Z-RESIDUAL:                                   -2

    146  146                        1.84     1.5        1.4
  OBSERVED: 1:  4    3    3    5    3     5    3
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:         -2

    195  195                         .73     1.4        1.4
  OBSERVED: 1:  2    4    2    3    4     5    3
  EXPECTED:     3.1  3.8  1.9  4.1  3.6   4.2  2.2
Z-RESIDUAL:

     32   32                        3.19     1.0        1.4
  OBSERVED: 1:  4    5    3    5    4     4    4
  EXPECTED:     4.0  4.7  2.9  4.8  4.5   4.9  3.2
Z-RESIDUAL:                              -2

    182  182                        3.19     1.0        1.4
  OBSERVED: 1:  4    5    3    5    4     4    4
  EXPECTED:     4.0  4.7  2.9  4.8  4.5   4.9  3.2
Z-RESIDUAL:                              -2

    110  110                        1.84     1.4        1.4
  OBSERVED: 1:  4    3    2    4    5     5    3
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:         -2

     80   80                         .00     1.4        1.4
  OBSERVED: 1:  3    3    2    3    4     3    3
  EXPECTED:     2.8  3.6  1.6  3.9  3.3   4.0  1.9
Z-RESIDUAL:

     27   27                         .73     1.3        1.4
  OBSERVED: 1:  4    3    2    3    4     5    2
  EXPECTED:     3.1  3.8  1.9  4.1  3.6   4.2  2.2
Z-RESIDUAL:

    102  102                         .73     1.3        1.4
```

```
  OBSERVED: 1:  4    4    2    5    3     3    2
  EXPECTED:     3.1  3.8  1.9  4.1  3.6   4.2  2.2
Z-RESIDUAL:                              -2

    198  198                    2.69     1.2        1.3
  OBSERVED: 1:  5    4    3    5    4     4    3
  EXPECTED:     3.8  4.6  2.7  4.8  4.3   4.8  3.0
Z-RESIDUAL:

     86   86                    1.46     1.3        1.3
  OBSERVED: 1:  4    4    2    5    4     5    1
  EXPECTED:     3.3  4.1  2.2  4.4  3.8   4.5  2.5
Z-RESIDUAL:                                   -2

    188  188                    1.09     1.2        1.3
  OBSERVED: 1:  4    4    2    5    4     3    2
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:                              -2

    192  192                     .73     1.2        1.3
  OBSERVED: 1:  2    5    2    4    3     5    2
  EXPECTED:     3.1  3.8  1.9  4.1  3.6   4.2  2.2
Z-RESIDUAL:

    163  163                    -.37     1.3        1.2
  OBSERVED: 1:  3    4    1    3    3     3    3
  EXPECTED:     2.7  3.4  1.5  3.7  3.1   3.8  1.8
Z-RESIDUAL:

    186  186                    1.09     1.2        1.2
  OBSERVED: 1:  3    4    3    4    4     3    3
  EXPECTED:     3.2  4.0  2.1  4.3  3.7   4.4  2.4
Z-RESIDUAL:                              -2

     22   22                    1.46     1.2        1.2
  OBSERVED: 1:  3    5    1    5    4     5    2
  EXPECTED:     3.3  4.1  2.2  4.4  3.8   4.5  2.5
Z-RESIDUAL:

     31   31                    1.84     1.2        1.2
  OBSERVED: 1:  3    5    1    5    4     5    3
  EXPECTED:     3.5  4.3  2.4  4.5  4.0   4.6  2.7
Z-RESIDUAL:              -2

     36   36                    1.46     1.2        1.2
  OBSERVED: 1:  3    5    1    5    4     5    2
  EXPECTED:     3.3  4.1  2.2  4.4  3.8   4.5  2.5
Z-RESIDUAL:

     39   39                    3.19     1.2        1.1
  OBSERVED: 1:  4    4    2    5    5     5    4
  EXPECTED:     4.0  4.7  2.9  4.8  4.5   4.9  3.2
Z-RESIDUAL:
```

Appendix J: Score Conversion Table from the full administration

TABLE OF SAMPLE NORMS (500/100) AND FREQUENCIES CORRESPONDING TO COMPLETE TEST

SCORE	MEASURE	S.E.	NORMED	S.E.	FREQUENCY	%	CUM.FREQ.	%	PERCENTILE
7	-8.03E	1.89	-131	119	1	.5	1	.5	1
8	-6.65	1.12	-45	71	0	.0	1	.5	1
9	-5.70	.87	15	55	0	.0	1	.5	1
10	-5.03	.78	58	49	0	.0	1	.5	1
11	-4.45	.73	94	46	0	.0	1	.5	1
12	-3.93	.71	127	45	0	.0	1	.5	1
13	-3.44	.70	158	44	0	.0	1	.5	1
14	-2.96	.69	188	44	0	.0	1	.5	1
15	-2.48	.68	218	43	1	.5	2	1.0	1
16	-2.02	.67	248	43	1	.5	3	1.5	1
17	-1.58	.66	276	42	1	.5	4	2.0	2
18	-1.15	.64	302	40	0	.0	4	2.0	2
19	-.75	.63	328	40	4	2.0	8	3.9	3
20	-.37	.61	352	39	5	2.5	13	6.4	5
21	.00	.61	375	38	4	2.0	17	8.4	7
22	.37	.60	398	38	13	6.4	30	14.8	12
23	.73	.60	421	38	10	4.9	40	19.7	17
24	1.09	.60	444	38	20	9.9	60	29.6	25
25	1.46	.61	467	39	24	11.8	84	41.4	35
26	1.84	.63	491	40	19	9.4	103	50.7	46
27	2.25	.65	517	41	24	11.8	127	62.6	57
28	2.70	.68	545	43	24	11.8	151	74.4	68
29	3.19	.73	577	46	22	10.8	173	85.2	80
30	3.75	.77	612	49	12	5.9	185	91.1	88
31	4.39	.83	652	52	11	5.4	196	96.6	94
32	5.13	.90	699	57	7	3.4	203	100.0	98
33	6.03	1.00	756	63	0	.0	203	100.0	100
34	7.22	1.22	831	77	0	.0	203	100.0	100
35	8.74E	1.94	927	122	0	.0	203	100.0	100