REFRAMING TEACHERS' CONCEPTIONS OF ASSESSMENT AND DATA USE

BY

NORA K. GANNON-SLATER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Thomas A. Schwandt, Chair
Professor Jennifer C. Greene
Professor Katherine Ryan
Professor Fouad Abd El Khalick

ABSTRACT

Teachers' use of assessment data for instructional decision making is one of several essential components of data-driven decision making (DDDM) as a reform strategy to improve teaching and learning. This thesis explores three different issues related to the use of assessment data for instructional decision making. The first paper examines the way instructional decision making is currently conceived in the DDDM discourse. The second paper uses qualitative data to highlight the broad spectrum of assessment strategies a group of high school science teachers employ to collect data for instruction, and it illustrates the complexity of using that data for instructional decision making. The third paper is a study of an instrument that measures pre-service teachers' conceptions of assessment. Qualitative pre-tests of the instrument and quantitative data analyses post-administration suggest that pre-service teachers' beliefs about assessment are conditional on the tensions present in the cultural and political assessment climate. Collectively, these papers support reframing conceptions of assessment and data in DDDM research to provider clearer explanations of what "use" of data for instructional decision making actually means.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

INTRODUCTION

Teachers' use of data for instructional decision making is one of several essential components of data-driven decision making (Coburn & Turner, 2011; Erikson, 2007; Hoover & Abrams, 2013). Across all levels of public education, from chief state school officers to classroom teachers, data-driven decision making (DDDM) has become a compulsory component of education practice (Mandinach, 2012). Public school choice and federal funding depend on the achievement data reported by public schools (Mandinach, Rivers, Light, Heinze, & Honey, 2006, Recesso & Zepeda, 2009). Federal policy initiatives like No Child Left Behind (U. S. Department of Education, 2002), reauthorization of the Individuals with Disabilities Act (U.S Department of Education, 2006), and Race to the Top (U.S. Department of Education, 2009) require that educators use data to monitor and improve school processes and student outcomes. Undeniably, "data have become the vehicle of choice for ensuring accountability" (Earl & Katz, 2002, p. 2).

The DDDM movement in education also draws on successful business practices of Total Quality Management (Deming, 1986) and organizational learning (Senge, 1990), whereby data facilitates a cycle of continuous improvement (Ingram, Louis & Schroeder, 2004). Research also highlights that data use is especially prevalent in higher performing schools (Datnow, Park & Wohlstetter, 2007; Supovitz & Klein, 2003; Tognieri & Anderson, 2003). The use of assessment data to improve instruction and learning also piggybacks on the promising insights from research on formative assessment as a critical practice for increasing student achievement (Black & Wiliam, 1998; Schneider & Andrade, 2013). The need to meet accountability requirements combined with a strong belief that using data in education practice facilitates improvement in instruction and learning has made DDDM a national priority in education reform with significant

investments of public and private monies (Coburn & Turner, 2012; Mandinach, Gummer & Muller, 2011; Means, Padilla & Gallagher, 2010; Spillane, 2012). Because of their proximity to student learning and achievement, teachers, and the ways they make use of assessment data in the classroom to improve instruction and, ultimately student learning, are especially subject to scrutiny.

Overview of the Three Papers

The three papers in this thesis examine different facets of teachers' use of data in the classroom. The first paper examines the concept of instructional decision making and how it is currently conceived in the dominant DDDM discourse. This paper reframes teachers' use of data for instructional decision making in ways that recognize the inherent ethical and moral demands of teaching. The second paper is a qualitative field study of what counts as data for a group of high school science teachers. The study provides key insights to the complexity of teachers' data use practices in their ongoing instruction including the wide spectrum of assessment practices they employ to collect data on student learning, how multiple sources of data interact in their instructional decision making, and their expectations for students to be users of their own data. The third paper is a study of an instrument that measures pre-service teachers' conceptions of assessment. The instrument was used in a larger study of how those conceptions are related to teachers' ideas about of teaching and learning, self-efficacy, and beliefs about assessment data that are useful to teachers for instruction. The findings suggest that pre-service teachers' beliefs about the purposes of assessment are conditional on the tensions present in the cultural and political assessment climate, leading to a way they think assessment currently happens and is used and a way they think it should happen or be used.

Reframing Teachers' Use of Data for Instructional Decision Making

Teachers are expected to use data to inform their instructional decision making. However, in the literature on how teachers' use data in the classroom, there is little discussion of how we conceive of the instructional decisions that teachers face in their practice. This paper uses key articles in the DDDM literature to illustrate that the dominant discourse of DDDM depicts teachers as making instrumental decisions – that is, decisions about the appropriate means to a given end – namely, student achievement in some content area. Focusing solely on instrumental decisions where the end is taken for granted promotes a view of teaching that is wholly technical in nature. This view ignores the fact that the instructional decisions teachers face are often as much about deciding the right ends as they are about determining the best means, requiring teachers' experiential knowledge and practical wisdom in addition to necessary technical knowledge and skills. This paper seeks to reframe the study of teachers' data use practices in ways that capture the complexities of their practice and enhances understandings of how teachers' practical wisdom, in concert with the use of data, contribute to and improve decision making in teachers' practice.

The target journal for this paper is the *Harvard Educational Review*. As an academic essay, this conceptual paper addresses both a popular and controversial topic in education today and provides a critical reflection on the practice of teaching. The paper is perhaps most relevant to scholars of DDDM research but other audiences, including district and school leadership or even teachers may find aspects of the article relevant to understanding their own challenges to using data in practice.

Broadening Our Conceptions Of Data: Understanding Secondary Science Teachers' Data Use In Ongoing Instruction

Teachers' instructional practice is constituted by a complex interaction between curriculum, pedagogy and their views of students in which simultaneous and not mutually independent considerations of data, goals, professional values and external pressures are in play (Schwandt, 2005; Young, 2005). Educational policymakers at federal, state, and local levels put considerable faith in the power of data – especially standardized assessment data – to improve teaching practices. Yet, teachers have reported time and time again that these kinds of data lack instructional value. The literature reveals that teachers are often skeptical of, and in some cases resistant to, the use of standardized assessment data for instructional decision making and often rely on other kinds of data for this purpose. However, there is very little empirical examination of the different kinds of data that teachers employ in making ongoing instructional decisions and why they use these data. This paper looks at the actual practice of a particular group of high school science teachers focusing specifically on how they rely on data other than standardized assessments for instructional decision making and how they appraise the value of that data. The findings from this study suggest that teachers rely more on informal versus formal assessment data but that they still collect a substantial amount of formal data with the expectation that students will use the data to monitor their own learning. This study argues that a broader framework for assessment and data is necessary to understand the complex nature of data use.

The target journal for this paper is *American Journal of Education*. This journal has published widely on the topic of data use including entire issues devoted to the phenomenon (see Coburn & Turner, 2012; Wayman & Springfield, 2006). While the sample for this study included only science teachers, it does not focus on what is unique to science classrooms with respect to

data and decision making and it is likely that teachers' in other disciplines also use a wide range of assessment data in their instructional decision making. However, some of the frameworks employed to analyze the data in this study come from science education researchers (e.g., formal and informal assessment, Ruiz-Primo & Furtak, 2007; assessment conversations; Duschle & Gitomer, 1997). I found these frameworks to be helpful for make sense of teachers' data use in ongoing instruction and it is my hope that by publishing in this journal, researchers in other content areas might find them useful as well.

The Way It Is And The Way It Should Be: Pre-Service Teachers' Conceptions Of Assessment

Teachers' dispositions toward assessment and learning can influence ways they use assessment information to determine effectiveness of instruction (Coburn & Talbert, 2006; Ingram et al., 2004; McMillan, 2003). International and cross-cultural studies of both pre-service and in-service teachers' conceptions of assessment provide evidence that their conceptions are related to the policy contexts in which they practice (Brown & Michaelides, 2011; Brown & Remesal, 2013). Additionally, scholars have theorized that pre-service teachers' conceptions of assessment may be an important component of understanding, and ultimately improving, pre-service teachers' development of assessment literacy, including the knowledge and skills needed to use data for instructional decision making (Brown & Hirschfield, 2011; Mandinach & Gummer, 2013). The purpose of this study was to examine United States pre-service teachers' (PSTs) conceptions of assessment at a large, mid-western university using the *Teachers' Conceptions of Assessment* (TCoA) framework (Brown, 2008), which has been studied in many different policy contexts. However, the previously validated instrument associated with this framework underwent considerable changes in this study as a result of qualitative pre-testing

including content analysis, expert review and most importantly, cognitive interviews. The findings from this study indicate that while this population of PSTs holds similar conceptions of assessment to those that have been reported elsewhere, U.S. PSTs also hold a conditional conception of assessment, believing there is a *way assessment currently happens or is used* and a way *assessment should be*. This study highlights the value in testing and evaluating validated instruments prior to administering to new populations. The findings also support prior research that suggests conceptions of assessment may be a function of the cultural origins of the research and that in different societies somewhat different intentions and patterns exist.

The target journal for this paper is *Assessment in Education: Policy, Principles and Practice*. This journal has published several studies on the TCoA framework albeit most studies do not modify the original instrument except to translate as necessary. This article, then, reflects a novel contribution to the current literature on the TCoA and will hopefully encourage researchers to continue examine how pre-service and in-service teachers' conceptions of assessment (using this framework or developing new ones) are shaped by the cultural and political contexts in which they learn and practice.

Significance

Each of these papers employ different perspectives and frameworks through which we can frame shared understandings of data, assessment, and instructional decision making. Collectively, they augment an ongoing research agenda in DDDM to describe and explain the inherent complexity of practices of data use (Coburn & Turner, 2012; Turner & Coburn, 2011) by enhancing understandings of myriad issues related to teachers' use of data for instruction. This work serves to bring a counter-voice to the dominant discourse of data-driven decision making. It is difficult to argue against the very *idea* of data-driven decision making; this thesis

serves to systematically unveil some of the assumptions of the presumably conflict-free, value-neutral idea of using data to inform instructional decision making. In addition, this work supports broader notions of assessment and data to be included in studies of teachers' data use. Put another way, this thesis emphasizes pluralistic and divergent views on the value and potential for DDDM as a critical strategy in education reform. As a result, education scholars, practitioners and policy-makers who subscribe to data-driven decision making may be more clear on what exactly they are subscribing to.

REFERENCES

Black, P., & Wiliam D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-73.

Brown, G. T. L. (2008). Conceptions of assessment: Understanding what assessment means to teachers and students. New York: Nova Science Publishers.

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice, 15*(1), 3–17.

Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education, 26*, 319–337.

Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. The Spanish Journal of Psychology, 15, 75–89.

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

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.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high performing districts use data to improve instruction for elementary school students*. Los Angeles, CA: Center on Educational Governance, USC Rossier School of Education.

Deming, W.E. (1986). *Out of crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study. Detente Possible?', Theory Into Practice, 44: 3, 234 — 244.

Duschl, R. & Gitomer, D. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, 4, 37–73.

Earl, L. & Katz, S. (2002). Leading schools in a data-rich world. In K. Leithwood, P. Hallinger, G. Furman, P. , Gronin, J. MacBeath, B. Mulford, & K. Riley (Eds.), *The second international handbook of educational leadership and administration*. Dordrecht, Netherlands: Kluwer.

Erikson, F. (2007). Some thoughts of proximal formative assessment of student learning (pp. 187-217). In P. A. Moss (Ed.), *Evidence and decision-making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell

Hoover, N. R. & Abrams, L. M. (2013) Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26(3),* 219-231.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision-making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, *42*(1), 30-37.

Mandinach, E. B., Gummer, E. S., & Muller, R.D. (2011). *The complexities of integrating data-driven decision making into professional preparation in schools of education: It's harder than you think*. Alexandria, VA, Portland, OR, and Washington, DC: CNA Education, Education Northwest, and WestEd.

Mandinach, E. B., Rivas. L., Light, D., Heinze, C., & Honey, M. (2006, April). *The impact of data-drive decision making tools on educational practice: a systems analysis of six school districts*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34-43.

Means, B., Padilla, C., Gallagher, L. (2010). *Use of education data at the local level from accountability to instructional improvement*. Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002)

Reauthorization of the Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446 108th Cong. 118 Stat. (2006).

Recesso, A. M. & Zepeda, S. J. (2009) Evidential reasoning and decision support in assessment of teacher practice (pp. 363-381). In T. J. Kowalski & T. J. Lasley II (Eds.) *Handbook of Data-based Decision Making in Education*. New York, NY: Routledge.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84.

Schneider, M. C. & Andrade, H. (2013). Teachers' and administrators' use of evidence of student learning to take action: conclusions drawn from a special issue on formative assessment. *Applied Measurement in Education, 26(3),* 159-162.

Schwandt, T. (2005) The Centrality of practice to evaluation. *American Journal of Evaluation, 26*(1), 95–105.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.

Spillane, J. (2012). Data in Practice: Conceptualizing the Data-Based Decision-Making Phenomena. *American Journal of Education*, *118*(2), 113-141.

Supovitz, J. A., & Klein, V. (2003). Mapping a course for improved student learning: How schools systematically use student performance data to guide improvement. Philadelphia: Consortium for Policy Research in Education.

Togneri, W., & Anderson, S.E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools*. Washington, D.C.: The Learning First Alliance and the Association for Supervision and Curriculum Development.

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

U.S. Department of Education (2009). *Race to the Top Program: Executive summary*. Washington, DC: Author.

Wayman, J. C. & Stringfield, S. (2006). Data use for school improvement: School practices and research perspectives. *American Journal of Education 112*(4), 463-468.

Young, V. M. (2005). *Teachers' use of data: Embedded Contexts and Institutional Logics* (Doctoral dissertation). Stanford University. Retrieved from *ProQuest Dissertations and Theses* (UMI Number 3186431)

PAPER 1

**REFRAMING TEACHERS' USE OF DATA FOR MAKING INSTRUCTIONAL
DECISIONS**

Data-driven decision making (DDDM), defined as "drawing on and interacting with
information in the course of decision making," (Coburn & Turner, 2012, p. 99) is a strategy that
assumes improvement in instructional practices and, ultimately, enhanced student achievement is
directly tied to teachers' use of data (achievement or otherwise) (Hamilton, et al., 2007; Ikemoto
& Marsh, 2007; Young, 2006). Educational policymakers at federal, state, and local levels put
considerable faith in the power of data to improve teaching practices. Federal mandates to use
data for school accountability (No Child Left Behind, 2002; Race To The Top, 2011), coupled
with research that has demonstrated that data use is especially prevalent in higher performing
schools (Datnow, Park, & Wohlstetter, 2007; Supovitz & Klein, 2003; Tognieri & Anderson,
2003), has prompted an unprecedented focus on the collection and use of data by schools and
districts (Marsh, Pane, & Hamilton, 2006). DDDM has become a prominent strategy for
education reform with significant investments of public and private monies (Coburn & Turner,
2012) and a national priority in education (Mandinach, Gummer, & Muller, 2011; Means,
Padilla, & Gallagher, 2010; Spillane, 2012).

The idea that educators should make decisions informed by empirical information is
perfectly logical. However, how we conceive of the kinds of instructional decisions that teachers
face in their practice is a critical consideration. Teachers are often pictured as making
instrumental decisions—that is, decisions about the appropriate means to achieve a given end.
The given end or objective is increased student achievement. Instrumental decisions require
pedagogical content knowledge to determine what instructional strategies will best increase
chances of meeting that objective (Shulman, 1987) and assessment literacy to determine whether

the objective has been met (Stiggins, 1991). These competencies are critical for successful practices in formative assessment (Young & Kim, 2010), arguably one of the most promising strategies in teachers' practice for increasing student learning (Black & Wiliam, 1998; Schneider & Andrade, 2013). While some decisions that teachers face may be instrumental in nature, focusing solely on the kinds of decisions where the end is taken for granted (e.g., student achievement on a standardized test) promotes a view of teaching practice that is wholly technical in nature. This view ignores the fact that the instructional decisions teachers face are often as much about deciding the right ends as they are about determining the best means. Teachers not only face dilemmas of *what to do* about some given end but also make moral and ethical decisions about *what is right to do* for students and for themselves as teachers.

This paper draws from key articles in the DDDM literature to illustrate that the current data-driven decision making discourse on teachers' use of data in the classroom assumes that teachers' instructional decisions are means-end decisions where the end (e.g., student achievement) is taken for granted. To illustrate this point, I use a fictional scenario that portrays instructional decision making in this way. On one hand, this discourse has burgeoned in light of accountability policies that emphasize such given ends and contributes immensely to understanding how organizations and individuals work toward those ends. On the other hand, I argue that focusing solely on these kinds of decisions renders teaching a technical enterprise and comes at the expense of understanding how teachers' experiential knowledge and practical wisdom, in concert with the use of data, contribute to and improve decision making in teaching practice. I propose to reframe the study of teachers' data use practices in service of enhancing teachers' professional judgment.

One Instance of Data Use

Ms. Johnson, a teacher with 15 years of experience, has spent the last week teaching her entry-level Chemistry students about organization of the periodic table of elements. She creates and administers a short, mid-unit quiz to gauge the students' abilities to identify and explain the various organizational concepts of the table (atomic number, electronic configurations, and chemical properties). After grading the quizzes, Ms. Johnson determines that less than half the class grasps this material. She examines the data more closely for possible relationships between missed multiple-choice and short-answer items within and across students to determine if they had trouble recalling information or held genuine misconceptions about the content. She knows she does not have much time to re-teach this material because of the constraints of her curriculum pacing guide but she also knows that mastering these basic concepts about the table are critical to student success in the rest of the unit. She also remembers from her weekly data log in the school's data management system that nearly three-quarters of her class was absent at some point last week. During her professional learning community (PLC) time that afternoon, she shares this data with her colleagues, getting feedback on her analysis of the data and gathers input on how to proceed.

She makes two instructional decisions. First, not wanting to punish students for poor performance as a possible result of attendance, she will hand the quizzes back and offer students the opportunity to write corrections and resubmit quizzes for a higher grade. She then prepares two learning activities for the first half of class. Students who did well on the quiz and showed mastery of concepts will use the time to read ahead in the textbook on the next set of concepts and attempt some practice problems. Students who failed the quiz will complete a worksheet that teaches students a mnemonic for remembering the key differences between the organizational elements of the periodic table. She plans to give short "concept check" the following morning to see if students were able to master and retain the material.

In this slightly fictionalized scenario, we find a typical dilemma teachers face on a day-to-day basis—making sense of evidence of student learning and moving forward[1]. Ms. Johnson had collected from data from an assessment about students' mastery of some concepts in Chemistry. When she determines that students have not yet mastered these concepts, Ms. Johnson must decide what to do. When situated in the literature on teachers' use of data for

---

[1] The scenario in this paper is composed from teacher interviews and classroom observations on high school science teachers' data use.

instructional decision making, this scenario serves as a prime example of means-ends instructional decision making.

Key Studies of Teachers' Use of Data for Instructional Decision Making

The scholarly discourse on DDDM is an interdisciplinary arena of research addressing a broad spectrum of concepts and issues related to data use and employing a variety of methodologies. A considerable amount of descriptive research portrays how the phenomenon of data use happens in a variety of contexts and at different levels (for reviews of this research see Coburn & Turner, 2011; Marsh, 2012; Young & Kim, 2010). The organizational and political contexts that constrain or enable data use practices are perhaps the most widely studied aspect of this phenomenon (Coburn & Turner, 2011). In contrast, there are few detailed descriptions of the processes teachers use to make sense of myriad assessment data in the classroom, especially in the current data-rich environment (Coburn & Turner, 2011; Little, 2012; Schneider & Andrade, 2013; Young & Kim, 2010).

The thirteen articles referenced in this paper are drawn from this body of research. Two studies focus on the implementation of data use interventions in districts (Ingram, Louis, & Schroeder, 2004; Kerr, Marsh, Ikemoto, & Barney, 2006). These studies are widely recognized for their identification of favorable and unfavorable organizational conditions related to teachers' use of data in the classroom and for school improvement (Coburn & Turner, 2011; 2012; Young & Kim, 2010). The remaining studies reflect some of the latest attempts at understanding teachers' use of data in the classroom. They are diverse with respect to the scope of study, school levels included, and methodologies employed. However, they all are concerned with identifying the necessary attributes, skills, and supports for effective practices of teacher data use. Three of these studies were selected from a recent special issue in *Applied Measurement in Education* on

"Teachers' and administrators' use of evidence of student learning to take action" (Schneider & Andrade, 2013). Three other studies include how high school teachers use data in instruction (Datnow, Park, & Kennedy-Lewis, 2012), the results from a yearlong workshop for teachers on data use (Murnane, Sharkey, & Boudett, 2005), and the results of a generalizability study of the relationships between teachers' data analytic skills, mathematical pedagogical content knowledge, and instructional decisions (Heritage, Kim, Vendlinski, & Herman, 2009). Two studies are technical reports from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) (Herman, Osmundson, Ayala, Schneider, & Timms, 2006; Herman, Osmundson, Dai, Ringstaff, & Timms, 2011) and another two studies are from research conducted at the Consortium for Policy Research in Education (CPRE) (Goertz, Nabor Oláh, & Riggan, 2009; Riggan & Nabor Oláh, 2011). "Teachers' ability to use data to inform instruction: Challenges and supports" (Means, Chen, DeBarger, & Padilla, 2011) is a sub-study of the larger *Study of Education Data Systems and Decision Making* sponsored by the U.S. Department of Education. Table 1 (at the end of this paper) provides an overview of each of these empirical studies including the central research question and research design.

Although the purposes of these studies varied, the theoretical and conceptual frameworks employed in the studies, along with their findings, provide a cogent narrative of how teachers' use of data for instructional decision making is largely considered a technical enterprise. The cycle of continuous improvement that prioritizes technical knowledge and skills for using data in instruction is a framework frequently employed in these studies. Embedded in this research are particular conceptions of assessment and data that privilege certain kinds of information over others in making instructional decisions.

Cycle of Continuous Improvement Framework

The cycle of continuous improvement is a popular framework for conceptualizing teachers' use of data to inform instructional decision making (Goertz et al., 2009; Ingram et al., 2004; Means et al., 2011; Riggan & Nabor Oláh, 2011). It is often depicted as a series of phases: plan, implement, assess, reflect, and repeat (Boudett, City, & Murnane, 2013; Means, Padilla, & Gallagher, 2010), or: define needs, set priorities and goals, plan interventions, and monitor progress (Goldring & Berends, 2009). Variants of this framework include the quality assessment framework (Herman et al., 2006), formative assessment theory of action (Schneider & Gowan, 2013), and formative feedback frameworks (Herman et al., 2011; Ruiz-Primo & Li, 2013). Another way of thinking about this framework is in the form of three questions: (1) Where are we going? (goals); (2) What evidence do we have? (data); (3) What are we going to do next? (instruction/intervention).

By starting with the known effects of educational processes (e.g., student test results), teachers can work backwards to experiment with possible causes in curriculum and instruction and then evaluate the effects of those experiments through another assessment of outcomes.

The continuous improvement framework and its variants are considerably tidy contexts for analysis whereby researchers can study how teachers participate in these cycles of inquiry, identifying points of deviation as cause for concern and further investigation (Goertz et al., 2009). For example, studies frequently focus on teachers' capacities to reflect on data, which include evidence of teachers' use of data to articulate student strengths and weaknesses, identify student misconceptions, or pose possible issues with instruction. These competencies are sometimes examined simultaneously with teachers' abilities to determine next steps in

instruction reflecting their pedagogical-content technical skills (Goertz et al., 2009; Heritage et al., 2009; Means et al., 2011; Ruiz-Primo & Li, 2013; Schneider & Gowan, 2013).

Ms. Johnson's dilemma is an example of the kinds of instructional decisions that are assumed to be happening in the cycle of continuous improvement. Ms. Johnson has a clear goal—mastery of organizational concepts of the periodic table of elements. She receives some data that indicates this goal has not been met and reflects accordingly for next steps. Her decision of what to do next is based on a careful analysis of the data (examination of within and across items and students), leading to a decision to differentiate instruction for students who failed to master concepts and those that were secure. These are the kinds of decisions that current research on teachers' use of data endeavors to examine, describe, and ultimately, theorize. Research on teachers' use of data focuses heavily on these isolated incidents of decision making (Goertz et al., 2009; Heritage et al., 2009; Means et al., 2011; Ruiz-Primo & Li, 2013; Schneider & Gowan, 2013) taking note of how teachers like Ms. Johnson analyze this singular source of data, draw inferences about student learning and decide next steps. Within this continuous improvement framework, researchers might also examine the nature of feedback Ms. Johnson provides to students on their quiz to determine the quality of feedback as it relates to students' accuracy of responses and receiving information on next steps (Herman et al., 2006; Ruiz-Primo & Li, 2013). They may also investigate how certain organizational contexts, including access to other data through the data management system, the support of her professional learning community and, the constraints of her curriculum-pacing guide, influence her instructional decisions (Ingram et al., 2004; Kerr et al., 2006).

Studies of teachers' data use have also recently started to evaluate the instructional decisions teachers make with data (Goertz et al., 2009; Heritage et al., 2009; Schneider &

Gowan, 2013). Within this framework, the conclusion thus far has been that teachers do not use data in expected ways and that instruction does not improve or change as a result. For example, Goertz et al. (2009) concluded that the depth to which teachers' analyzed data strongly influenced their instructional responses. The study suggests that, most of the time, teachers' did not analyze the data deeply and employed only organizational changes in teaching (e.g., deciding what or who to teach) rather than actually changing the way they presented material to students. In our scenario, we see that Ms. Johnson's skills in assessment literacy help her create (or select) a mid-unit quiz that is well-aligned with the learning outcomes and provides evidence of student thinking. Her analytic skills help her identify patterns of student error to determine the cause of low performance. Finally, her pedagogical content expertise guides her choice of strategies in differentiating instruction and provides struggling students with a learning tool to help them master the material. Ms. Johnson's choice to differentiate instruction, including presenting material in a new way to students who struggle, would likely reflect a commendable change in instruction as a function of data use.

Conceptions of Data and Assessment

A considerable amount of the research on teachers' use of data for instruction tends to focus exclusively on how teachers utilize one singular source of information in their instructional decision making (Goertz et al., 2009; Heritage et al., 2009: Means et al., 2011; Schneider & Gowan, 2013). In some instances, researchers rate the quality of an instructional decision based on the use of this single source of data (Goertz et al., 2009; Heritage et al., 2009). To be sure, understanding how teachers do make sense of single sources of evidence is valuable; for example, teachers do need certain statistical skills to analyze quantitative data. However, drawing conclusions about the act of constructing instructional implications from a single source

of data, particularly in ongoing instruction, stands in sharp contrast to research that indicates that teachers draw from a wide variety of data in their instructional decision making (Coburn & Talbert, 2006; Datnow et al., 2012; Goertz et al., 2009; Herman et al., 2006; Hoover & Abrams, 2013; Ingram et al., 2004; Kerr at al., 2006). Research has also suggested that multiple measures of student performance from a variety of sources may enhance data use by allowing for triangulation of findings, providing greater balance, and reducing the risk of making a wrong decision resulting from relying on a single assessment (Copland, 2003; Herman, 2002). In our example, Ms. Johnson also considers two other sources of data (e.g., previous student performance and attendance) in making sense of her students' performance on this present quiz, and these data collectively contribute to her decisions of allowing students to submit corrections for a higher grade and differentiating instruction.

The more important matter, however, is how data are defined or what counts as data. It is not uncommon for research to all but ignore any other information teachers take into account in their instructional decision making, dismissing it as data from "not typically conceived assessment sources" (Ruiz-Primo & Li, 2013) or "external factors" (Goertz et al., 2009) that interfere with the data-driven decision making process. Ingram et al. (2004) concluded that teachers' pre-existing beliefs about the data, including whether they perceive it as useful for instruction and valid measures of educational goals, act as potential barriers to teachers using the data for instruction. They also noted that when teachers and administrators base their decisions on experience, intuition, and anecdotal information rather than on data, this, too, is a potential barrier to establishing school cultures supportive of DDDM. These kinds of claims rest on an assumption that renders data, as that which is not experience, intuition, and anecdote, as an authority for decision making. Even more so, these claims assert that learning cannot be validly

assessed through firsthand observation and that formal assessments provide better (more reliable, more valid) evidence of student learning than teachers' classroom observations of their own students (Erikson, 2007). Data then, by default, is narrowly defined as that which can be systematically collected and recorded. Information that cannot be ascertained in these ways, like intuition, experience, or anecdote, is essentially considered *not* data. Research on teachers' data use perpetuates this definition, examining mostly formal assessments that produce a written record of student work. For example, researchers conclude that in situations where teachers use a combination of data and experience, intuition, and anecdote, it is difficult to determine the extent to which DDDM occurred (Ingram et al., 2004). Our scenario illustrates this point as well. It focuses on Ms. Johnson's use of her quiz in making instructional decisions, how other systematic data (e.g., attendance, previous performance) contributes to her understanding of that data, and omits any mention of how her fifteen years of experience teaching chemistry, student behavior, or any other information not systematically observed may have played into her decision.

Research on teachers' use of data for instructional decision making is heavily focused on data, as systematic records of student learning, and on the technical knowledge and skills teachers need to properly use that data in decision making. The continuous improvement framework is only sustainable for studying teachers' data use as long as we assume that the ends of instruction are a given and all a teacher needs to do is decide the best means to meet those ends. This research tends to perpetuate the view that data forthcoming from assessment tools are authoritative for decision making in the classroom and depicts the practice of teaching as an application of instrumental reasoning.

Two Considerations of Teachers' Practice

Research focused on data use in terms of instrumental reasoning, guided by a cycle of continuous improvement, is valuable for understanding the technical knowledge and skills needed to make sense of data and construct implications for instruction. However, it promotes the view of teaching as a technical activity where the teacher's role is that of an interventionist in service of a given end. That view is starkly different from a view of teaching as an exercise in professional judgment where the teachers' task includes determinations of means as well as ends. To illustrate this difference, I draw on the Aristotelian distinction between two kinds of knowledge for making practice decisions, *techne* and *phronesis* (Flyvbjerg, 2001; Schwandt, 2005). As an illustration of *phronesis,* we will return to our scenario with Ms. Johnson.

The kind of knowledge required for making means-ends decisions is what Aristotle called *techne (*Schwandt, 2005). *Techne* can be understood as the application of technical knowledge and skills (i.e., analytic and instructional know-how) according to a rationality (i.e., cycle of continuous improvement) governed by a conscious goal (i.e., student achievement) (Flyvbjerg, 2001). The story of Ms. Johnson reflects this kind of practice and illustrates how she determined the best course of action for her students in light of previous knowledge about their performance and attendance, and illuminates how her decision was both afforded and constrained by organizational contexts. Possessing the technical know-how for analyzing data and determining next steps in instruction are valuable skills for teachers, especially in the current data-rich environment. However, when the primary focus of inquiry surrounding the use of data for instructional decision making is teachers' technical knowledge, then we are led to assume that the professional practice of teaching is primarily a matter of intervention (Biesta, 2007). Teachers' actions in the classroom are interventions in service of "what works", mirroring that of

a doctor administering a round of radiation to a patient to minimize the cancerous tumor. This is a causal model of professional action "based on the idea that professionals do something—they administer a treatment, they intervene in a particular situation—in order to bring about certain effects" (Biesta, 2007, p. 7). Because the outcomes of professional action are given—the goals (e.g., increased student achievement) are set—the only relevant questions teachers need to ask are about what interventions are most effective and efficient for achieving those ends. Schwandt (2005) explains further: "Thus, for example, the [DDDM] enterprise seeks to equip a classroom teacher with a kind of knowledge of fact or procedure in such a way that every problem she faces, from class size, to a decision about inclusion, to deciding the best curriculum for reading, is a technical problem to be solved. If this effort succeeds, we have made a technically adept teacher" (p. 22).

The causal model of professional action ignores the fact that teaching is fundamentally an interaction with students: "If teaching is to have any effect on learning, it is because of the fact that students interpret and try to make sense of what they are being taught. It is through processes of (mutual) interpretation that education is possible" (Biesta, 2007, p. 8). Teaching does not simply involve decisions about what works but also decisions about what is appropriate and right for students. Just because the most effective intervention can be adopted, it does not actually answer the question of whether it should be (Biesta, 2007). To suggest data can replace "normative professional judgment is…to deny educational practitioners the right *not* to act according to evidence about 'what works' if they judge such that such a line of action would be educationally undesirable" (Biesta, 2007, p. 11).

It must be understood that teachers have the ability to decide how to achieve a given end but also an ability to reflect upon and determine what are good ends. In other words, they must

simultaneously consider both means and ends in their decision making. The kind of knowledge required in these situations is what Aristotle called *phronesis* (Schwandt, 2005). *Phronesis* concerns the analysis of values—things that are good or bad for students—as a point of departure for action. It is a sense of the ethically practical; a person possessing *phronesis* or practical wisdom has knowledge of how to behave in each particular circumstance. Conceived in this way, teachers' practice decisions are an exercise in professional judgment. Schwandt (2005) explains further:

> [E]quipping the teacher with our best science and decision rules will do little to
> help her grasp the fact that in making evaluation decisions she is morally
> accountable. In others, she is obliged to defend as appropriate and right the
> decision she makes in arguing for reducing her class size this semester, for
> excluding Billy who has cerebral palsy from her classroom, and for deciding that
> Mary's ability in reading will be best enhanced by teaching her phonics. In sum,
> efforts to enhance the teachers' ability to manage and control her practice, over
> time, alienate the teacher from the responsibility that inevitably befalls her for
> the moral-political decisions she makes as a teacher. (p. 23)

A teacher's practical wisdom addresses the particularity of things and situations, identifying distinctive conditions so that she can decide how to move in a "morally framed direction" (Eisner, 2002, p. 381). Put another way, the guiding questions of practice conceived in this way (in contrast to the technical framework of teaching) might be: (1) Where are we going? (2) Is this desirable? (3) What should be done? (Flyvbjerg, 2001, p. 60). To answer these questions, teachers need a grasp on the distinctive features of their students and classrooms to make good decisions.

Current research on DDDM largely ignores a view of teaching as an exercise of professional judgment about appropriate educational ends. It focuses instead on instructional decisions where the ends are given and suggests that teachers' intuitions, experience, and anecdotes, which comprise teachers' *phronesis*, *hinder* teachers' use of data in instructional decision making. *Phronesis* does not replace or supplant *techne*; rather, it involves exercising that technical knowledge in view of careful reflection upon and determination of good ends. DDDM research has indicated that teachers do not agree about which student outcomes are most important (Coburn & Talbert, 2006; Ingram et al., 2004). This is evidence of teachers' consideration of both means and ends in their practice; unfortunately, it is not investigated beyond this observation. Data can be interrogated for what they reveal about both means and ends. To illustrate this point, we return to Ms. Johnson.

Another Instance of Data Use

Ms. Johnson, a teacher with 15 years of experience, has spent the last week teaching her entry-level Chemistry students about organization of the periodic table of elements. She creates and administers a short, mid-unit quiz to gauge the students' abilities to identify and explain the various organizational concepts of the table (atomic number, electronic configurations, and chemical properties). After grading the quizzes, Ms. Johnson determines that less than half the class grasps this material. She examines the data more closely for possible relationships between missed multiple-choice and short-answer items within and across students to determine if they had trouble recalling information or held genuine misconceptions about the content. She also considers that nearly three-quarters of her class was absent last week at some point, and the students have not had a day off in nearly three weeks and spring break is still two weeks away. During her lunch hour on the same day, Ms. Johnson overhears the Algebra II teacher talking about the recent poor performance of some of her students in his classroom. She knows she does not have much time to re-teach this material because of the curriculum-pacing guide but she also knows from teaching this unit for the last ten years that mastering these basic concepts about the table are critical to student success in the rest of the unit. Her instincts tell her, however, that handing these grades back will diminish student morale and she is wary of asking them engage the material again—in the past, when this class performs poorly, they tend to just shut down. During her PLC time that afternoon, she shares the student performance data, as well as her memories and feelings

about the whole situation with her colleagues. They talk with her about the challenges of keeping students engaged after failing an assessment and how important it is to balance content coverage with maintaining supportive relationships with students. They share their own experiences of how they dealt with such situations in the past and encourage her to listen to her instincts, but still find a way to cover the content.

When the students show up the next day, with anxiety in their faces over their grades, Ms. Johnson announces that yesterday's quiz was a freebie and today they will work in pairs on an assignment. Each group will be given several fake elements and a set of rules and will need to assemble a new periodic table of elements based on these rules. Their performance on this group assessment will replace their quiz grades. Ms. Johnson knows that evidence of individual student mastery of these concepts will not be accessible from this group assessment but she forgoes that advantage in favor of keeping students engaged in the material and reinforcing classroom norms of peer support for learning. She smiles to herself as students start to pair up and eagerly get to work.

This fictionalized scenario is in keeping with a different way of understanding teacher decision making. However, as was the case in the first scenario, Ms. Johnson determines that students have not yet mastered these concept. She must decide what to do and draws on a variety of data sources to inform this decision. However, her decision to discount their poor performance on the quiz and give a group assessment instead relies heavily on her experience, intuition, and anecdotal information *as well as* the data. In this scenario, she considers whether the end in view—individual student mastery of basic chemistry concepts—is appropriate in light of the current situation. In her deliberations, she is considering not only *what to do* about this apparent lack of mastery but also *what is right to do* for these students and in light of her own sense of being a teacher. This kind of decision, where both the means (how to move student learning forward) and the ends (what are the desired outcomes) are not given, requires Ms. Johnson's analysis of her own values as a teacher and making decisions based on that analysis. Data play a significant part but are not authoritative in these decisions.

Reframing Research on Assessment, Data, and Practice

Young (2006) defines practice as "both the source of teachers' appetite for particular types of data and the lens through which they judge the appropriateness and usability of the data" (p. 521). Current research on teachers' use of data for instructional decision making is divorced from this idea of practice. Privileging knowledge as *techne* in DDDM research comes at the expense of understanding how teachers' experiential knowledge and practical wisdom contribute to and improve instructional decisions, especially those where teachers wrestle with both means and ends.

Scholars of DDDM have already recognized the need to turn toward practice, particularly as it pertains to the study of organizations implementing DDDM initiatives (Coburn & Turner, 2012). Current empirical work on practice, defined by Cook and Brown (1999) as the "coordinated activities of individuals and groups in doing their 'real work' as it is informed by a particular organizational or group context" (as cited in Coburn & Turner, 2012, p. 386–87), provides key insights into how data use interventions 'play out on the ground' within the environments in which they are implemented (Daly, 2012; Honig & Venkateswarn, 2012; Spillane, 2012). I seek to extend this work on practice by building on useful insights from the literature, including potential theoretical frameworks and methods for studying individual teachers' data use for instructional decision making. I also propose three aspects of practice that are worth investigating for their potential to shed light on how data adds to the improvement of teaching and learning.

Theoretical Frameworks

Organizational research on data use has provided insights into the processes by which people interpret and make decisions with data (Coburn, Toure, & Yamashita, 2009; Spillane &

Miele, 2007) and how teachers learn when they talk with one another about student work and other forms of data (Gearhart et al., 2006; Gearhart & Osmundson, 2009; Horn & Little, 2010; Kazemi & Franke, 2004; Little, 2007). These lines of inquiry serve as reminder that practice is a socio-cognitive undertaking where perceptions and understanding are already shaped by worldviews and mental models that serve as filters for new knowledge and experiences (Spillane & Miele, 2007). Put another way, teachers construct understandings of new information and enact interpretations from their environment within the confines of their current cognitive frameworks (e.g., knowledge and beliefs), which are shaped by the collective understandings within an organization (Coburn, 2001; Spillane, Reiser, & Rimer, 2002).

Leaving techno-rational theories (i.e., continuous improvement) aside, researchers must pursue other theories and perspectives with which to study teachers' data use on the ground. Frameworks used in examining educational reform policy that might be useful in studying teachers' use of data for instruction include sense-making and co-construction perspectives (Coburn, 2001; Coburn & Talbert, 2006; Datnow et al., 2012; Spillane et al., 2002). These perspectives place implementers (e.g., educators) at the "forefront of reform efforts, highlighting the process by which they interpret, adapt, or transform policy" (Datnow et al., 2012, p. 251). Individual beliefs and actions are guided by notions of appropriate, natural, or legitimate behavior and these norms also define the limits of possible action "in the form of taken-for-granted roles, rules, or ways of doing things" (Coburn & Talbert, 2006, p. 472). Additionally, these perspectives privilege the roles that social learning and cognitive capacities play in implementing initiatives, assuming people's actions cannot be understood apart from the contexts in which they happen. These frameworks are already being applied in DDDM research, helping to refine our understandings of how organizations navigate the implementation of

DDDM interventions (Coburn & Turner, 2012). However, such inquiry might extend further, to better understand the individual beliefs and organizational norms and routines that inform teachers' normative decision making. In this way, DDDM research would help practitioners see that their respective practices involve moral purposes and intentions, not to be constructed as ends to which practice is a technical means but "commitments that can only be realized in and through the lived reality of practice" (Schwandt, 2005, p. 24).

There are other frameworks outside of DDDM research that might also improve our understandings of teachers' use of data for instructional decision making. Improvement Research is a research and development infrastructure currently underway at the Carnegie Foundation for the Advancement of Teaching. According their website, this approach "allows us to cull and synthesize the best of what we know from *scholarship and practice"* (http://www.carnegiefoundation.org/improvement-research/approach) and points to improvement science as a framework of inquiry to accelerate learning about practices that support student achievement. This approach adheres to six principles, the first of which includes engaging key participants (e.g., teachers) early and often to identify specific problems and to help work toward user-centered solutions. DDDM research would also benefit from modeling responsive, participatory and empowerment evaluation approaches to inquiry, endeavoring to be collaborative and action-research oriented. Where participants (teachers, administrators, students, and so on) encounter one another, listen to each other, learn, and deliberate a course of action, people and situations are more likely to change (Schwandt & Burgon, 2006).

Method

Research on teachers' data use practices will also benefit from current methodological trends in DDDM research in organizations. For example, Little (2012) reviewed five exemplary

studies of teachers' data use to frame new directions of organizational research on data use

(Coburn & Turner, 2012). She names these studies "micro-process" studies, requiring close

attention to patterns of on-the-ground interaction. She explained that:

> Although interviews, surveys, and self-report logs and diaries supply *ex situ* accounts of
>
> practice and point usefully to salient dimensions of interaction and context, it seems
>
> unlikely that a robust understanding of practice can be achieved absent the strategic use
>
> of methods that capture the detail, nuance, and patterning of social interaction. (p. 146)

The studies she reviewed are characterized by sustained and systemic observation of both verbal

and nonverbal actions that expose the "patterned and consequential aspects of data use practice

that elude the blunter instruments of interview and survey" (Little, 2012, p. 151).

Research on teachers' data use practices should rely on audiotaped and videotaped

records of two primary work spaces: (1) classrooms to capture the subtlety, consistency,

collective skill, and tacit understandings that constitute the social and intellectual fabric of

classrooms; and (2) professional learning environments (e.g., professional development, grade

level meetings) to "trace the trajectory of teacher learning and its relationship to teachers'

encounters with evidence of various kinds" (Little, 2012, p. 157). Heritage and Heritage's (2013)

study of a single teacher's questioning practices and her decisions *in situ* reflect this direction of

research. DDDM research will benefit from conceptualizing teachers as individuals, each with

differing perceptions and needs. Approaching research in these ways makes it more likely that

teachers' practical wisdom will become more salient as it bears on teachers' individual and

collective deliberations and decisions, and more likely that research will capture critical insights

into the roles that data play in the consideration of both means and ends.

Studying Expertise

In their introduction to a special issue on formative assessment, Schneider and Andrade (2013) recognized that "the skills needed by teachers and administrators to be efficient, effective users and leaders of formative assessment are complex and likely require years of practice to acquire" (p. 161). This suggests there are already experts out there from whom we can learn. Previous research on expert and novice teachers has shown that the expert teacher's understanding of classrooms is more comprehensive, characterized by Copeland, Birmingham, DeMuelle, and D'Emidio-Caston (1994) as an "increase in the 'quantity and complexity of linkages' among ideas and by a shift in the focus of these linked ideas toward issues more central to classroom teaching" (p. 176; see also Westerman, 1991). Admittedly, identifying expert teachers is difficult to do as experience does not equal expertise and cultural differences abound (Berliner, 2001). However, the propositions about expert teachers put forth by Berliner (2001) portray both the teacher as scientist and artist, proficient in her knowledge, skill, and craft, possessing both technical knowledge and practical wisdom. Although expert teachers may be challenging to identify in research, there is much to gained in learning how their knowledge as *techne* and knowledge as *phronesis* interact in the course of instructional decision making.

Understanding Resistance to Change

Inquiry on teachers' data use practices requires systematic investigation of the reported resistance and skepticism teachers express toward enacting data use routines in their daily practice (e.g., Ingram et al., 2004; Marsh et al., 2006; Young, 2006). It is not likely the case that teachers are data-phobic, but rather they do not have recent experience in working with data to improve specific classroom practices (Ingram et al., 2004). As mentioned earlier, teachers' perceptions of proposed change to their practices are filtered through their worldviews. Each

individual teacher, therefore, filters information about a change or innovation through his or her own reality, which is also shaped by the collective understandings within their grade level or department, school, and district. Regardless of the nature of the change being made, individual teachers may exhibit resistance based upon their own perceptions and circumstances. These factors can result in individuals responding to changes or innovations in ways that may be viewed as irrational (Duffy & Roehler, 1986). Studies of classrooms suggest that teachers use a variety of normative and pragmatic criteria for selecting classroom procedures, typically *adapting* rather than merely *adopting* policy initiatives (Diamond & Cooper, 2007; Doyle & Ponder, 1977). For example, such inquiry might examine how teachers evaluate the practicality of data use interventions in light of their own practice, which includes multiple, competing demands on time and resources as well as competing conceptions of what is right to do. Some scholars have focused on how teachers' data use practices are filtered through their own sense of self-efficacy and anxiety related to data use (Dunn et al., 2013). Furthermore, research on teachers' conceptions of assessment (Brown, 2008; DeLuca, Chavez, & Chao, 2012) suggests that teachers' beliefs about the purposes of assessment in education likely influence the methods of assessment (and subsequent data) they find useful for teaching and helpful to students.

Finally, Doyle and Ponder (1977) note that teachers are concerned with the immediate contingencies and consequences of adopting change. They argued that teachers use three criteria to judge the practicality of adopting a change initiative in their practice: (1) the extent to which the change proposal depicts classroom contingencies—that is, teachers must be able to *see* how it would look in practice; (2) the extent to which the change proposal is congruent with what they are already doing in their classroom and; (3) the extent to which the amount of return is balanced by amount of investment (Doyle & Ponder, 1977). Using these criteria, we might hypothesize

that DDDM interventions suggest new routines for determining instructional strategies and assessment practices that are wholly different than the practices teachers enact in the classroom on a daily basis. Teachers' reluctance to engage in data use routines might then be related to the fact that their self-image and preferred way of relating to students is threatened, or that they fail to see the return in experimenting with new instructional strategies if the reward structure is contingent upon results (i.e., increased student achievement) rather than innovativeness (i.e., trying new things). Teachers' concerns as to how data use mediates their own efforts to meet the ecological demands of the classroom as well as fulfill their moral obligations as a teacher warrant systematic investigation.

Examining the Role of Students in Teachers' Data Use Practices

Currently, the dominant discourse of data use largely omits one important group, namely the students who generate the data and which are the focus of data use practices (Hamilton, 2011). As Hamilton (2011) stated, "As co-constructors of their educational experiences, students play a key role in influencing the quality and nature of the learning activities in which they engage both inside and outside of the classroom" (p. 207). Similarly, Heritage and Heritage (2013) wrote "assessment needs to engage students in active reflection about what it means to be a learner and in participation with others, in constructing shared knowledge" (p. 188). Where students are mentioned in DDDM research, it is generally implied that they would know what to be able to do with the feedback (i.e., data) provided (Datnow et al., 2013; Ruiz-Primo & Li, 2013). Ingram et al. (2004) found that some teachers overlooked potentially useful data because they viewed the data only as feedback to students with respect to students' performance and not indicative of teachers' instructional effectiveness. Another study laid out an explicit theory of action for formative assessment in which a key assumption was that the student uses the

teachers' feedback to close gaps in learning and demonstrate mastery (Schneider & Gowan, 2013). Yet, it is not clear how students are expected to use that information, what opportunities do teachers need to provide, and what bears on both teachers' and students' deliberations of whether such information is useful and actionable. However, if teachers are assuming that students are self-regulating their own learning, then how does that assumption influence what kinds of data teachers make available to students? What are their beliefs about and expectations for how students will use that data? What do they feel are their obligations to help students gain ownership of their learning? Are there differences in teachers' expectations for elementary, middle, or high school students' use of data or between content areas and what warrants those differences? In order to understand better how data contributes to teachers' instructional decisions, researchers will need to focus on the dynamic social interactions between students and teachers and the influence of organizational and political contexts on those interactions.

Conclusion

The professional practice of teaching involves both technical skill in learning and applying a general teaching strategy to a specific case as well as professional judgment. Current research on teachers' use of data for instructional decision making overemphasizes the technical skill of teaching at the expense of understanding how teachers' experience, anecdotal knowledge, and intuitions may enhance teachers' instructional practices and improve student learning. To the extent that we do not modify our assumptions about teachers' instructional decision making, change our approaches to studying teachers' data use, and continue to exclude teachers' practical wisdom in efforts to explain effective data use practices, we will continue to conclude that teachers' changes to instruction as a result of data use will fall short of our expectations. DDDM inquiries that honor teachers' practice as primarily a matter of exercising professional judgment

about both means and ends should go beyond just identifying the variables that matter in getting

teachers to successfully engage in data use practices, and examine how data use interventions

enhance their ability to exercise judgment in practice (Schwandt & Burgon, 2006).

REFERENCES

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research*, *35*(5), 463-482.

Biesta**,** G.J.J. (2007). Why 'what works' won't work. Evidence-based practice and the democratic deficit of educational research. *Educational Theory 57*(1), 1-22.

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

Brown, G. T. L. (2008). *Conceptions of assessment: Understanding what assessment means to teachers and students*. New York: Nova Science Publishers.

Coburn, C. E. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23(2)*: 145-170.

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

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.

Coburn, C. E., Toure, J., & Yamashita, M. (2009). Evidence, interpretation, and persuasion: Instructional decision0making in the district central office. *Teachers College Record, 111(4)*: 1115-1161.

Copeland, W. D., Birmingham, C., DeMeulle, L., D'Emidio-Caston, M., & Natal, D. (1994). Making meaning in classrooms: An investigation of cognitive processes in aspiring teachers, experienced teachers, and their peers. *American Educational Research Journal, 31*(1), 166–196.

Copland, M. (2003). Leadership of Inquiry: Building and Sustaining Capacity for School Improvement. *Educational Evaluation and Policy Analysis* 25 (4): 375–95.

Datnow, A., Park, V., & Wohlstetter, P. (2007). *Achieving with data: How high performing districts use data to improve instruction for elementary school students*. Los Angeles, CA: Center on Educational Governance, USC Rossier School of Education.

Datnow, A., Park, V., Kennedy-Lewis, B. (2012). High school teachers' use of data to inform instruction. *Journal of Education for Students Placed At Risk*, 17, 247-265.

DeLuca, C., Chavez, T., & Cao, C. (2012). Establishing a foundation for valid teacher judgments: The role of pre-service assessment education. *Assessment in Education: Principles, Policy and Practice, Special Issue: Moderation Practice and Teacher Judgment,* 1-20.

Diamond, J.B., & Cooper, K. (2007). The uses of testing data in urban elementary schools: Some lessons from Chicago (pp. 241-263). In P. A. Moss (Ed.), *Evidence and Decision-Making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell.

Doyle, W., & Ponder, G. A. (1977). The practicality ethic in teacher decision-making. *Interchange*, *8*(3), 1-12.

Duffy, G., & Roehler, L. (1986). Constraints on teacher change. *Journal of Teacher Education, 37,* 55-58.

Dunn, K. E., Airola, D. T., Lo, W., 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(1),* 87-98.

Eisner, E. W. (2002). From episteme to phronesis to artistry in the study and improvement of teaching. *Teaching and Teacher Education*, *18*(4), 375-385.

Erikson, F. (2007). Some thoughts of proximal formative assessment of student learning (pp. 187-217). In P. A. Moss (Ed.), *Evidence and decision making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell

Flyvbjerg, B. (2001). *Making social science matter*. Cambridge, UK: University Press

Gearhart, M., & Osmundson, E. (2009). Assessment portfolios as opportunities for teacher learning." *Educational Assessment, 14 (1),* 1–24.

Gearhart, M., Nagashima, S., Pfotenhauer, J., Clark, S., Schwab, C., Vendlinksi, T., Osmundson, E., Herman, J., & Bernbaum, D. J. (2006). Developing expertise with classroom assessment in K–12 science: Learning to interpret student work. Interim findings from a two-year study." *Educational Assessment 11(3–4),* 237–63.

Goertz, M. E., Nabors Oláh, L., & Riggan, M. (2010). From testing to teaching: The use of interim assessments in classroom instruction. (CPRE Research Report No. RR-65). Philadelphia, PA: Consortium for Policy Research in Education.

Goldring, E. & Berends, M. (2009). *Leading with data: Pathways to improve your school*. Thousand Oaks, CA: Corwin.

Hamilton, L.S. (2011). The role of students in data use: Commentary on Coburn and Turner's Research on Data Use: A Framework and Analysis. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 207–210

Hamilton, L.S., Stecher, B.M., Marsh, J.A., Sloan McCombs, J., Robyn, A., Russell, J., et al., (2007). *Standards-based accountability under No Child Left Behind*. Santa Monica, CA: Rand Corporation.

Heritage, M. & Heritage, J. (2013). Teacher questioning: The Epicenter of Instruction and Assessment. *Applied Measurement in Education, 26(3),* 176-190.

Heritage, M., Kim, J., Vendlinski, T. and Herman, J. (2009), From evidence to action: A Seamless process in formative assessment? *Educational Measurement: Issues and Practice, 28,* 24–31.

Herman, J. L. (2002). *Instructional Effects in Elementary Schools*. Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J., Osmundson, E., Ayala, C., & Schneider, S. (2006). The Nature and Impact of Teachers' Formative Assessment Practices. *National Center for Research on Evaluation, Standards and Student Testing*, *1522*(310), CSE Technical Report 703. Retrieved from http://works.bepress.com/michael_timms/7/

Herman, J., Osmundson, E., Dai, Y., Ringstaff, C., & Timms, M. (2011). *Relationships between teacher knowledge, assessment practice, and learning-Chicken, egg, or omelet?* (CRESST Report 809). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Hoover, N. R. & Abrams, L. M. (2013) Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26(3),* 219-231.

Horn, I. S. & Little, J. W. (2010). Attending to problems of practice: routines and resources for professional learning in teachers' workplace interactions. *American Educational Research Journal* 47 (1): 181–217.

Ikemoto, G. S., & Marsh, J. A. (2007). Cutting through the "data-driven" mantra: Different conceptions of data-driven decision-making. In P. A. Moss (Ed.), *Evidence and Decision-Making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Kazemi, E., & Franke, M. L..(2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education* 7, 203–35.

Kerr, K. A., Marsh, J. A., Ikemoto, G. S., & Barney, H., (2006). Strategies to promote data use for instructional improvement: Actions , outcomes , and lessons from three urban districts. *American Journal of Education*, *112*, 496-520.

Little, J W. (2007). Teachers' accounts of classroom experience as a resource for professional learning and instructional decision making. (pp 217-240) In P. A. Moss (Ed.), *Evidence and Decision-Making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell.

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

Mandinach, E. B., Gummer, E. S., & Muller, R.D. (2011). *The complexities of integrating data-driven decision making into professional preparation in schools of education: It's harder than you think*. Alexandria, VA, Portland, OR, and Washington, DC: CNA Education, Education Northwest, and WestEd.

Marsh, J. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11), 1-48.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making Sense of Data-Driven Decision Making in Education: Evidence from Recent RAND Research*, Santa Monica, Calif.: RAND Corporation, OP-170-EDU, 2006.

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011) *Teachers' Ability to Use Data to Inform Instruction: Challenges and Support.s* U.S. Department of Education, Office of Planning, Evaluation and Policy Development, , Washington, D.C., 2011.

Means, B., Padilla, C., Gallagher, L. (2010). *Use of education data at the local level from accountability to instructional improvement*. Report prepared for U.S. Department of Education, Office of Planning, Evaluation and Policy

Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using Student-Assessment Results to Improve Instruction: Lessons From a Workshop. *Journal of Education for Students Placed at Risk*, *10*(3), 269–280.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002)

Riggan, M. & Nabors Oláh, L. (2011). Locating Interim Assessments Within Teachers' Assessment Practice. *Educational Assessment*, 16(1), 1-14.

Ruiz-Primo, M. A. & Li, M. (2013) Analyzing teachers' feedback practices in response to students' work in science classrooms. *Applied Measurement in Education, 26(3),* 163-175.

Schneider, M. C. & Andrade, H. (2013). Teachers' and administrators' use of evidence of student learning to take action: conclusions drawn from a special issue on formative assessment. *Applied Measurement in Education, 26(3),* 159-162.

Schneider, M. C. & Gowan, P. (2013). Investigating teachers' skills in interpreting evidence of student learning. *Applied Measurement in Education, 26(3),* 191-204.

Schwandt, T. (2005) The Centrality of practice to evaluation. *American Journal of Evaluation, 26*(1), 95–105.

Schwandt, T., Burgon, H. (2006). Evaluation and the study of lived experience. In I. F. Shaw, J. C. Greene & M. M. Mark (eds) *The Sage Handbook of Evaluation.* Sage: London.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, *57*(1), 1-23.

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

Spillane, J. P. & Miele, D. B. (2007). Evidence in practice: A framing of the terrain (pp. 46-73). In P. A. Moss (Ed.), *Evidence and decision making.* 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell.

Spillane, J., Reiser, B.J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, *72*(3), 387–431.

Stiggins, R. J. (1991). Assessment Literacy. *Phi Delta Kappan*, *72*(7), 534-39.

Supovitz, J. A., & Klein, V. (2003). Mapping a course for improved student learning: How schools systematically use student performance data to guide improvement. Philadelphia: Consortium for Policy Research in Education.

Togneri, W., & Anderson, S.E. (2003). *Beyond islands of excellence: What districts can do to improve instruction and achievement in all schools.* Washington, D.C.: The Learning First Alliance and the Association for Supervision and Curriculum Development.

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

U.S. Department of Education (2009). *Race to the Top Program: Executive summary.* Washington, DC: Author.

Young, V. M. (2005). *Teachers' use of data: Embedded Contexts and Institutional Logics* (Doctoral dissertation). Stanford University. Retrieved from *ProQuest Dissertations and Theses* (UMI Number 3186431)

Young, V. M. (2006). Teachers' Use of Data: Loose Coupling, Agenda Setting, and Team Norms. *American Journal of Education, 112(4),* 521-548

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: a literature review. *Education Policy Analysis Archives, 18* (19). Retrieved [date], from http://epaa.asu.edu/ojs/article/view/809

Table 1. Overview of Key Studies On Teachers' Use of Data

Citation	Identified Research Problem	Purpose of Study/Research Questions	School Level and Content Area	Theoretical/Conceptual Framework	Design/Methods/ Data sources
Datnow, Park & Kennedy-Lewis (2012)	Little is known about how teachers, situated within current policy environments and organizations, make instructional decisions using data	Examine how high school teachers situated within policy and work contexts use data to inform instructional decision	High school; Multiple content areas	Sense-making and co-construction theories	Qualitative case study of four public high schools; interviews, observations, document analysis of 90 participants
Goertz, Nabor Oláh & Riggan (2009)	There has been a proliferation of interim assessments but little is known abut how they are used by teachers in the classroom.	How do teachers learn from interim assessments and apply to instruction?	Elementary; Math	Cycle of continuous improvement	Mixed-methods case study of nine elementary schools; observations, interviews, surveys, data use assessment [a]
Heritage, Kim, Vendlinski & Herman (2009)	There is little research on whether teachers have the knowledge and skills to adapt instruction to meet student learning needs	How accurate are teachers in determining next steps in instruction?	Upper elementary; Math	Formative assessment and feedback	G-study of 118 teachers; assessed using "teacher knowledge measures" for ability to identify next steps in instruction

(continued)

41

Table 1 (continued)

Citation	Identified Research Problem	Purpose of Study/Research Questions	School Level and Content Area	Theoretical/Conceptual Framework	Design/Methods/Data sources
Herman, Osmundson, Ayala, Schneider, & Timms (2006)	The role of quality assessment in effective instruction and impact on student learning is unknown.	What is the quality of teachers' assessment practices? What are the links between quality assessment and student learning?	Middle school; Science	Quality assessment framework includes quality goals, assessments, interpretation and use	9 teachers from 9 different schools; observations and survey surveys
Herman, Osmundson, Dai, Ringstaff, Timms (2011)	Few studies have examined the relationship between teachers' pedagogical content knowledge (pck) and teachers' assessment practices and student learning	What is quality of teachers' pck and the relationships between pck and teacher knowledge and student learning	Upper-elementary, Math	Quality formative assessment practices are dependent upon teachers' foundational pck	Randomized field trial two cohorts of teachers (N=39) implemented new assessment; measures of teachers' pck, assessment practices, and student outcomes on end of year assessment

Table 1 (continued)

Citation	Identified Research Problem	Purpose of Study/Research Questions	School Level and Content Area	Theoretical/Conceptual Framework	Design/Methods/ Data sources
Hoover & Abrams (2013)	There is a lot of summative assessment data available to teacher but very little research on how such information is used in instructional-decision making; no information on how its used	Examine extent to which teachers report using summative assessment data in formative ways	Elementary, middle and high school; Multiple disciplines	Types and frequency of assessment types; Methods and frequency of quantitative data analysis; Changes to instructional practices	Survey (mostly close-ended, some open-ended questions) administered to 4000 teachers in large urban district; 650 responded
Ingram, Louis & Schroeder (2004)	The assumptions that underlie current accountability legislation have not been examined	What is the relationship between school culture and implementation of continuous improvement practices?	High schools, all content areas	Continuous improvement and Organizational Learning	Longitudinal study of high school implementing continuous improvement practices
Kerr, Marsh, Ikemoto, & Barney (2006)	It is unknown how different district efforts to promote DDDM increase quality of instruction and student learning outcomes	What strategies do districts employ to promote DDDM for instruction? What constrained or enabled district efforts?	All levels and content areas	Cycle of continuous improvement	Comparative case study of three districts; principal and teacher interviews and surveys

Table 1 (continued)

Citation	Identified Research Problem	Purpose of Study/Research Questions	School Level and Content Area	Theoretical/Conceptual Framework	Design/Methods/ Data sources
Means, Chen, DeBarger & Padilla (2011)	Teachers likelihood to use data is related to confidence about knowledge and skills in data analysis and interpretation	How to teachers individually or collectively reason through data for instruction?	Elementary; Math	Five critical skill areas for data use: data location, comprehension, interpretation, instructional decision making and posing questions	Student scenarios containing performance data; administered to 50 individual teachers and 72 groups of teachers
Murnane, Sharkey & Boudett (2005)	There is a challenge of providing adequate support to allow teachers and admins to make constructive use of student results	What makes support for teachers' data use valuable? How do teachers' use data when support is provided?	Elementary, middle and high school teachers; Math	Conceptual data use (Weiss, 1977) to enrich dialog about what students know, what they are able to do, and how effective instruction has been	Yearlong workshop that involved group discussion of cases; individual and team memos; team reports
Riggan and Nabor Oláh (2011)	Explosion of interim assessment but little is known how they fit in with other assessment practices	Do different types of assessment serve different purposes in the classroom and how do they connect to each other?	Elementary; Math	Cycle of continuous improvement	Classroom observations and interviews (part of larger study by Goertz et al., 2009)[a]

(continued)

44

Table 1 (continued)

Citation	Identified Research Problem	Purpose of Study/Research Questions	School Level and Content Area	Theoretical/Conceptual Framework	Design/Methods/ Data sources
Ruiz-Primo & Li (2013)	Feedback is a vital component of formative assessment but we don't know much about the nature of their feedback	What is the frequency and nature feedback to teachers provide in response to student assessment questions?	Elementary and middle school; Science	Quality formative feedback needs to be accurate and include, elements that lead to and guide subsequent student and instructional actions	Document analysis of 2000 pieces of feedback recorded in 200 student science notebooks from 26 teachers
Schneider & Gowan (2013)	There is limited information about teachers' skills to accurately measure student learning, interpret student responses, or adapt instruction based on student misconceptions	Do teachers math pdk differ in figuring what an item measures, analyzing student work, providing feedback, and determining next instructional steps;	Upper elementary; Math	Three assumptions of the formative assessment classroom theory of action	Randomized control trial of three cohorts of teachers (N=23); teachers scored student work without a rubric, with rubric, or with rubric and 30-minute professional development; assessment of teachers' analysis of student work

[a] Exact number of teachers in study are not reported.

PAPER 2

BROADENING OUR CONCEPTIONS OF DATA: UNDERSTANDING SECONDARY

SCIENCE TEACHERS' DATA USE IN ONGOING INSTRUCTION

Educational policymakers at federal, state, and local levels put considerable faith in the power of data—especially standardized assessment data—to improve teaching practices. Yet, teachers have reported time and time again that these kinds of data lack instructional value (Hamilton et al., 2007; Young 2006). Analogously, scholars have suggested that standardized assessment data, including the interim, benchmark, and common assessments administered at some frequency to all students at a grade level or in a content area, do not provide teachers with information they can use for ongoing instruction (Heritage, 2007). Shepard (2005) suggested that, at best, these kinds of assessments might be useful for more macro-level decisions, such as adoption of instructional programs or remedial placements, rather than the micro-level decisions that teachers make on a daily basis in response to individual student behavior and needs. The literature on data use reveals that teachers are often skeptical of, and in some cases resistant to, the use of standardized assessment data for instructional decision making and often rely on other kinds of data for this purpose (Ingram, Louis, & Schroeder, 2004; Coburn & Talbert, 2006; Young, 2006). However, there is very little empirical examination of the different kinds of data that teachers employ in making ongoing instructional decisions and why they use these data.

This study looks at the actual practice of a particular group of high school teachers focusing specifically on how they rely on data other than standardized assessments for instructional decision making and how they appraise the value of that data. "Practice" is a key notion here because, as Young (2006) has noted, it serves as "both the source of teachers' appetite for particular types of data and the lens through which they judge the appropriateness

46

and usability of the data" (p. 521). The findings of this study lend support to the claim that standardized assessment data are not useful for instruction and that data from more informal (Ruiz-Primo, 2011) assessment events (Torrance & Pryor, 1998), such as observing behaviors and engaging in question and answer sessions with students, are more useful for making instructional decisions than any other kind of data, standardized or otherwise. However, the findings also reveal that teachers administer a variety of formal assessments to students and hold an expectation for students to use the data from these assessments to monitor and advance their own learning. Research on data use requires a broader framework of understanding both assessment and data use, and further investigation is required on the possible hidden curriculum that underlies teachers' expectations for students to be users of their own data. I conclude by suggesting that research on data use must continue to highlight the complexity of such a phenomenon that does not rely on any particular assessment instrument or task.

Literature Review

At least three major conclusions can be drawn from the literature on teachers' conceptions of data use. First, studies of teachers' conceptions of data suggest teachers do not find standardized assessment data useful in their daily instruction for a number of reasons, including assessment being misaligned with instruction, the data not being received in a timely manner or in a inaccessible format, and doubts about the credibility of the data (Hamilton et al., 2007; Coburn & Talbert, 2006; Ingram, Louis, & Schroeder, 2004). Second, teachers draw from a wide variety of data in their instructional decision-making (Coburn & Talbert, 2006; Datnow et al., 2012; Goertz et al., 2009; Herman et al., 2006; Hoover & Abrams, 2013; Ingram et al., 2004; Kerr at al., 2006). Third, data from teacher-developed, classroom-based assessments are not usually considered in data use research because the data are not standardized outside the

classroom (e.g., Mandinach, Honey, & Light, 2006; Wilson, 2004). In fact, all assessment data other than *standardized* assessment data are typically lumped into an "other" category that includes information arising from practice, habits, intuition, and anecdote (i.e., not gathered in systematic ways) and that reflects a teaching practice that is predominantly ideologically driven, subject to fads and the "method du jour" (Lasley II, 2009; Supovitz & Klein, 2003). The reality is that the notion of assessment reflects a broad array of different instruments and processes that use student-generated data to inform a range of educators (teachers, administrators, policymakers) for different purposes (Supovitz, 2012).

Defining Assessment and Data

Formative assessment is a pervasive topic in the DDDM literature and many have written about teacher data use under this term (Supovitz, 2012). However, it is not the definition of assessment used in this paper because "this term is overly used and research on it entangles many important distinctions about the instruments, the processes, and the purposes of assessment" (Supovitz, 2012, p. 3). Rather, conceptualizing assessment as collecting information continuously through unconventional means, in addition to conventional, formal, and often less frequent means, broadens the meaning of the term, since this conceptualization does not necessarily rely on associating the term with any particular assessment instrument or task (Moss, 2008). This conceptualization thereby effectively renders assessment as ''a normal, ubiquitous part of all social interaction'' (Jordan & Putz, 2004, p. 346). Data, then, are the product of such interactions.

There is value in parsing out different types of assessment to make sense of the kinds of assessment data teachers find useful for improving teaching and learning. Maria Ruiz-Primo and

Erin Furtak (2007) describe assessment in service of those purposes as a range of activities that exist along a

> continuum determined by 1) the premeditation of the assessment moment, 2) the formality of means used to make explicit what students know and can do, and 3) the nature of the action taken by the teacher (the characteristics of the feedback). The continuum then goes from *formal* assessment in one end to *informal* assessment on the other (p. 58, emphasis added).

Formal assessments are typically thought of as curriculum-embedded assessments that focus on some specific aspect of learning. They are documented assessments (Jordan & Putz, 2004) that generate a record of student learning and often inform student grades. The timing of these kinds of assessments may be more frequent within a given unit, such as homework, in-class worksheets, and quizzes, or less frequent, such as performance assessments, laboratory exercises, and unit tests (Bell & Cowie, 2001; Ruiz-Primo & Furtak, 2007; Supovitz, 2012). They enable teachers to step back at key points in instruction to check student understanding and plan next steps (Ruiz-Primo & Furtak, 2007).

Conversely, *informal* assessments are more improvisational, can take place in any student-teacher interaction at the whole-class, small-group, or one-on-one level, and are embedded in and arise out of the learning and teaching activities at hand (Bell & Cowie, 2001; Moss, 2008; Black, 2009; Ruiz-Primo, 2011; Leahy et al., 2005). Data gathered from informal assessments take many forms including a) oral (e.g., student questions, comments and responses, group discussion), b) behavioral (e.g., observation of students carrying out lab procedures, body language of students during lecture) and, c) written (e.g., notes in science notebooks, bell-ringers or quick-writes including graphic evidence) (Eisenkraft, 2004). More likely than not, informal

assessments are unrecorded (Jordan & Putz, 2004; Ruiz-Primo, 2011) and data on student performance are not counted in students' grade (although some behaviors and interactions may be counted toward grades not related to academic achievement like participation grades). The timeframe for interpretation is more immediate when compared to formal formative assessments; an incorrect response or unexpected question can trigger an immediate adjustment in instruction that may include further assessment events (e.g., following up on student questions) or modifying instructional approach (e.g., repeating activity, conducting a demonstration). While informal formative assessment does not focus on conventional, standardized means of collecting and interpreting information about students, it is consistent with the purposes of sound educational assessment practices and, in that, it supports instructional decisions based on inferences made about students (Moss, 2008; Ruiz-Primo, 2011).

Influences on the Kinds of Data Teachers Find Useful

It has been estimated that teachers make over 1,000 decisions per day in the course of teaching (Jackson, 1968). While teachers' ongoing curriculum adaptations may seem to focus on instructional decisions that pertain to the immediate goals of a particular activity or class (e.g., deciding whether to spend an extra day on a concept or re-group students differently), these decisions are reflective of their understanding of a larger learning trajectory in the context of unit or yearly goals (Ruiz-Primo, 2011), underlying teaching/pedagogical philosophies and goals (Clough, Berg, & Olson, 2009), beliefs about the appropriateness of content and materials (Olson & Clough, 2011), theories of students learning (Clough et al., 2009), and teaching experience and expertise (Clough et al., 2009; Copeland, Birmingham, DeMeulle, D'Emidio-Caston, & Natal 1994; Westerman, 1991).

Not surprisingly, the factors that influence teachers' instructional decisions are often the

same factors that influence teachers' perceptions of data they find useful for instruction. Coburn and Turner's (2011) literature review references studies on teachers' attention to and interpretation of data, explaining that new information is always understood through the lens of what they already know and believe, which, in turn, influences how data is encoded, organized, and interpreted (Greeno, Collins, & Resnick, 1996; Spillane & Miele, 2007). There is an ample amount of research that shows individuals attend more closely to test scores or assessment information that confirms pre-existing beliefs and discount the data that challenges these beliefs (Coburn, 2001; Coburn et al., 2009; David, 1981; Young & Kim, 2010). For example, in one study about a workshop to improve teachers' data use, the authors found that teachers' beliefs about the appropriateness of content covered on an assessment influenced their interpretation of the utility of data (Murnane, Sharkey, & Boudett, 2005). In another study, Ingram et al. (2004) suggested that teachers' conceptions of the evidence that is useful for instruction are related to their own benchmarks of teaching effectiveness and personal goals for the education of students. These indicators included student behavior and feedback in class and student success in college and beyond, leading them to discount data that did not provide evidence of these goals. Research on data use also suggests that teachers' attention to and interpretation of data is constrained by a number of organizational and political contexts including time, district leadership, curricula and pacing guides, and history of accountability systems (Coburn & Turner, 2011; Ingram et al., 2004; Marsh, Pane, & Hamilton, 2006). However, because the focus of much data use research is on standardized assessment data, the connections between these influences and other kinds of data remain virtually unstudied.

Of particular interest in this study, and of little mention in the data use literature, are teachers' beliefs about the role of students in data use. The Institute for Educational Sciences'

practice guide on *Using Student Achievement Data to Support Instructional Decision Making* (Hamilton et al., 2009) recommends that educators teach students to examine their own data and set learning goals as an effective means to monitor student learning. Scholars stress that in order for formative assessment to work, students must have clear expectations and criteria for performance and receive feedback that is timely, specific, and constructive (Black & Wiliam, 1998; Heritage, 2007; Shepard, 2005). Tittle (1994) recognized the importance of both students and teachers as users of assessment information:

> A cognitive constructivist perspective … suggests that teachers and learners construct schemas or integrate representations from assessments into existing views of the self, of teaching and learning, and of the curriculum, broadly construed. These interpretations include knowledge and beliefs and may also result in intents to use and actual use of assessments (p. 151).

While there is considerable research on the nature of feedback provided by teachers to students and some literature on how students perceive and use their own data, there is little empirical evidence of the expectations for data use that teachers have for students and how it relates to the kinds of data teachers both generate and use in their instruction as well as share with students.

Clearly, there is still much to learn about the kinds of data teachers find useful in their instruction and why. This paper contributes to this line of inquiry by explicitly (a) broadening the definition of assessment and data to be inclusive of any data teachers might find useful in their instruction and (b) examining what makes such data useful or not useful. I also propose that there is a hidden curriculum of data use evident in high school classrooms that warrants further investigation. Finally, I conclude that the future of data use research must treat data use as a practical problem, best understood from the ground up.

Method

The data reported here come from a comprehensive field study of secondary science teachers from two high schools in a medium-sized district (~12,000 students) in the Midwest. The overarching aim of the study was to examine the phenomenon of instructional decision making from the teachers' perspective. All 23 science teachers at both schools were invited to participate in the study and 16 agreed. Two of the sixteen teachers requested only to be observed and not interviewed. Each teacher was observed teaching for 2–4 class periods (each class period was 55 minutes) and each interview lasted approximately 45–60 minutes. In addition, I spent several hours throughout the study in informal conversation with teachers over lunch and in planning periods talking about their instruction.

Participants

Teachers that participated in this study spanned all major content areas and levels of high school science education including Introduction to Physical Sciences (IPS), Biology (Bio) I, Bio-technical engineering, Advanced Placement (AP) Bio, Chemistry (Chem) I, Chem II and AP Chem, and Physics I and AP Physics. Some of these teachers were teaching a specific content area for the first time while others had many years of experience in one or more content areas. Most teachers taught within one content area but typically two different courses (e.g., Chemistry and AP Chemistry) or teachers taught different tracks of the same course (e.g., Chemistry I and Honors Chem; Bio I and Bio for ESL students). Five teachers spanned two content areas (e.g., Physics and AP Chem; IPS and AP Bio). Nine out of the sixteen teachers were female and teaching experience ranged from 1 to 33 years. The average number years of total experience was 14 years and six teachers had taught more than 20 years. All but three teachers had a master's degree and one held a Ph.D. in education. With the exception of one first-year teacher,

all teachers had taught within the district for at least two years, and most had been there for the majority of their careers.

Context

I chose high school teachers because they are an understudied population in both assessment and data use. Because extensive literature on data use emphasizes the role that organizational and political contexts play in enacting data use practices, I chose to study within one district and across one department because of the uniformity of the curricula and assessment policies across both science departments within the schools. Participation in this study was voluntary and I was not privy to teachers' interactions with other teachers who chose not to participate in the study, as well as teachers' relationships with teachers in other departments or the administrative office. This study recognizes salient relationships within the organizational and political environment in the findings as appropriate, but purposely shies away from making empirical links between specific dimensions of the context and teachers' perceptions of data, as I did not have access to the entire department.

At the time of this study, as part of a district initiative to create a unified district curricula, each content area in the science department was in the process of either developing or administering the first year of common assessments in the entry-level science courses. Additionally, all teachers teaching the same entry-level science courses (e.g., Chem I, Bio I, and Physics I) followed the same curricula with a pacing guide that prescribed the approximate number of days they were to spend on each unit. Each unit was broken down into several learning targets and teachers were expected to teach each learning target. Finally, the district mandated that 75% of any multiple choice assessment administered at the end of a unit should be the same or 'common' across all teachers of that course. To the best of my knowledge, this was

the only assessment policy in this department and prior to the implementation of this policy,

teachers had considerable individual autonomy to teach their courses as they saw fit, as well as to

develop and administer whatever unit assessment they preferred.

At the time of this study, the district was in its third year of rolling out 'early start days'

for teachers. By the third year, teachers from each content area met biweekly before school as

part of an organizational routine (Spillane, 2012) to talk—in principle—about the curricula for

the entry-level courses and, correspondingly, the data from common assessments. In addition to

the fatigue experienced from coming into work at 6:45 AM, an hour earlier than usual, most

teachers in the science department did not find much value in these collaborative spaces for two

primary reasons: 1) they felt the district had sold the idea to them that it would be a time to talk

about instruction and student learning, but that had not yet happened, and 2) the early start days

cut into 20% of their instruction as school started an hour later on those days. Some teachers also

expressed a loss of professional autonomy as a result of these meetings, feeling that establishing

the common curricula and assessments had undermined their decisions of what and how to teach.

Despite the widespread frustration, many teachers, including some of those who expressed

concern for their autonomy, remained optimistic that once a whole year of common assessments

had been administered, the focus would shift to meaningful discussions of instruction and

learning.

Data Collection and Analysis

Data collection began in October 2012 and ended in April 2013. Observations were

strategically planned to observe teachers enacting different instructional strategies (e.g., lecture,

labs, write-up days, test review, group work, etc.) with the primary criteria being that some

instructional interaction between teachers and students occurred during the class period (i.e., I

did not observe on a test day). I also observed different types of classes (e.g., honors versus regular chemistry or biology classes, larger class sizes versus smaller class sizes, lower grade levels versus upper grade levels) and at different points within a unit (near the beginning, toward the middle, prior to unit assessment). The average class size was 23 students but classes ranged from 14 to 31 students with larger classes associated with lower level and non-honors courses. Teachers usually taught five periods a day, seeing approximately 115 different students each day, possibly more if they were not teaching as many AP or honors-tracked courses. During observations, I was looking for evidence of teachers attending to or eliciting data through the course of instruction, paying attention especially to if and how teachers responded to student questions, student behavior, and whether any documented assessments were administered during instruction. See Appendix A for a sample observation protocol.

Semi-structured interview protocols guided all interviews and were tailored based on observations (See Appendix B for a sample interview protocol). All teachers were asked to describe how they knew students were learning in their class and that their instruction was going well. Follow-up questions were asked to learn what data informed their judgments and if instructional decisions were made in response to those judgments. Observations helped tailor some of these follow-up questions. For example, I observed most teachers elicit or attend to oral data from students so I asked these teachers to describe the kinds of questions they asked and what they were looking for in student answers. Similarly, teachers who appeared to be frequently scanning their classroom were asked "when you are looking around the class during instruction, what are you looking for?" Another way I inquired about the data they found useful was to ask them to "tell me about a time when you started class doing one thing but ended up doing something else," probing their responses with "how did you know that you needed to change?"

and "what was it about [that data] that lead you think that changing instruction to [a different strategy] would be better?" All teachers were asked to explain what formal assessment data they collected and for what purposes. All teachers were also asked what influenced their attention to and use of different data sources in instruction with probing questions that included class size, content being taught, preference for certain instructional strategies, student ability, and maturity.

Data collection and analysis occurred simultaneously throughout the study year (Miles & Huberman, 1994). Data analysis addresses two key questions: 1) What data do teachers find useful in ongoing instruction? 2) What accounts for differences in the kinds of data they find useful? Field notes were written for each observation session and coded descriptively for two reasons: 1) identify the types of data teachers elicited or attended to during class (e.g., oral data, written data, behavioral data, homework, quizzes, labs, performance assessments, and common assessments) and, 2) identify emerging themes and highlight areas for additional data collection. For example, since early observations indicated that oral and behavioral data were important data sources for teachers, later observations attended more closely to those instances and interview questions were designed to ask about the oral and behavioral data teachers were collecting. As another example, I observed a teacher engage in different data collection strategies across different classes and courses, so I shared this observation with this teacher and asked them to reflect on why certain data sources in one class are more or less helpful than in other classes. Types of data were classified as either informal or formal according to Ruiz-Primo's (2011) typology. Teachers' explanations of why they found data sources helpful or not were coded into different categories (e.g., efficiency of data collection, evidence of teaching goals, credibility of data, and beliefs about students) using grounded theory concepts and strategies (e.g., Corbin & Strauss, 1990) to make sure the interpretive process remained as close to the current data as

possible. By comparing episodes of informal and formal assessment data types with categories of explanations, the analytic theme of "expectations of students as users of data" emerged. Descriptive and analytic codes were entered into a qualitative database using DeDoose (www.dedoose.com). The names of teachers used in this article have been replaced with pseudonyms.

Results

The findings from this study illustrate a practice of data use in which teachers pursue information from multiple sources of data. As evidenced by the rich information yielded from the data analyses, it is apparent that teachers' ongoing instruction is a practice in which their goals for instruction and learning are varied; the means by which they attain those goals are also diverse, sometimes systematic and sometimes not. These results are organized beginning with informal and moving toward more formal assessments. I then discuss how expectations for students as users of data are a possible explanation for the relative emphasis on informal assessment data versus formal assessment data.

What Data Do Teachers Find Useful in Their Ongoing Instruction?

Most teachers are often mining multiple data sources for evidence of the progress or success of any given lesson or multiple lessons within a unit. The data are elicited by teachers as well as generated by students and comprise a spectrum of formal and informal assessment strategies. Every teacher reported collecting oral or behavioral data. In addition, teachers also referred to student-written data in the form of short and frequent assessments administered during class. Finally, all teachers collected data from formal assessments, including homework, quizzes, in-class work, labs, performance assessments, and unit assessments. Six teachers openly admitted that they also relied on their intuition, reporting the success of class as a "feeling I get,

that it couldn't have gone any better" (male teacher, 6 years of experience) and that "it just seems to go better when I [teach this way]" (female teacher, 1 year of experience). Notably, each of these teachers also recognized that they could not document their intuition as readily as other data sources; while "it's more intuitive than you might want to admit," teachers with more years of experience valued their experiential wisdom, because after so many years, as Mrs. Noble, a teacher with 30 years of experience explained, "you become less focused on what you're doing and more focused on what they're doing."

Informal Assessment Data

Oral data. Oral data was a, if not *the*, primary data source for teachers in their ongoing instruction. Although oral data included listening in on students working in pairs or groups, or even having whole class discussion, it was oral data elicited through question and answer strategies that mattered most to teachers. As one teacher with 25 years of experience explained, "Questioning has really become a priority for me mainly because it is one of the fastest and most efficient methods of formally assessing the students and making changes in the instruction instantly based on their responses." The answers to questions posed by teachers, the questions generated by students, and the conversations that often followed were valuable data for every teacher in this study.

The use of *strategic* questioning strategies to elicit student thinking has a substantive literature base in formative assessment and notably in science education (Duschle & Gitomer, 1997; Ruiz-Primo & Fertak, 2006, 2007; Ruiz-Primo, 2011). Duschle and Gitomer (1997) coined the term "assessment conversation" to describe specific patterns of planned interactions between teachers and students designed to reveal both conceptual and epistemological understandings. My study did not focus on assessment conversations and thus I will not make conclusions about

teachers' practices as it relates to such patterns. However, the assessment conversation framework is useful for thinking about teachers' questioning practices. Specifically, Duschle and Gitomer explained that upon *receiving* information from a student (by asking a question), the next step was for teachers to *recognize* the information provided in student answers as evidence of the cognitive and epistemic orientations of learning. I use this two-step process of *receiving* and *recognizing* information as a way to make sense of the extent to which teachers might be strategic in their question and answer strategies (even if it is not for the purposes put forth in the assessment conversation literature). Specifically, teachers' instruction was guided by information *received* from *who* was asking or answering and *recognized* data based on *how* student questions or answers were expressed.

First, data analyses signaled that teachers *received* information from particular students, based on whether or not those students asked or answered questions. Although observations suggested such purposeful instructional dialogue was present, the data collection was not long enough to confirm the existences of such patterns across all classes or teachers. Yet, teachers readily admitted in interviews that they considered some students as gauges in their class and frequently turned to such students to determine their current instructional effectiveness. Practically speaking, teachers explained that, especially in the larger classes, time and classroom space prevented them from talking to each student individually or even groups of student about their comprehension of the material. Four teachers also explained that they purposely did not call on certain students because they felt it would embarrass them, a sentiment left over from the teachers' personal experiences as students or lessons learned from earlier in the year. This strategy of identifying gauges in the classroom was also related to teachers' beliefs about the representation of student abilities in the classroom; teachers believed that certain students

reflected a particular tier of ability or achievement in the classroom and *recognized* that the response of those students reflected the understanding of all students judged to be like them. One teacher explained that she does not get a lot of questions in her class but when she does, they tend to come from the "top kids" and she knows "if they aren't getting it, then no one is getting it." Another teacher gauged student comprehension of material on whether his "high participation" students were asking questions or willing to try to answer the questions he posed. One teacher described that thinking:

> In my mind I kinda have a sense for, okay, you're typically a student that picks up on things pretty well and you're typically a student that doesn't do so hot. That sounds bad, but it's the way it is. I mean it's the way it is, honestly and truly. When I am delivering instruction, I observe cues from those different types of students.

In addition, data analyses suggested that both *how* students answered questions and what kinds of questions students posed provided data as evidence of learning and subsequently informed ongoing instruction. Some teachers indicated that they were only looking for right answers to the questions they asked. Furthermore, several teachers interpreted the absence of questions from students as evidence that students were comfortable with the material and ready to move on. Teachers expected students to speak up for themselves in class and, even if a student was not comfortable speaking in front of the entire class, teachers explained that students should use time before and after class or school to engage in instructional dialogue with teachers. At least nine teachers in the study described in detail the kinds of questions or answers they expected from students, as well as the strategies employed to gather that information. Three of those nine teachers focused on the data supplied by questions students asked while the remaining six valued the data from the answers students provided to the questions those teachers posed. In

all cases, teachers were looking for a certain level of cognitive complexity in student questions and answers that provided evidence of mastery. The following examples illustrate these two foci.

One teacher described an instance from an individual lesson on the anatomy of the human arm where behavioral data signaled that students were struggling. This teacher then began posing specific questions to students, starting with one that asked students to apply the information they were already supposed to know about how the anatomy of the arm is related to its motion. When students failed to apply, he asked another question that gauged their comprehension of different muscles and bones. When students could not demonstrate comprehension, he then asked them to recall the different muscles and bones. By strategically employing questions at different cognitive levels, this teacher was able to isolate the specific gap in student knowledge. This not only appropriately modified instruction for that particular class but also provided a compass (Jennings, 2012) for starting instruction in the same course later in the day. In another example, Mrs. Noble explained simply that "if a student cannot tell it to me 'grandma style'," that is, if a student could not explain the concept to her in a way that his/her grandmother could understand, then she knew mastery had not been attained and attention to that particular concept was required.

Another teacher, who focused on the questions students asked, described her desire for students to ask a specific type of question that illuminates student thinking. She explained,

I want a specific question, rather than 'I don't understand any of this;' [I want] 'I don't know the difference between membrane potential and active potential—I'm getting those two confused.' I've instructed them to give me a specific question [because] it shows me you are actively engaged in processing what it is you're looking at.

This teacher not only employed this strategy to inform her ongoing instruction but also because she wanted students to actively reflect on their own learning and to be able articulate the gaps in their knowledge. Three teachers in this sample who taught students in both entry level and advanced courses suspected that the capacity to expect a certain level of questions and answers from students might be related to student maturity and that, admittedly, they tended to ask more direct, lower-level cognitive questions and attended more so to right answers than explanations of reasoning in their entry level courses than their advanced courses.

  Behavioral data. Overall, student behavioral data was used by all but one teacher to gather evidence of motivation, engagement, and participation as a proxy for student learning and a compass for instruction. Teachers reported that high participation and engagement, as described by a number of behavioral cues, signaled to teachers that students were learning. These verbal and non-verbal behaviors included eye contact with teacher or with notes on screen; facial expressions in response to what students saw on the screen, heard from the teacher, or read in front of them; whether heads were up or down on the desk; nodding, murmuring, talking to other students out of turn; and so on. If behavioral data provided evidence of engagement, motivation, and participation as a proxy for student learning, then observations of teaching should highlight teachers' attention to such data. Classroom observations confirmed teachers' belief in this data; teachers focused a lot of energy in their instruction in keeping students engaged and motivated. Teachers used humor, spoke encouragingly to students, related topics to teenage culture, and conducted lively demonstrations to keep their attention. Teachers also called students' attention to particular behaviors (e.g., telling a student to put her math notebook away) that might be interfering with students' learning. Observations and interviews also revealed that teachers used

behavioral data to slow the pace of instruction within a class and to inform the pace of instruction for the same class throughout the day.

Some teachers, notably those with considerably more years of experience, valued behavioral data to a significant degree over other forms of data and could describe with remarkable detail specific, non-verbal behaviors they were looking for to inform their instruction. For example, Mrs. Noble, a veteran teacher of 30 years at elementary, middle, and high school levels, provided the following description of behavioral data she found important about any one particular student:

> If you said, 'Give me any data I had on [a student],' here's what I would say. I could tell you, is she on time for class? Is she prepared for class? Is she attentive in class? Does she ask questions in class? Is this student turning in assignments? Are they neat? Does she or he pull them from a specific location? Do they know where the work is at all times? Are they on task during the class period? ...This person's bothering people. I guess that would be a form of being on task. Does this person seem to have visual problems because of where they sit in the room? Are they always asking me to repeat? Do they follow oral directions well or not? Do they follow written directions well or not? [On a quiz], for example, if there's a good answer first, but the best answer is last. I can tell who picks quickly. Sometimes, I can tell, just by the number of erasures, they're constantly erasing and changing answers that they're indecisive.

Mr. Huntsman, another veteran teacher with 25 years of experience, distinguished between two types of 'silence' in his room, one of which signaled that learning was happening while the other indicated that students were no longer engaged. He described the silence in which learning occurs:

"You just know that the silence in the room, it's a different silence. There's more motion. There's more activity—which doesn't necessarily mean standing up and moving around activity, but there's eye contact with me, there's looking down at the paper and looking back up, there's nodding, there are smiles. If I say something and oh, there's a smile, okay, so-and-so was listening and caught that joke. It might have been very subtle. It might have been not so subtle. There's more movement. There's more energy."

Written work. Six teachers described students' written work, such as bell ringers, question of the day, quick-writes, exit slips, and class polling strategies, as useful for guiding ongoing instruction. These data collection methods were sometimes routinized—administered at the beginning or end of every class—or were spontaneously administered as an outgrowth of instruction and embedded within particular lessons. These quick assessments provided some documentation of student learning but were rarely collectedly and hardly ever used in students' grades except for completion or participation points. These teachers found value in these assessments because they provided immediate feedback on what students learned that very day or could remember from the day before, providing information about what their instruction had accomplished and where to start the next lesson. Teachers often used these assessments to activate prior knowledge as well as engage students in peer assessment by having students solve a few problems or answer some questions and then share their answers with a classmate.

Formal Assessment Data

Homework, quizzes, and in-class work. Only two teachers talked about homework as a useful tool for their instruction and only one of those teachers stated that student performance on homework guided their instructional decisions. Mr. Truman, a veteran teacher with 30 years of experience, explained that he rarely attended to student answers but instead focused on student

behaviors throughout the process of completing homework, including 1) how many students came to see him before or after school with questions regarding homework, 2) how many students were looking at the answer keys in class and what keys they were looking at (e.g., procedural or numerical), and 3) how many students did not turn it in. Otherwise, teachers did not find homework useful because they could not be sure who completed it and often suspected cheating. Furthermore, many teachers felt homework, especially homework that was given and collected on a regular basis, penalized students who did not learn as quickly as others and created an unrealistic amount of work for them to grade on a daily basis. As a result, three teachers in the study rarely, if ever, gave homework, while the remainder gave it but only intermittently or handed it all out at the beginning of the unit and collected it right before the unit test to be graded for completion.

All teachers administered quizzes throughout the unit on a daily, weekly, or less frequent basis. However, only four teachers found quizzes to be moderately useful to guide ongoing instruction and utility was limited by a belief that students "just study for the test " and a recognition that quizzes were more helpful in the beginning of the year when teachers did not yet know their students. Six teachers talked about student in-class work (e.g., worksheets) as being useful for their instruction because they saw the direct application of what was learned and had "complete control over the environment in which I'm assessing." The data from these more frequent assessments guided immediate and future instruction, as well as gauged student mastery and ability. Many teachers also explained that quizzes and homework were used to determine ability level of students in the classroom early on and thus guide the behavioral and oral data collection strategies they used throughout the year.

Labs. Teachers were distinctly divided on the value of data from science labs. Approximately half the teachers described labs as useful sources of information for instruction. These teachers conducted frequent labs within a unit, often using the lab as an instructional strategy to introduce concepts and help students make connections between ideas within or across units. Thus, they gathered valuable information about student misconceptions to help inform instruction. Other teachers found the use of labs conducive to diverse learning styles and used them towards the end of a unit to provide students an alternative way to demonstrate mastery of material and inform the teacher what skills and concepts needed review before the unit assessment. These teachers reported that observing students conduct the lab, as well as their write-up of the process, analysis, and conclusions, was helpful for instruction.

The rest of the teachers described labs as a poor source of information for instruction. Teachers explained that labs were too time consuming to grade for anything other than completion and expressed the concern that labs were "overdone" and thus useless because "students don't know what they are doing them for." They only saw value in doing labs and demonstrations because it "made science come alive" and thus engaged students in instruction. These teachers also recognized that some students learned better with labs but did not find them particularly useful in their ongoing instruction. There were no other salient differences in these teachers (e.g., experience, content area, teaching philosophies, preference for other data types) that could readily explain this difference in preference for labs.

Performance assessments. Not all teachers administered performance assessments, but those that did found them to be less useful for ongoing instruction and more pertinent to instructional planning in future units or the same unit in the next year. Performance assessments

were generally administered at the end of a unit and teachers had little or no time to re-teach any concepts or skills not mastered before moving on to the next unit.

Common assessments. All science teachers administered unit tests at the end of a unit, or at least twice a semester. For teachers of advanced level or AP courses, they did not find unit test data useful in their instruction. In the entry-level physics and chemistry courses, teachers administered the *common assessment* at the end of every unit (biology was still in the process of developing their assessment). Although the teachers in the department had developed the common assessments, and thus 'owned' the data (Shepard, 2005), teachers expressed considerable frustration with the way the common assessments had been implemented (they were administered on a computer) and were wary of the district using the results to evaluate teachers, even though they were told this would not be the case. Not surprisingly then, none of the teachers found the data from common assessments useful for their ongoing instruction. However, teachers did report that having a common assessment helped them ensure their instruction was aligned to both the assessment and the learning targets. This was especially helpful for beginning teachers.

Overall, the exercise of alignment initiated discussions about instruction. Several teachers expressed hope that one day these assessments would guide remediation and enrichment opportunities under the target-based learning framework and, perhaps, allow teachers to focus instruction on missed items from the previous year. However, the teachers in honors-level courses did not share this sentiment because they interpreted their students' exceptionally high scores on the assessments as evidence that the assessment did not challenge their students and only tested the "bare minimum." Three teachers dismissed the common assessments altogether,

explaining that the test did not do anything but confirm what they already knew about their students.

Much like previous research on data use, these findings support the claim that standardized test data (e.g., common assessments) are not useful for instruction and illustrate that the data teachers find useful for instruction vary widely and are dependent upon a number of factors. Some factors are quite obvious—time (as it always does) constrains their capacity to collect and use data, whereas other factors, like teaching philosophies, personal experiences, and beliefs about data, are much more nuanced and warrant further examination. However, one particular finding that stands out is the uneven priority given to informal assessment data over formal assessment data. That is, despite the fact that teachers administer homework, quizzes, labs, performance assessments, unit tests, and so on to students on a regular basis, the teachers, but with few exceptions, do not find data from these assessments to be regularly useful for ongoing instruction. This finding is quite contradictory to the expectations in the data use literature and more broadly to the mission of assessment in general, as an integrative process to inform teaching and learning (National Research Council, 2001). Admittedly, teachers gave many assessments to motivate students and provide grades. Yet, more often than not, the data that *teachers* did not find useful for their instruction were the very data they expected *students* to use to monitor their own learning. This finding emerged very early on, after my first day of observation, when I asked a teacher what the point was of the target quiz she had just administered; she told me, "Oh, it's for them. Student accountability."

Students as Users of Data

All teachers administered many different kinds of formal assessments because they assumed the students should be and were users of their own data. Teachers spoke of a variety of

assessments that catered to different learning styles and met students' need to demonstrate learning in different ways, all with the hope that students would be able to use assessment data to monitor their own learning. Teachers linked these expectations to their educational goals for students, that included becoming resourceful learners independent of the teacher, developing as critical thinkers capable of identifying what is important in a given content area, and, practically speaking, getting them ready for college. As one teacher told her students, "This is for you. This is to help you get ready for college. In a couple years, you will not have anyone to help you."

"It's not for me; it's for them," explained one teacher as she told me about her completion policy on homework. The teachers that gave homework expected students to self-monitor their completion of the homework by identifying areas where they were struggling, bringing questions to teachers outside of class, checking the answer keys available, or using reserved time within the classroom to get help with their learning. Quizzes, administered once or twice a unit, or on a weekly or even daily basis, were graded for accuracy only to hold students accountable and generate grades for students. Ultimately, though, teachers explained to their classes or told me in interviews that they expected that students would analyze their performance to identify gaps in learning and focus their efforts in closing those gaps for the unit summative assessment. Expectations were the same for labs, performance assessments, and even unit tests.

While all sixteen teachers expressed these expectations, at least seven teachers assumed responsibility for teaching or modeling for students how to become users of their own data. These teachers described how they designed their assessments to provide feedback to students or made their expectations for student performance explicit and transparent. These teachers sought to close the feedback loop for students, providing them with information about the gap between their current learning and learning targets. For example, six teachers, building off the target-

based curricula that were recently implemented in each content area, administered their quizzes with learning targets printed above each question aligned with a target. The targets were worded in the form of "students can" or "I can" statements followed by a particular verb and object that articulated the expectation for mastery. When students answered items incorrectly on quizzes, they could refer back to the target the item was intended to address and, in theory, have instant feedback about the targets they had not yet mastered.

On top of conveying data for student use, two veteran teachers described in detail how they modeled their *own* data use practices to help students become effective users of their own data. Mrs. Noble, for whom behavioral data was her biggest driver in instruction, made her expectations for behavior that exemplified students taking responsibility for their own learning explicit to students at the very beginning and throughout the year. She explained:

> I train them. I model them. Here are the steps, A, B, C, D. When it comes to the fact that they end up coming into my classroom after school, then I personally say, "Show me. Show me. Show me. Show me." I list all these things. They either can or they cannot. One of them being the agenda [I will ask a student to show it to me], and if they say, "Well, I never fill it out," well, problem A. Then, I say, "Show me the very last paper we did in class." If they can produce it, there's organization. If there's scuffling around their bag tossing papers, they don't have that. If they cannot tell me their login to the online website; they should be able to tell me just like that. What's my password? That gives me data.

Ms. Simpson, a teacher with 25 years of experience and for whom oral data heavily informed her instruction, also described at length her expectations for students to compose a specific type of question that clearly identified the current gap(s) in their learning. She said her

students know that she will not answer broad and vague questions. In her interview, she explained how she talked to students:

> If you have questions, be specific about what you are saying. I need to know what you're thinking about what I'm saying because if you just say a broad statement then you're asking the same type of question a kid would ask if they hadn't been there for 50 minutes, and I'm going to reteach all of that to you. And you're a few months away from college and you're not going to go to your professor and say this. You need to be more specific about what you are saying and go back.

Ms. Simpson explained that students usually struggled at first to identify gaps in their learning but that it "absolutely improved" as the year progressed. By the end of the year, she typically finds that questions from students are less frequent because once students can identify the gaps in their learning, they can figure out the answer to their questions on their own. Regardless of how explicit their expectations were, because teachers assumed students would be users of their own data, they also then assumed students should and could advocate for themselves. As one teacher explained it, "if you don't ask questions, I'm left assuming that we're good to go and we can move on to the next thing." The scope of this study did not include asking students about the use of their own data and so it is unknown what the likely consequences are of making such assumptions.

Discussion

The findings from this study highlight the fact that teachers' instructional practices constitute a complex interaction between curriculum, pedagogy, and their views of students in which simultaneous, and not mutually independent, considerations of data, goals, professional values, and external pressures are in play (Schwandt, 2005; Young, 2005). In recent years,

scholars have become especially keen on purposefully examining the inherent complexity of data

use. Coburn and Turner have spearheaded many of these efforts, starting with their literature

review in 2011 that synthesized various strands of research and theory related to data use in

schools, highlighting the complexity of pathways between the adoption of a data-use intervention

and the attainment of desired outcomes, and clarifying the ways in which different groups of

actors and stakeholders interact with data and one another. Coburn and Turner (2012) also edited

a special issue of the *American Journal of Education*, titled the "Practice of Data Use*"*,

comprised of six articles that presented research on the practice of data use at multiple levels.

This issue attended closely to the role of organizational and institutional contexts and the

dynamics of social interaction and highlight potentially useful theoretical frameworks and

methodological approaches to studying data use. In a 2012 special issue of *Teachers College

Record*, Turner and Coburn commissioned several reviews of existing empirical literature on

data use interventions to expand our understanding of the relationships between processes,

contexts, and outcomes of these interventions in hopes, again, to move the field toward more

systematic and theoretically grounded research. This article contributes to this research agenda

by highlighting the patterns and nuances in the kinds of data teachers find useful.

Broadening Our Framework of Assessment and Data

It is not enough, however, to just recognize that teachers rely on multiple kinds of data in

their instruction or to conclude that efforts to implement data use interventions are mediated by

the fact that teachers do not value standardized assessment data in instruction. Rather, research

on data use must be intentionally inclusive of multiple data sources in order to understand how

data, in any form, are translated through individual practice or collaborative interactions into

actionable evidence for instruction. Similarly, we must resist the temptation to categorize

assessment practice purely by broad purposes (e.g., formative or summative). Assessment takes many forms, formal and informal (Ruiz-Primo, 2011), documented as well as inherent in interaction (Jordan & Putz, 2004). Furthermore, in reality, teachers (and other actors and stakeholders) use the same assessment for many different purposes, both directly and indirectly related to instruction. For example, on homework assignments, Mr. Truman valued his observation of student behaviors in completing the homework rather than the actual student responses. In another example, Mr. Heinz's multiple choice and short answer mid-unit quizzes provided a necessary grade for students in the middle of the unit. While it was only the short answer questions that provided him data about student thinking, he expected the multiple choice questions to provide information to students about learning targets they had not yet mastered. Also, Ms. Simpson's question and answer strategy served the multiple purposes of helping students identify their own learning needs, motivating students to engage in their own learning, and guiding her instruction.

One way to better capture the means by which data become useful and are used is to focus exclusively on expert teachers. Previous research on expert and novice teachers has shown that an expert teacher's understanding of classrooms is more comprehensive, characterized by Copeland et al. (1994) as an "increase in the 'quantity and complexity of linkages' among ideas and by a shift in the focus of these linked ideas toward issues more central to classroom teaching" (p, 176; see also Westerman, 1991). In this study, teachers with substantially more years of experience (e.g., Ms. Simpson, Mrs. Noble, Mr. Truman, and Mr. Huntsman and others) demonstrated a more nuanced attention to data than their less experienced counterparts.

Broadening our framework of assessment and data also invites the critique of systematic data collection by teachers as a principle of quality data use and a key step in improving

instruction (Boudett, City & Murnane, 2013; Mandinach, Gummer & Light, 2006; Kowalski, 2009). Currently, the data use literature highlights the promises of systematically collecting data in order to effectively engage in a cycle of continuous improvement, a concept borrowed from manufacturing principles (Deming, 1985; Senge, 1990); but systematically collecting data does not in and of itself lead to improved instruction if the data being collected is not being used to inform instruction. Given that teachers utilize multiple data sources, attended to at varying frequencies, coupled with the practical realities of teaching, the assumption of systematic data collection is problematic: 1) Is it even a realistic expectation for ongoing instruction? 2) If it is, what does it mean to be systematic? 3) How does being systematic contribute to improved instruction and learning independent of other practices we know to be related to improved teaching and learning (e.g., feedback)? For example, if teachers rely heavily on question and answer data but direct their attention to particular types of students, believing them to be effective gauges for the progress of the class, then how might we systematize that process and connect it to student achievement?

Examining the Hidden Curriculum of Data Use

Currently, the data use discourse largely omits one important group—the students who generate the data, which are the focus of data use practices (Hamilton, 2011). As Hamilton (2011) stated, "As co-constructors of their educational experiences, students play a key role in influencing the quality and nature of the learning activities in which they engage both inside and outside of the classroom" (p. 207). Whereas Hamilton suggested that the very same norms and climates that impact teacher data use practices may also mediate students' data use practices, I also contend that the assumption of data use by students as suggested in this study, especially at the high school level, mirrors a common phenomenon observed in higher education known as the

hidden curriculum. According to Skelton (1997) the hidden curriculum is "[t]hat set of implicit messages relating to knowledge, values, norms of behavior and attitudes that learners experience in and through educational processes. These messages may be contradictory, non-linear and punctuational and each learner mediates the message in her/his own way" (p. 188). Scholars in higher education identify the message of the hidden curriculum in any education system by its assessment procedures as

> every act of assessment gives a message to students about what they should be learning and how they should go about it. Assessment messages are coded, not easily understood and are often read differently and with different emphases by staff and by students (Boud, 1995, p.39).

Here the hidden curriculum is that students should be able to interpret and use assessment information to improve their achievement. Students who know how to play the game understand this implied message but for many the expectation may be unknown.

That teachers' preferences for certain kinds of data are linked to their expectations of students as users of their own data has implications for understanding teachers' data use practices. This finding warrants closer examination of the nature of data teachers make available to students, their beliefs about and expectations for how students will use that data, and the ways in which they make those beliefs or expectations explicit. Such an examination will also augment our understanding of what data teachers find useful in instruction and how they pursue that information. Some of these issues have been addressed in studies of teacher feedback to students, but the majority of research that suggests that, at best, feedback is hardly used (Shepard, 2005). In order to better understand how the hidden curriculum of data use is enacted and experienced in schools (both at elementary and secondary levels) researchers will need to focus on the

dynamic social interactions between students and teachers and the influence of organizational

and political contexts on those interactions (Coburn & Turner, 2012).

Conclusion: **Data Use as a Practical Problem**

This article builds on the efforts of previous scholars to make sense of teachers' data use

practices from the ground up, so to speak—that is, from a teacher's viewpoint. It provides

empirical evidence of the need to expand our thinking about what data matters to teachers'

instructional decision making and why. These findings underline the complexity of individual

teacher data use practices and suggest, like many other studies of data use, that institutional,

organizational, and even professional norms are influencing the kinds of data teachers find useful

and why. Unlike of studies of data use, however, this study does not have a theoretical objective.

There is no attempt to theorize teachers' data use practices. Rather, the intent is to reflect the

reality of daily teaching practices and the myriad of strategies teachers use to identify issues in

teaching and learning which serve a broad spectrum of goals for both teachers and learners.

'What works' for one teacher in identifying gaps in student learning and assessing the efficacy of

teaching does not serve other teachers equally well, even though they may rely on similar

methods and data from time to time. In contrast, much of the literature on data use assumes

purposes for using data can be readily identified and categorized and, more importantly, that the

goals of teaching are all (or nearly) the same (e.g., Mandinach, Honey & Light, 2006; Supovitz,

2012). Data use, then, is simply a process by which empirical evidence guides the selection of

means (e.g., instructional strategies) to solve problems in order to achieve clearly articulated

goals. This study challenges this view, suggesting that the process of data use itself is vetted

through internal and external influences that cannot be separated from the process itself. There is

no normative process of data use to which teachers might ascribe; rather, much like the practice

of teaching itself, it is practical, it is bounded by time and place, and it requires simultaneous

consideration of material, social, cultural, and historical contexts (Spillane & Miele, 2007); and

individual world views.

REFERENCES

Bell, B., & Cowie, B. (2001). *Formative assessment and science education*. Dordrecht, The Netherlands: Kluwer.

Black, P. (2009). From frequent testing to interactive dialogue: Diversity and challenge. Guest's editor introduction. *Journal of Multidisciplinary Evaluation*, *6*(12), iii–vi.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5(1), 7–73*.

Boud, D. (1995). *Enhancing Learning Through Self-assessment*. London: Kogan Page.

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 edition. Cambridge, MA: Harvard Education Press.

Clough, M. P., Berg, C. A. & Olson, J. K. (2009). Promoting Effective Science Teacher Education and Science Teaching: A Framework for Teacher Decision-Making. *International Journal of Science and Mathematics Education*, 7(4), 821–847.

Coburn, C. E. (2001). Collective sense-making about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145–170.

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., & Talbert, J. E. (2006). Conceptions of Evidence Use in School Districts : Mapping the Terrain. *American Journal of Education*, *112*(4), 469–495.

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

Copeland, W. D., Birmingham, C., DeMeulle, L ., D'Emidio-Caston, M., & Natal, D. (1994). Making meaning in classrooms: An investigation of cognitive processes in aspiring teachers, experienced teachers, and their peers. *American Educational Research Journal, 31*(1), 166–196.

Deming, W.E. (1986). *Out of crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study. Detente Possible?', *Theory Into Practice, 44*(3), 234–244.

Duschl, R. & Gitomer, D. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, 4, 37–73.

Eisenkraft, A. (2004, *April*). *How do we know what they know*. Paper presented at the *Florida Association of Science Supervisors Meeting*, Orlando, FL.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.) *Handbook of Educational Psychology* (pp. 15–46). New York, NY: 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). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/.

Hamilton, L.S. (2011). The role of students in data use: Commentary on Coburn and Turner's Research on Data Use: A Framework and Analysis. *Measurement: Interdisciplinary Research & Perspective*, *9*(4), 207–210

Hamilton, L.S., Stecher, B.M., Marsh, J.A., Sloan McCombs, J., Robyn, A., Russell, J., et al., (2007). *Standards-based accountability under No Child Left Behind*. Santa Monica, CA: Rand Corporation.

Heritage, M. (2007). Formative Assessment : What Do Teachers Need to Know and Do? *Phi Delta Kappan*, *89*(2).

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision-making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258–1287.

Jackson, P. (1968) *Life in classrooms*. New York: Holt, Rinehart & Winston.

Jennings, J. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record Volume 114*(11), 1–23

Jordan, B., & Putz, P. (2004). Assessment as practice: Notes on measures, tests, and targets. *Human Organization, 63*, 346–358.

Kowalski, T. J. (2009) Evidence and decision making in professions (pp. 3-20). In T. J. Kowalski & T. J. Lasley II (Eds.) *Handbook of Data-based Decision Making in Education*. New York, NY: Routledge.

Lasley II, T. J. (2009) Using data to make critical choices (pp. 243-258) In T. J. Kowalski & T. J. Lasley II (Eds.) *Handbook of Data-based Decision Making in Education*. New York, NY: Routledge.

Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational leadership, 63*(3), 18–24.

Mandinach, E. B., Honey, M., & Light, D. (2006, April). *A Theoretical framework for data-driven decision making*. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making Sense of Data-Driven Decision Making in Education: Evidence from Recent RAND Research*, Santa Monica, Calif.: RAND Corporation, OP-170-EDU, 2006.

Moss, P. A. (2008). Sociocultural implications for the practice of assessment I: Class- room assessment. In P. A. Moss, D. Pullin, J. P. Gee, E. H. Haertel, & L. J. Young (Eds.), *Assessment, equity, and opportunity to learn*. Cambridge University Press: New York.

Murnane, R. J., Sharkey, N. S., & Boudett, K. P. (2005). Using Student-Assessment Results to Improve Instruction: Lessons From a Workshop. *Journal of Education for Students Placed at Risk*, *10*(3), 269–280.

National Research Council. 2001. *Knowing what students know: The science and design of educational assessment*. Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., and Glaser, R., editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press

Olson, J.K. & Clough, M. P. (2001). Technology's tendency to undermine serious study: A cautionary note. *The Clearing House*, *75(*1), 8–13.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The Role of instructional dialogues in assessing Students for science learning. Special Issue in Assessment for Learning. *Studies of Educational Evaluation, 37*(1), 15–24.

Ruiz-Primo, M. A. & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment*, *11*(3–4), 205–235.

Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84.

Schwandt, T. (2005) The Centrality of practice to evaluation. *American Journal of Evaluation, 26*(1), 95–105.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.

Shepard, L. A. (2005, October). *Formative Assessment: Caveat Emptor*. Paper presented at the ETS Invitational Conference 2005, New York.

Skelton, A. (1997). Studying hidden curricula: Developing a perspective in the light of postmodern insights. *Curriculum Studies*, *5*, 177–193.

Spillane, J. (2012). Data in Practice: Conceptualizing the Data-Based Decision-Making Phenomena. *American Journal of Education*, *118*(2), 113–141.

Spillane, J. P. & Miele, D. B. (2007). Evidence in practice: A framing of the terrain (pp. 46–73). In P. A. Moss (Ed.), *Evidence and decision making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell.

Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: How schools systematically use student performance data to guide improvement*. Philadelphia: Consortium for Policy Research in Education.

Supovtiz, J. (forthcoming) Getting at student understanding-The key to teacher use of assessment data. *Teachers College Record*, *114*(11), 1–29.

Tittle, C. K. (1994). Toward an educational psychology of assessment for teaching and learning: Theories, contexts, and validation arguments. *Educational Psychologist*, *29*(3), 149–162.

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

Westerman, D. (1991). Expert and novice teacher decision making. *Journal of Teacher Education, 42*(4), 292–305.

Young, V. M. (2005). *Teachers' use of data: Embedded Contexts and Institutional Logics* (Doctoral dissertation). Stanford University. Retrieved from *ProQuest Dissertations and Theses* (UMI Number 3186431).

Young, V. M. (2006). Teachers' Use of Data: Loose Coupling, Agenda Setting, and Team Norms. *American Journal of Education, 112(4),* 521–548.

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: a literature review. *Education Policy Analysis Archives, 18* (19). Retrieved from http://epaa.asu.edu/ojs/article/view/809.

PAPER 3

THE WAY IT IS AND THE WAY IT SHOULD BE: PRE-SERVICE TEACHERS'

CONCEPTIONS OF ASSESSMENT

Teachers' conceptions of assessment are key factors that influence classroom decisions, including the ways they use assessment information to determine the effectiveness of instruction (Coburn & Talbert, 2006; Ingram, Louis, & Schroeder, 2004; McMillan, 2003) and improve educational processes (Borko & Putnam, 1996; Shepard, 2000). Research also indicates that teachers' beliefs about assessment (Delandshere & Jones, 1999) and student performance data (Coburn & Talbert, 2006) mediate reform efforts to improve teaching and learning. Further, international and cross-cultural studies of teachers' conceptions of assessment provide evidence that teachers' conceptions are related to the policy contexts in which they practice (Brown & Hirschfield, 2008; Brown, Hui, Yu, & Kennedy, 2011; Brown, Lake, & Matters, 2011). More recently, scholars have highlighted that understanding *pre-service* teachers' conceptions of assessment may be an important factor in pre-service teachers' assessment literacy (Brown & Hirschfield, 2008; Brown, 2011; Daneen & Brown, 2011; DeLuca, Chavez, & Cao, 2012; Smith, Cowie, Gilmore, & Hill, 2012).

Scholars have attempted to model both pre-service and in-service teachers' conceptions of assessment as well as to pinpoint the links between those conceptions and (a) the policy contexts in which they work (Brown, 2008, 2011; Brown & Michaelides, 2011; Brown & Remesal, 2012; Gebril & Brown, 2013; DeLuca et al., 2012, Munoz, Palacio, & Escobar, 2012); (b) individual perspectives on teaching and learning (Brown, 2008); (c) assessment practices (Brown & Hirschfield, 2008; Brown, Kennedy, Fok, Chan, & Yu, 2011; McMillan, 2001); and (d) academic achievement (Brown & Hirschfield, 2008). While there are multiple published

frameworks that describe possible conceptions of assessment (e.g., DeLuca et al., 2012; Harris &

Brown; 2009; Leighton, Gokiert, Cor, & Heffernan, 2010), this paper focuses on the use of the

widely cited framework developed by Gavin Brown (2008), Teachers' Conceptions of

Assessment (TCoA), that was developed through studies of teachers in New Zealand and

Australia but has since been applied in many international contexts. The purpose of this study

was to examine a sample of pre-service teachers' (PSTs) in the United States to understand their

conceptions of assessment using the TCoA framework and a modified version of the TCoA-III

inventory (Brown, 2008). Substantive changes to the instrument were made because of several

qualitative pre-tests conducted prior to administration. The findings from this study indicate that

while this sample of PSTs in the United States holds conceptions of assessment similar to those

that have been reported elsewhere, including in similar policy contexts, PSTs in the U.S. also

hold a conditional conception of assessment, believing there is a way *assessment currently is* and

a way *assessment should be*. This study highlights the value in testing and evaluating validated

instruments prior to administering to new populations and supports prior research that suggests

conceptions of assessment may be a function of the cultural origins of the research and that in

different societies, somewhat different intentions and patterns exist.

Literature Review

The term *conception* is used to capture all that a teacher thinks about the nature and

purpose of an educational process and practice (Thompson, 1992; Brown, 2008). One of the

difficulties in researching teachers' conceptions of assessment is that they appear to hold

multiple and sometimes contradictory conceptions without being disturbed by such

contradictions (Cizek et al., 1995; Kahn, 2000). A major factor in this plurality of conceptions is

that assessment itself serves multiple purposes, which also may be complementary or

contradictory. For example, many would consider the belief that assessment is for improved teaching and learning to be opposed to the belief that assessment is for accountability or evaluation purposes, but this may not be the case if teachers accept the legitimacy of accountability mechanisms (Brown, Lake, & Matters, 2011).

Teachers' Conceptions of Assessment

Gavin Brown's (2008) formative work on New Zealand and Queensland teachers' conceptions of assessment produced the *Teachers' Conceptions of Assessment Framework*. Brown's framework expands on previous research on teachers' conceptions of assessment (see for example Philippous & Christou, 1997; Saltzgaver, 1983; Torrance & Pryer, 1998). The original framework offers four inter-correlated, intention-oriented conceptions of assessment that may loosely be categorized as three 'purposes' and one 'anti-purpose' (Brown, 2008); the three purposes include: 1) Assessment as Improvement of Teaching and Learning (*Improvement)*; 2) Assessment as Making Schools and Teachers Accountable for their Effectiveness (*School Accountability)*; and 3) Assessment as Making Students Accountable for their Learning (*Student Accountability)*. What Brown (2008) terms as an anti-purpose reflects the belief that that 4) Assessment is fundamentally irrelevant to the life and work of teachers and students (*Irrelevant*). All told, there are nine distinct factors in the framework, four which contribute to *Improvement*, two toward *Student Accountability* and three toward *Irrelevant*. Refer to Table 2 to see each of the factors that contribute to each purpose or anti-purpose.

Table 2. Original Teachers' Conceptions of Assessment (TCoA) framework (Brown, 2008).

Major Conception	Summary	Contributing Factors
1. *Improvement*	Assessment is used to improve teaching and student learning by providing information to students and teachers about the effectiveness of teaching and mastery of learning taking place.	a. Assessment is *valid* b. Assessment *describes* student learning c. Assessment improves *teaching* d. Assessment improves *learning*
2. *School accountability*	Assessment is used to hold educators and education systems accountable for producing a quality education for students by tying student performance on assessment to consequences.	
3. *Student accountability*	Assessment is used to hold individual students accountable for their learning through their performance on assessment.	a. Assessment *certifies* student learning b. Assessment has *consequences* for students (e.g., grades)
4. *Irrelevant*	Assessment as a formal process of evaluating students and teaching is irrelevant to a teacher's responsibility to students and his/her profession	a. Assessment is *bad* b. Assessment is *useless* c. Assessment should be *ignored*.

Improvement. The conception of assessment as improvement, sometimes known as assessment for learning or formative assessment (Brown, Lake, & Matters, 2011), is tied to assessment research that demonstrates positive impacts on educational outcomes (Black & Wiliam, 1998). In order for assessment to lead to improvement, assessment must be believed to be a) *valid* and b) *descriptive* of student learning so that improvements in c) *teaching* and d) *learning* can take place. Assessment as improvement relies on a wide range of assessment techniques and strategies (Linn & Gronlund, 2000) that are carried out in the process of instruction and provide appropriate feedback to the learner and instructor (Sadler, 1989).

School accountability. *School accountability* refers to the use of assessment results to publicly demonstrate that schools and teachers are effective and imposes negative or positive

consequences for schools or teachers on account of those results (Firestone, Mayrowetz, & Fairman, 1998). The rationale is that schools and teachers have to be able to demonstrate they are delivering quality instruction (Smith & Fey, 2000) and that by holding teachers and schools accountable, instruction will improve (Linn, 2000).

Student accountability. In this purpose of assessment, students are individually accountable for their learning through their performance on assessments (e.g., graduation exams, grade promotion, college entrance exams) (McMillan, 2007). In this way, assessment then a) *certifies* student learning, providing information about how much students have learned, often against a set of criteria or standards. Accordingly, that information then has b) *consequences* in the form of grades, retention or promotion, and selection for further education or employment. Some believe that this type of assessment places a necessary and motivating pressure on students (Kahn, 2000); whereas, others believe that assessment, particularly in the form of high-stakes testing, adversely effects students, causing unwarranted worry and anxiety (Smith, 1991).

Irrelevant. An 'anti-purpose' of assessment, this conception refers to the belief that assessment as a formal, organized process of evaluating student performance has no place in teaching and learning. Teachers' intuitive, intimate, and continuing knowledge of students, content, and pedagogy precludes the need to carry out formal assessment (Torrance & Pryor, 1998). Further, negative consequences, again largely from high-stakes testing environments, make assessment (a) *bad*, reducing teacher autonomy and professionalism and distract from the real purpose of teaching—student learning (Firestone et al., 1998; Hamilton et al., 2007). Teachers may also consider assessment (b) *useless* because they find it to be inherently subjective (Hall, 2000). Finally, teachers who find assessment irrelevant also believe it should be

(c) *ignored* because assessment is just a game that students have to negotiate in order to get a fair result.

The Teachers' Conceptions of Assessment Inventory

Brown used the TCoA framework to develop and validate the *Teachers' Conceptions of Assessment Inventory* (TCoA-III)—a tool for the assessment of teachers' conceptions of assessment. The TCoA-III contains 27 items, with three items each loading onto the nine factors in the framework. Each item is a statement about assessment (e.g., "Assessment keeps schools honest.") and respondents are asked to rate their agreement with the statement using a positively packed 6-point agreement (e.g., strongly agree, usually agree, moderately agree, slightly agree, usually agree, strongly disagree), which is known to generate discrimination in contexts of social desirability (Brown, 2004). The TCoA-III was extracted from a longer instrument of 40 items, which also mapped onto the same nine factors but in the interest of creating an shorter instrument to reduce response fatigue (Brown, 2008), the strongest 3 items that loaded onto each of the nine factors was retained to create the current 27-item instrument (Brown, 2008).

The 27-item instrument was originally piloted with 525 elementary teachers (K-8) from New Zealand and validated with 784 primary teachers from Queensland, Australia. Through confirmatory factor analyses and structural equation modeling, Brown (2008) revealed an invariant factor structure of these elementary teachers' conceptions of assessment that became the anchor model against which other researchers test their results in different contexts. The model, as described in Table 2, posits four main conceptions of assessment, three of which are informed by two or more lower-order factors. Among those major conceptions, *improvement* is the dominant purpose for assessment (i.e. highest mean rating), followed by *school accountability, student accountability*, and lastly *irrelevant*. When testing the framework with

Queensland high school teachers (n=614), Brown (2008) found the same factor structure. However, *student accountability* tended to be endorsed by high school teachers somewhat more than elementary school teachers, while the use of assessment to hold schools accountable elicited very weak agreement from high school teachers.

  Inter-factor relationships. Regarding the relationship between factors, Brown used effect size differences to describe the relative importance teachers give to each of the different conceptions of assessment and inter-factor correlations to illustrate how teachers' ratings of one conception are linearly related to others. Table 3 provides a summary of these relationships as reported in Brown (2008) with effect size differences reported on the upper diagonal and correlations reported on the lower diagonal.

  Using Cohen's (1988) guidelines for interpreting effect sizes, medium to large effect sizes are observed for nearly all inter-factor effect size differences, suggesting significant and large differences in how strongly teachers agree with each conception. The exceptionally large effect sizes ($d > 1.0$) between *improvement* and *school accountability* and *improvement* and *irrelevant*, highlights teachers' strong belief in assessment for the use of improvement over and above those two purposes, while a small effect size difference ($d = .39$) between *improvement* and *school accountability* indicates similar beliefs in the use of assessment for those purposes. Likewise, the difference between school accountability and student accountability was large ($d = .86$) and the different between student accountability and irrelevant was close to large ($d = .70$). In contrast, the relationship difference between *school accountability* and *irrelevant* was only trivial ($d = .13$).

Table 3. Summary of correlational[a] relationships and effect size differences between New Zealand teachers' conceptions of assessments.

Conception	1	2	3	4
1. Improvement	--	1.29	.39	1.11
2. School accountability	.60	--	.86	.13
3. Student accountability	.30	.60	--	.70
4. Irrelevant	-.75	-.12	.30	--

Note. Statistics are from Brown (2008). Factors are listed in order from highest mean rating to lowest. Inter-factor correlations for the four factors are presented below the diagonal, and effect size differences between the four factors are presented above the diagonal.
[a] All inter-factor correlations were significant at p<.01 except for *school accountability (2)* and *irrelevant (4)*.

Looking at inter-factor correlations, when New Zealand teachers believe assessment is about *improvement*, then it is highly unlikely they will consider assessment as *irrelevant* and are more likely to believe assessment is connected to *school accountability*. If teachers think assessment is about *school accountability*, then they may or may not believe that assessment is *irrelevant*; belief in one is independent of belief in the other. Finally, teachers who believe assessment is for *student accountability* will likely consider assessment *irrelevant*, because it is bad for students or inaccurate, and thus they can safely ignore it.

Pre-service teachers' conceptions of assessment. The available research on pre-service teachers' conceptions of assessment is small but insightful. Since PSTs do not have substantial experience in using assessment to guide teaching and learning, their conceptions of assessment are more likely to reflect their assessment experiences as high school and post-secondary students rather than that of practicing teachers (Brown, 2011a). For example, a small study of U.S. PSTs noted that the formative purposes of assessment they observed in their practicum courses was not what they had experienced as high school or college students (Goc Karp & Woods, 2008). Brown (2011a) evaluated the appropriateness of the TCoA framework with 324

PSTs at a university in Auckland, New Zealand using the TCoA-III, short version. Using a combination of exploratory and confirmatory factor analyses, Brown concluded that the four-factor model originally validated for practicing teachers was inappropriate for pre-service teachers and suggested that pre-service teachers actually held five different conceptions of assessment: 1) *improvement*, 2) *measures school quality validly*, 3) *ignored and inaccurate*, 4) *bad*, 5) *grades students*. Similar to studies of New Zealand practicing teachers, the dominant conception of assessment held by PSTs in Brown's study was *improvement*. The conception of *irrelevant* was divided into two distinct conceptions: assessment as unimportant or something teachers *ignore,* and assessment as a negative experience (*bad)*. Similar to practice teachers, the more PSTs conceived of assessment to be *ignored*, the more likely they viewed assessment as used to *hold students accountable*. However, their perception of whether assessment was *bad* was independent of their conception of using assessment to hold students accountable.

Brown and colleagues research on New Zealand teachers, pre-service teachers, and high school students illustrates some trends in changing conceptions of assessment from the role of the student to the role of the teacher. Brown and Hirschfield (2008) summarized in their findings of New Zealand high school students' conceptions of assessment that as students move through high school and into teacher preparation, they begin to value the role assessment plays in improving both teaching and learning. The emphasis on *student accountability* seemingly wanes as PSTs and teachers begin to understand how assessment contributes to measure of school quality and view it as more relevant. Unlike high school students, both pre-service and practicing teachers who view assessment as *irrelevant* are less likely to conceive of assessment as holding students accountable, suggesting an acceptance of their assessment responsibilities as teacher.

Applications of the TCoA-III in different contexts. A number of studies on both

practicing and pre-service teachers in many different policy contexts have used the TCoA-III to

try and recover four main factors of the TCoA framework. Notably, the findings from these

studies suggest that recovery of the statistical invariant four-factor structure is dependent in part

on how similar the assessment and education policy contexts are to the original study of New

Zealand teachers. New Zealand is widely recognized for giving priority to low-stakes, teacher

judgments about student performance, as opposed to high-stakes, test-driven education systems

(Brown, 2008; 2011b). Not surprisingly then, previous studies with the TCoA inventory found

statistical invariance between New Zealand primary and secondary teachers (Brown, 2011b) and

between New Zealand and Queensland primary teachers (Brown, 2008), with Queensland having

a similar context for assessment. However, Brown and Harris (2009) argued that the recent shifts

in New Zealand schools for school-wide, evidence-based improvement led to significant shifts in

New Zealand teachers' conceptions of assessment; that is, they had a much stronger emphasis on

school accountability than previous teachers had reported. Similarly, cross-cultural research with

the TCoA has suggested that strong historical and cultural acceptance of the importance and

value of public examinations impacts teacher beliefs systems. For example, in Hong Kong,

(Brown et al., 2009), China (Brown et al., 2011), and Egypt (Gebril & Brown, 2013), only a

three-factor model was recovered from the TCoA-III inventory, with the *student accountability*

factor subsumed into a larger factor of *accountability* and *accountability* having the highest mean

agreement among teachers (as opposed to *improvement*). Similarly, the five-factor model for

PSTs was recovered in a comparative study of PSTs in Spain (Brown & Remesal, 2012), which

is another low-stakes assessment environment. However, the tracking strategies employed in

Spain's schools were offered as a possible explanation for why *assessment is bad* was rated significantly higher than *assessment is for improvement*.

The statistically significant differences between cross-cultural samples suggests that, while the current TCoA inventory translated into different language can work, a new set of items, factors, and structures may be needed to capture the full range of beliefs of prospective teachers, and possibly also of practicing teachers in other contexts. It has become apparent that modeling teacher conceptions with just four intentions may be a function of the cultural origins of the research and that in different societies somewhat different intentions and patterns exist.

The purpose of this study was to examine whether the four-factor model originally devised by Brown (2008) was appropriate for a sample of U.S. PSTs. However, rather than just administer the TCoA-III and check for statistically equivalent models, I employed three phases of survey development to evaluate the instrument (Presser et al., 2004) prior to administration. The process included item selection through content analysis of the TCoA-III short and long forms and a similar survey that been administered to PSTs in New Zealand (Smith et al., 2012), expert review with survey design and assessment experts, and two rounds of cognitive interviews. These pre-testing measures lead to a considerably different instrument than the TCoA-III with most notably, the suggestion that PSTs hold a conditional conception of assessment that bound all other conceptions, believing there is a way *assessment is* and a way *assessment should be*.

Research Design

Data for this study come from a comprehensive survey that contained several items and instruments designed to measure undergraduate PSTs' conceptions of assessment, and influences on those conceptions including beliefs about types of assessment useful for instruction,

conceptions of teaching and learning, self-efficacy, and self-reported influences on conceptions

of assessment in the current U.S. context (See the full instrument in Appendix C). The full

survey development process underwent three phases of pre-testing prior to administration: 1)

content review, 2) expert review, and 3) cognitive interviews. The final survey was administered

to upper-level PSTs at a large, mid-western university. Exploratory factor analyses and inter-

factor effect size differences and correlations were computed to describe the factor structure of

these PSTs' conceptions of assessment, as well as provide substantive comparisons to previously

reported conceptions of assessment in practicing and pre-service teachers. Effect size differences

between individual items and comparative models support the presence of PSTs' conditional

conception of assessment. Only the pre-testing steps as they relate to the TCoA framework are

described here.

Content Review

In order to select items for the conceptions of assessment instrument used in this study, I

conducted a content review of the the TCoA-III short form (n = 27 items), additional items from

the TCoA-III long form (n = 40 items) (Brown, 2008), and the Pre-service Teachers Beliefs

about Assessment questionnaire (n = 19 items) (Smith et al., 2012). The questionnaire

administered by Smith et al. (2012) was also an adaptation of the TCoA-III and, at the time of

this study, the authors shared the instrument but had not yet completed structural analyses (Hill,

personal communication). One of the intentional changes made in adapting these instruments for

this study was to change the structure of some items that aligned with the *improvement*

conception. For example, one item on Brown's instrument read "Assessment provides feedback

to students about their performance" and on my instrument, the item equivalent read, "Students

use feedback from assessments to improve their learning." This change was to reduce the

possible effects of response bias as a result of "yea-saying" (Smith, 2004) in this particular conception since it is consistently rated the highest among all four conceptions and may be subject to social desirability. While the focus of the item is arguably on the user and not the use of assessment itself, it indirectly captures the respondents' conception because if a PST does not believe that assessment as a process can be used to provide feedback in service of improvement, then it is also likely they will not believe students are going to use assessment for that purpose as well. Five items were changed for this purpose.

Using the TCoA framework, I compared all items on each version of the survey under each major conception of assessment (*improvement*, *school accountability*, *student accountability*, *irrelevant*). In the end, I created a 39-item instrument, selecting 20 out of 27 items from the TCOA-III short form, 2 additional items from the long form, and 11 from the PST survey by Smith et al. I also wrote 6 items of my own. In devising this initial instrument, I created a sub-factor of school accountability with four new items to reflect the current trends in U.S. education policy of using student performance data to evaluate teachers. The items were functionally equivalent to the school accountability items and, wherever possible, the word school was simply replaced by teacher (e.g., "Assessment is a good way to evaluate a school." became "Assessment is a good way to evaluate a teacher.").

Expert Review

After creating the initial instrument, the instrument along with the rest of the questionnaire underwent expert review. Four reviewers with expertise in both survey design and assessment in U.S. K-12 contexts were invited to review the survey. The reviewer form asked reviewers to comment on the strengths and weaknesses or limitations of the survey, completeness in addressing the conceptions of assessment, and other constructs in the survey.

The expert reviewers offered several suggestions that improved the quality of the instrument including feedback on content, formatting, and editorial issues. In particular, expert reviewers indicated that they felt the items aligned well with the conceptual framework but identified three items that appeared redundant with other items in the instrument. These three items initially remained in the instrument until evidence from the cognitive interviews also supported that they were in fact redundant and removed from the instrument.

Cognitive Interviews

Two rounds of cognitive interviews were conducted with third (n = 7) and fourth (n = 2) year pre-service teachers from the same population of PSTs that received the survey in February and March of 2013. Cognitive interviews function as an exploration of the respondents' understanding of the items on a questionnaire revealing if the understanding of the constructs are shared or conflicted with the conceptual framework and the intent of the researcher (Desimone & le Floch, 2004). Data from cognitive interviews can be used to increase consistencies within and across participants in interpreting the content of the instrument as well as improve the researcher's interpretation of the data collected, increasing aspects of validity within the instrument.

The cognitive interviews were conducted utilizing a hybrid model (Ryan, Gannon-Slater, & Culbertson, 2012), combining both think-aloud and verbal probing approaches. Participants were asked to provide an account of what they were thinking (e.g., think-aloud) as they responded to an item or just after responding to the item (Tourangeau, Rips, & Rasinski, 2000) and I probed their responses based on what was said using semi-structured probes (e.g., Can you tell me more about that? Why did you respond that way?). The cognitive interviews contributed to many editorial changes of the survey and suggested that the order of questions might influence

responses, leading to the decision to randomize items during administration. It was also during the cognitive interviews that the conditional conception of *the way assessment is* and *the way assessment should be* appeared.

During the first round of cognitive interviews, the interviewees expressed an inability to select one response option that appropriately reflected their judgment (Tourangeau, 2000). This was because as they read the items, they felt that their response would be different when they considered it from the perspective of *the way assessment is* and the way they thought *assessment should be*. For example, during one interview, a respondent read the item, "Assessment challenge students to do their best," and then asked me, "Do you want me to tell how it is or, like, how I think it should be?" I first instructed the respondent to just answer the item as though I was not there. He then told me that he would just leave the item blank because he did not know how to answer. I asked this pre-service teacher to tell me more about his response and he responded that "If you want me to tell you how it is, I would say 'strongly disagree' but if you want me to tell how I think it should be, I would say 'strongly agree.'" After the first two interviews, I purposely probed the interviewees for evidence of this internal struggle for several items throughout the instrument and in nearly all instances, respondents indicated a difference between the two scenarios.

During the second round of cognitive interviews, the instrument was redesigned so that participants first responded to conceptions of assessment items with explicit directions to consider how they believe assessment *is used* or *currently happens*. Towards the end of the survey, respondents were asked respond to the same set of items but this time report their conceptions of assessment as they believe it *should be used* or *should happen*. During this round of cognitive interviews, I focused specifically on whether the interviewees could respond to both

sets of items, whether their responses were different, and probed their responses as necessary. Data from the second round of cognitive interviews confirmed that this change to two sets of items addressing the contrasting conceptions was appropriate. Also, because of these conditional beliefs about assessment, a positive-packed agreement scale was no longer tenable as the PSTs expressed greater degrees of distinction in the negative end of the scale.

The final instrument was 36 items and nine factors for *the way assessment is* (WAIS) and 28 items and six factors for *the way assessment should be* (WASB), and utilized a six-point Likert-scale ranging from strongly disagree to strongly agree. Table 4 lists all items and corresponding factors for *the way assessment is*. The only difference between this instrument and *the way assessment should be* instrument is that *irrelevant* items were omitted from the WASB instrument and each item was restated to elicit beliefs about what the purposes of assessment should be. For example, the first item in the table, "Assessment helps identify the particular learning needs of any student" read "Assessment *should* help identify the particular learning needs of any student" on the WASB instrument. The instrument was administered online and items were randomly presented to participants to reduce response bias. Also, participants were not provided with a definition of assessment prior to answering the items but instead were asked to "respond to a set of statements about assessment, whatever that term means to you."

Table 4. Final factors and items for pre-service teachers' conceptions of assessment.

Factor	Sub-factor	Items
Improvement	Assessment describes process of learning	Assessment identifies how students think.
		Assessment challenges students to do their best.
		Assessment shows whether students can analyze and think critically.
		Assessment measures higher-order thinking skills.
	Students use	Students use feedback from assessments to improve their learning.
		Assessment helps students identify their strengths and weaknesses.
		Students use assessments to evaluate their own learning.
		Students use assessments to help them prepare for tests.
	Teachers use	Assessment helps teachers identify the particular learning needs of any student.
		Assessment is integrated with teaching practice.
		Teachers use assessment information to modify on-going teaching of students.
		Assessment allows different students to get different instruction that fits their needs.
School and teacher accountability	School	Assessment provides information on how well schools are doing.
		Assessment determines if schools are being effective.
		Assessment is a good way to evaluate a school.
		Assessment measures the worth or quality of schools.
		Assessment keeps schools honest.
	Teacher	Assessment is a good way to evaluate a teacher.
		Assessment measures the worth of a teacher.
		Assessment results reflect the quality of teaching.
		Assessment determines if teachers are being effective.
Student accountability	Certify student learning	Assessment determines how much students have learned.
		Assessment is comparing a student's performance to a set of goals, standards, or criteria.
		Assessment is used to determine if students have met state standards.
		Assessment is useful when reporting a student's progress and achievement to parents.
	Judge students	Assessment is assigning a grade to student work.
		Assessments are useful to determine if students pass or fail a grade.
		Assessment selects students for future education or employment.
Irrelevant	Bad	Assessment interrupts students' learning.
		Assessment is a stressful activity for students.
		Assessment is unfair to most students.
	Useless	Assessment is of little use to teachers on a day-to-day basis.
		Assessment takes away time that should be spent on teaching.
		Teachers pay attention to assessment only when the stakes are high.
		Teachers ignore assessment information even if they collect it.
		Assessment has little impact on teaching.

Note. The item listed here are WAIS items; WASB items are identical but omit the *irrelevant* factor and each item includes the word "should" in front of the action verb of the sentence. These items were randomized for administration.

Participants

All PSTs (n = 731) enrolled nine education programs offered at a large, research-intensive, mid-western university were invited to participate in an online survey in May 2013. In total, 119 PSTs completed at least 90%[2] the survey for a response rate of 16.3%. Although the response rate is unfortunately low, the response sample is representative of the population proportions of different education programs overall with a slight oversampling of secondary education students (+5%) and under-sampling of elementary education students (-8%). Further, the response rate was substantial enough to afford meaningful analyses including exploratory factor analyses (Guadagnoli & Velicer, 1988), reliability analyses for scale dimensions, and other mean comparisons. The timing of the survey administration was purposefully chosen to coincide with close to the end of spring semester, when third-year PSTs had completed 10 weeks of observation and the majority of their teaching methods courses over the course of the academic year and fourth-year PSTs were beginning to take over their classrooms in their semester long student teaching experiences. That being said, the timing of the administration may also be an attributing factor to the low response rate. Table 5 illustrates demographic characteristics of the participants including whether they had completed assessment training (n = 75, 65.3%) and were currently student teaching (n = 63, 54.8%). For the final analytic sample, respondents were categorized into one of four possible education programs: elementary/early childhood (n = 34), secondary education (n = 40), special education (n = 15), and other (e.g., agricultural education, visual arts, music, and physical education students, n = 25).

---

[2] Typically, 90% of the survey means that the respondent answered all questions except for demographics at the end; 5 of the 119 students did not answer these questions.

Table 5. Pre-service teacher participant demographics (n = 119).

Demographic	Frequency	%[a]
Female	88	75.6
Programs		
Elementary/Early Childhood (ECE)	34	29.6
Secondary Education (SE)	40	34.8
Special Education (SPED)	15	13.0
Other (e.g., music, visual arts, physical education, agricultural education)	25	21.7
Currently Student teaching	63	54.8
Receiving training in assessment [b]	75	65.2

[a] Percentages are calculated from 115 participants; 4 participants omitted responses to demographics. [b] Participants were asked if they had completed any training in assessment including courses, workshops, units in other classes, or other and then to describe their training.

Analyses

Data analyses were conducted to 1) identify pre-service teachers' conceptions of *the way assessment is* (WAIS) and conceptions of *the way assessment should be* (WASB) and 2) compare differences between these conditional conceptions. First, exploratory factor analysis (EFA) and descriptive statistics using means, correlations, and paired t-tests suggest somewhat distinct factor structures of these PSTs' conceptions of assessment. Individual item mean differences also suggest significant differences between WAIS and WASB. However, the similarities between the factor structures of each of the models supports collapsing some factors of the WAIS model so that it mirrors the WASB model and comparing shared items within those models to report significant mean differences between factors of *the way assessment is* and *the way assessment should be*.

Factor Structures

Exploratory factor analyses were conducted on 28 of the 36 items[3] in the WAIS

instrument and all 28 items of the WASB instrument. EFA was conducted using a principal

components extraction (PCA) for WAIS data while Principal Axis Factoring (PAF) was used for

the WASB data because violations of normality were detected in the WASB items (Fabrigar,

Wegener, MacCallum, & Strahan, 1999). For both models, I employed an oblique rotation

method (Promax) since the factors were conceptually related (Lattin, Douglas, & Green, 2003).

Factors were retained based on eigenvalues greater than 1 and if the variance explained by the

factor was greater than or approximately equal to 5%. EFA on the 28 WAIS items suggested a

five-factor model that retained 27 items[4]. In contrast, EFA on the WASB items indicated that a

three-factor model was present with only 23 items retained. Table 6 shows the results of the PCA

where five factors and 59.31% percent variance explained overall by five factors in the WAIS

model and 66.72% of the variance in the WASB model could be explained by three factors. The

factor *measures effectiveness* contains the same items in each model.

---

[3] Items written for the construct of assessment is *irrelevant* were omitted from the WAIS EFA because there was no
comparison set of items in WASB.

[4] Items that had poor fit characteristics were identified and dropped from subsequent analyses; this poor fit included
items with loadings below .40, those with cross-loadings greater than .40 on another factor, those with
communalities below .40, and those that had poor theoretical fit with the other items in the factor.

Table 6. Exploratory Factor Analysis Results for WAIS and WASB.

Factors	Eigenvalue	% of variance	Cumulative % of variance
Assessment the way it is (WAIS, 27 items)			
Measures effectiveness	8.75	32.40	32.40
Useful to teachers	2.67	9.89	42.29
Useful to students	1.72	6.37	48.67
Compares students	1.56	5.78	54.45
Judges student qualifications	1.31	4.90	59.31
Assessment the way it should be (WASB, 23 items)			
Used for improvement by teachers and students	10.53	47.38	47.38
Measures effectiveness	3.12	13.57	60.95
Compares and judges student qualifications	1.33	5.78	66.72

Scale scores were computed by taking the unweighted averages of items within each factor and reliability analyses were conducted to confirm the coherence of each conception. (Reliability analyses were only conducted on the *irrelevance* items). Table 7 provides scale statistics for each of the factors in the WAIS and WASB models. Each scale was fairly reliable, very close to or greater than the acceptable level of Cronbach's alpha (greater than 0.6; DeVellis, 2003), which could not be improved by any item deletion.

Table 7. Summary of final survey scales for WAIS and WASB.

Model	No. Items	Mean	SD	α
		Scale Statistics		
Assessment the way it is				
1. Compares students	4	4.44	0.78	0.59
2. Useful for teachers	7	4.02	0.92	0.84
3. Useful for students	4	3.67	0.95	0.79
4. Judges student qualifications	4	3.60	0.93	0.58
5. Irrelevant	7	3.19	0.8	0.76
6. Measures effectiveness	8	3.15	1.03	0.91
Assessment the way it should be				
1. Useful for improvement	10	5.22	0.82	0.93
2. Compares and judges student qualifications	5	3.91	0.95	0.80
3. Measures effectiveness	8	3.50	1.08	0.93

Overall, the factors in each of the models contain many of the same items from the TCoA framework but the names are written to reflect the nuanced character of items within. For example, factor 6 in WAIS and factor 3 in WASB is called *measures effectiveness* because the items within reflect the potential for assessment to provide information about quality of schools and teachers. These factors are the same for both models and similar to New Zealand PSTs conception that *assessment measures school quality validly* (Brown, 2011b). Unlike previous studies of PSTs using this framework, PSTs in this sample do not distinguish between assessment as bad and to be ignored or inaccurate. In the WAIS model, the highest mean agreement is *comparing students* whereas in the WASB model, *useful for improvement* is rated highest.

Inter-factor Relationships

Tables 8 and 9 highlight the effect size differences (upper diagonal) and the linear relationships (lower diagonal) between factors within models. Overall, significant, mostly positive correlations of moderate size are observed between conceptions of assessment within

WAIS and WASB as well as large effect size differences between factors within models. For example, the more likely PSTs believe that assessment *measures effectiveness* or *should measure effectiveness*, then the more likely PSTs are to endorse any other conception of assessment, except *irrelevant*. Notably, the lack of relationship between *irrelevant* and *measures effectiveness* ($r = -0.11, p = .22$) suggests that, for better or worse, these PSTs accept that assessment is used to evaluate schools and teachers. The relationships between factors in the WAIS and WASB models are similar to the linear relationships observed in Brown (2008) except that where there was a positive relationship between *irrelevant* and *student accountability* in New Zealand teachers, there are two negative relationships between *irrelevant* and *judges student qualifications* and *irrelevant* and *compares students*. This suggests that PSTs who agree assessment is used to make determinations about student learning and qualifications are less likely to find assessment *irrelevant*, meaning that they accept the use of assessment for these purposes.

Although the effect size differences in the WAIS model range from nearly nonexistent between *useful for teachers* and *useful for students (d = .06)* to exceptionally large between *compares students* and *irrelevant (d = 1.57)*, nine of the fifteen effect size differences reported here are medium or large, suggesting relatively significant and large differences in how strongly teachers agree with each conception. Similarly, especially large effect sizes ($d > 1$) are reported in Table 9 for the difference between *useful for improvement* and the other two factors. However, the smaller effect size differences, such as that between *irrelevant* and *measures effectiveness* signal similar strength of agreement.

Table 8. Inter-factor relationships for WAIS.

The Way Assessment Is						
Factor	1	2	3	4	5	6
1. Compares students	--	.47	.86	.94	1.39	1.57
2. Useful for teachers	.39***	--	.38	.45	.99	.90
3. Useful for students	.32***	.63***	--	.06	.53	.57
4. Judges student qualifications	.34***	.17	.26**	--	.47	.50
5. Measures effectiveness	.46***	.48***	.53***	.38***	--	.03
6. Irrelevant	-.21*	-.52***	.18*	-.37***	-.11	--

Note. Inter-factor correlations are on the lower diagonal and effect size differences are on the upper diagonal. *p<.05. **p<.01. ***p<.001

Table 9. Inter-factor relationships for WASB.

The Way Assessment Should Be			
Factor	1	2	3
1. Useful for improvement	--	1.40	1.79
2. Compares and judges students	.50***	--	.40
3. Measures effectiveness	.47***	.69***	--

Note. Inter-factor correlations are on the lower diagonal and effect size differences are on the upper diagonal. ***p<.001

Comparing WAIS and WASB

Data from the cognitive interviews suggested that teachers might hold different conceptions of *the way assessment is* and *the way assessment should be*. Statistical analyses provide supporting evidence of this conditional conception of assessment as well. Individual mean differences between common items (n=23) in the WAIS and WASB models show that 68% of items on the survey (n = 19) were significantly different and, of those mean differences, over 90% of them yield effect size differences that are either moderate ($0.5 < d < 0.8$, n = 7), large ($d > 0.8$, n = 3) or exceptionally large ($d > 1$, n = 8). As Table 10 shows, most significant mean differences are found in items reflecting use of assessment for improvement by students and teachers, suggesting that PSTs perceive a vast difference between the current practice of

assessment for the purpose of improvement and how they perceive it should be. Because of the number of statistical tests performed for this analysis, Type-I error inflation becomes an issue of concern. P-values are reported in this table to allow readers to make their own determination of the strength of statistical evidence.

Table 10. Individual items differences between shared items of WAIS and WASB (n = 115).

Item	WASB		WAIS		Test statistics		
	Mean	SD	Mean	SD	t	p	d
Assessment challenges students to do their best.	5.33	0.97	3.77	1.25	10.46	.000	1.37
Assessment measures higher-order thinking skills.	5.00	1.19	3.37	1.25	9.92	.000	1.39
Students use feedback from assessments to improve their learning.	5.33	0.92	3.53	1.12	13.24	.000	1.72
Assessment helps students identify their strengths and weaknesses.	5.15	1.02	4.02	1.26	7.46	.000	0.99
Students use assessments to evaluate their own learning.	5.17	1.01	3.65	1.27	9.91	.000	1.32
Students use assessments to help them prepare for tests.	4.79	1.19	3.81	1.27	5.98	.000	0.83
Assessment helps teachers identify the particular learning needs of any student.	5.39	0.94	4.47	1.2	6.43	.000	0.84
Assessment is integrated with teaching practice.	5.35	0.98	4.52	1.23	5.54	.000	0.74
Teachers use assessment information to modify on-going teaching of students.	5.38	0.99	4.65	1.06	5.36	.000	0.71
Assessment allows different students to get different instruction that fits their needs.	5.37	0.91	3.79	1.49	9.68	.000	1.24
Assessment provides information on how well schools are doing.	3.97	1.33	3.56	1.33	2.33	.021	0.33
Assessment determines if schools are being effective.	3.66	1.35	3.49	1.31	1.00	.320	0.14
Assessment is a good way to evaluate a school.	3.55	1.41	3.03	1.28	2.90	.004	0.42
Assessment measures the worth or quality of schools.	3.13	1.22	3.02	1.23	0.66	.512	0.09
Assessment is a good way to evaluate a teacher.	2.98	1.27	3.14	1.32	-0.89	.376	-0.12
Assessment measures the worth of a teacher.	3.37	1.29	2.45	1.21	5.49	.000	0.78
Assessment results reflect the quality of teaching.	3.86	1.35	3.16	1.34	3.82	.000	0.56
Assessment determines if teachers are being effective.	3.54	1.32	3.48	1.38	0.34	.735	0.05
Assessment determines how much students have learned.	4.54	1.17	3.86	1.28	4.14	.000	0.57
Assessment is comparing a student's performance to a set of goals, standards, or criteria.	4.21	1.35	4.52	1.22	-1.79	.074	-0.26
Assessment is used to determine if students have met state standards.	3.91	1.21	4.74	1.08	-5.36	.000	-0.76
Assessment is assigning a grade to student work.	3.21	1.36	3.38	1.49	-0.91	.366	-0.13
Assessments are useful to determine if students pass or fail a grade.	3.74	1.33	3.98	1.41	-1.32	.189	-0.19

Note. Items shown here are shared items between the WAIS and WASB instruments. Items are identical but in the WASB instrument, the word "should" is in front of the action verb of the sentence (e.g., Assessment should challenge students to do their best).

Admittedly, item-level analyses do not provide reliable evidence of the complex

constructs being measured in this study. However, results from the factor analyses and inter-

factor relationship comparisons indicate that the underlying factor structures and relationships

between factors are similar enough between factors to support reducing the WAIS model from

five factors to three by collapsing *useful for teachers* and *useful for students* into one factor—

useful for improvement and combining *compares students* and *judges student qualifications* into

one factor—*compares and judges students*. Mean differences between common items within

each conception are shown in Table 11. As might be expected, the effect size difference between

WAIS-*useful for improvement* and WASB-*useful for improvement* is exceptionally large. The

small but significant effect size difference between *measures of effectiveness* suggests that PSTs

not only accept that assessment is used to measure the quality of schools and teachers but also

believe that it should be used for that purpose as well, although not as strongly as using

assessment for improvement.

Table 11. Differences between factors of WAIS and WASB.

Conception of Assessment	*Way it should be*		*Way it is*			
	Mean	Standard Deviation	Mean	Standard Deviation	*t-test*	*Effect Size*
Useful for improvement	5.22	.82	3.94	.82	12.95***	1.54
Compares and judges students	3.92	.95	4.09	.95	1.94	-.20
Measures of effectiveness	3.50	1.07	3.15	1.07	3.97***	.33

***p<.001

Discussion

The purpose of this study was to determine if Brown's four-factor model explaining

teachers' conceptions of assessment was appropriate for a sample of U.S. pre-service teachers.

Only a qualitative comparison is permissible given the differences in the instruments used to

measure conceptions of assessment in previous literature and the one administered in this study.

The results suggest that PSTs hold substantively similar conceptions of assessment as other

populations that have completed this instrument. There is evidence that PSTs do conceive of assessment in terms of improving teaching and learning (*use for improvement*), as well as holding school and teachers (*measures effectiveness*) and students (*judges student qualification, compares students to expectations*) accountable for student performance. Inter-factor relationships between factors within both WAIS and WASB are similar to previous research as well. Furthermore, the three-factor model that emerged from the WASB items is similar to other three-factor models in Chinese, Hong Kong, and Egyptian contexts, except that here *use for improvement* has the highest rate of agreement rather than *judges and compares students* (i.e., *student accountability*). The major difference, then, is that these U.S. PSTs perceive conditional differences in assessment. That is, these PSTs clearly believe that there is a way assessment is currently used and practiced and away that it should be. Why? As suggested in previous studies this conditional difference in conceptions of assessment could be influenced by the context in which it operates.

U.S. Context

A typical PST today, who went straight from high school to college, would be a direct recipient of one of the most significant education policies in United States history, the No Child Left Behind Act (No Child Left Behind [NCLB], 2001). The elementary and secondary school experiences of these PSTs were directly impacted by NCLB. Extensive research highlights how NCLB was accompanied by a continued narrowing of the curriculum, increased testing, proliferation of test preparation, and test anxiety as consequence of increased emphasis on accountability (e.g., Darling-Hammond, 2004; Hamilton et al., 2007; Stecher & Chun, 2001). As students, these same PSTs would also have been the recipients of significant paradigm shifts in curriculum, pedagogy, and assessment occurring in the last three decades, particularly the

emphasis on using assessment to improve teaching and learning (Shepard, 2000; Black & Wiliam, 2005; National Research Council, 2001).

As PSTs recount their assessment experiences and engage in their teacher education programs, they may also reflect on the tensions present in assessment policies and practices today as indicative of a *way assessment is* and a *way assessment should be*. These PSTs would likely begin forming conceptions of assessment prior to their teacher education programs and reflect on their experiences and understanding of those experiences of assessment as students. As reflected in the results, these PSTs readily accept that assessment is used to compare students to state standards, to assign grades or help student prepare for tests, and by teachers to modify instruction. Yet, their instructors and teacher education curriculum may augment the present tensions in assessment by privileging theories of curriculum and pedagogy that emphasize particular roles assessment should play in the teaching and learning process and/or supply a particularly critical lens through which accountability uses of assessment, especially in the evaluation of schools or teachers, are considered. This is reflected in the large effect size difference between *useful for improvement* in WAIS and WASB suggesting that PSTs believe assessment is capable of doing more for students and teachers than they themselves have experienced or observed in the schools. Similarly, the proliferation of testing as a result of test-based accountability systems (Linn, 2000) may explain the mild agreement that assessment is used to compare students to criteria and attach consequences like promotion or retention since that resonates with PSTs elementary and secondary school experiences. However, learning about multiple forms of assessment in their teacher education program or even recounting negative experiences with standardized testing may explain the non-significant difference between *compares and judges students* in WAIS and WASB. Similarly, the small effect size difference

between *measures of effectiveness* in WAIS and WASB suggest that while PSTs only slightly agree that assessment is currently being used for measuring the quality of teachers and schools, they do not necessarily think that it should be used any more for this purpose either.

Implications

Recognizing previous efforts to create an instrument that is both efficient and readily applicable in various contexts, the first step is to administer the same instrument to a similar population of PSTs and examine whether the models of WAIS and WASB can be recovered. This conditional belief may be key to understanding how pre-service teachers, who are both preparing to become educators but at the same time are still students, understand the purposes for assessment and as well as other beliefs measured in this study including teaching and learning, self-efficacy, and the kinds of data they believe are most useful to teachers. Another way to examine the issue is to look at how larger or smaller differences between these conditional conceptions relate to their beliefs about teaching, learning, self-efficacy or data. For example, do pre-service teachers' who perceive large differences between the way assessment is used for improvement and the way they think it should be have a higher sense of self-efficacy because they view assessment as a potentially powerful tool to help meet students needs?

Still other questions of interest include whether those conditional conceptions of assessment exist at the beginning of their undergraduate education or are invoked by teacher education curricula. As one example, previous research on PSTs' conceptions of assessment suggests that conceptions begin to change over the duration of an assessment course (Brown, 2011; Deneen & Brown, 2011; Smith et al., 2012). However, it is unknown whether those conceptions strengthen over time or wane considerably at the overwhelming prospect of being a first-year teacher and all that comes with that experience. Studying first and second year

undergraduates who have not had extensive training in curriculum, pedagogy or assessment might suggest entirely different conceptions of assessment and the conditional aspect might not be present at all. Similarly, a longitudinal study of pre-service teachers could track their conceptions of assessment upon entering the program, at key points throughout (e.g., after observations, before and after student teaching) and follows them into their first few years of teaching. This research would permit a closer examination of what exactly raises the notion that the way it is, is not the way it should be, and whether that belief persists or wanes throughout their teacher education program and early career. It may also be worth exploring whether PSTs beliefs about assessment are subject to representational biases (Tversky & Kahneman, 1974) that strongly, and perhaps erroneously, inform their judgment. For example, if teachers' conceptions of assessment are likely to be learned as students in school (Pajares, 1992), it might be the case that PSTs are largely drawing on just a few especially traumatic or rewarding experiences with assessment to make those judgments.

Finally, if pre-service teachers' conceptions of assessment do influence their assessment literacies, then it is necessary to examine the relationship between this conditional conception and their competencies. The evidence that pre-service teachers believe there is a *way assessment is* and *a way it should be* could both facilitate or hinder the acquisition and application of assessment knowledge and skills. On one hand, students who see the potential for assessment to bring about improvement may eagerly engage in learning about how to use assessment to improve teaching and learning. On the other hand, students who fixate on the negative consequences of assessment may believe that just because assessment "should be" used in particular ways for improvement does not mean that it will ever be given current trends and thus fail to engage in learning best practices. Should these kinds of relationships exist, the

implications for teacher education training in assessment are to be inclusive of pre-service teachers' critical reflection on their conceptions of assessment and to tailor assessment training to directly address both their conceptions and related influences that might be especially negative or counter-productive to learning sound assessment practices.

Conclusion

Although inclusive of a small sample of PSTs and reflective of only one education program in the U.S., this study lays the groundwork for future research in understanding PSTs' conceptions of assessment. At the outset, this study reminds readers of the importance of carefully pre-testing a previously validated instrument when it is going to be administered to a new population in a new context. The use of cognitive interviews revealed that these U.S. pre-service teachers struggle internally with a conditional conception of assessment, believing that the way it is currently practiced is not necessarily the way it should be. Examining whether respondents' interpretations of self-reported items are consistent with intended meanings is fundamental for judging whether instrument results provide valid interpretations (Ryan et al., 2012). Without cognitive interviews, there would have been no way of knowing this conditional conception existed and given the expressed difficulty in choosing a response option during the interviews, data quality would have been a greater issue in this study.

As PSTs complete teacher training and enter the work force, they will draw on both their training and their personal experiences with assessment to navigate and reconcile tension against assessment policy and practices. They need to be aware of the controversy surrounding the use of external, standardized assessments to hold schools, teachers, and students accountable, inform public school choice, qualify for federal funding, and, perhaps most controversially, evaluate teachers (Mandinach, Rivers, Light, Heinze, & Honey, 2006; Recesso & Zepeda, 2009).

Assessment literacy may be at risk if teachers fear assessment and testing, have false perceptions about assessment, or succumb to strong pressures to practice assessment in service of meeting accountability demands over and above improvement. Helping new teachers wrestle with these internal struggles, especially if they perceive a way assessment is and a way it should be, is critical to developing sound assessment practices early on in their careers.

REFERENCES

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74.

Black P. & Wiliam, D. (2005). Lessons from around the world: how policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal*, *16*(2), 249–261

Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, *94*, 1015–1024.

Brown, G. T. L. (2008). Conceptions of assessment: Understanding what assessment means to teachers and students. New York: Nova Science Publishers.

Brown, G. T. L. (2011a). New Zealand prospective teacher conceptions of assessment and academic performance: Neither student nor practicing teacher. In R. Kahn, J. C. McDermott, & A. Akimjak (Eds.) *Democratic access to education* (pp.119–132). Los Angeles, CA: Antioch University Los Angeles, Department of Education.

Brown, G. T. L. (2011b). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*, 45–70.

Brown, G. T. L., & Harris, L. R. (2009). Unintended consequences of using tests to improve learning: How improvement-oriented resources heighten conceptions of assessment as school accountability. *Journal of MultiDisciplinary Evaluation*, *6*(12), 68–91.

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice, 15*(1), 3–17.

Brown, G. T. L., Hui, S. K. F., Yu, W. M., & Kennedy, K. J. (2011). Teachers' conceptions of assessment in Chinese contexts: A tripartite model of accountability, improvement, and irrelevance. *International Journal of Educational Research*, 50, 307–320.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice,16*(3), 347–363.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*, 210–220.

Brown, G. T. L., & Michaelides, M. P. (2011). Ecological rationality in teachers' conceptions of assessment across samples from Cyprus and New Zealand. *European Journal of Psychology of Education, 26*, 319–337.

Brown, G. T. L., & Remesal, A. (2012). Prospective teachers' conceptions of assessment: A cross-cultural comparison. The Spanish Journal of Psychology, 15, 75–89.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Darling-Hammond, L. (2004). Standards, Accountability, and School Reform. *The Teachers College Record*, *106*(6), 1047–1085.

Delandshere, G. & Jones, J. H. (1999) Elementary teachers' beliefs about assessment in mathematics: A case of assessment paralysis. *Journal of Curriculum and Supervision, 14*(3), 216–240.

DeLuca, C., Chavez, T., & Cao, C. (2012). Establishing a foundation for valid teacher judgments: The role of pre-service assessment education. *Assessment in Education: Principles, Policy and Practice, Special Issue: Moderation Practice and Teacher Judgment,* 1-20.

Deneen, C. C., & Brown, G. T. L. (2011, October). *The persistence of vision: an analysis of continuity and change in conceptions of assessment within a teacher education program*. Paper presented to the 37th annual meeting of the International Association for Educational Assessment, Manila, Philippines.

Desimone, L. M., & le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Education Evaluation and Policy Analysis, 26*, 1–22.

DeVellis, R. F. (2003). Scale development: Theory and applications (2nd ed.. Thousand Oaks, CA: Sage.

Fabrigar, L R., Wegener, D. T., MacCallum, R. C., Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods 4*(3): 272–299.

Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: the effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis, 20*(2), 95–113.

Gebril, A. & Brown, G.T.L (2013). The effect of high-stakes examination systems on teacher beliefs: Egyptian teachers' conceptions of assessment. *Assessment in Education: Principles, Policy & Practice*. *doi:* 10.1080/0969594X.2013.831030

Goc Karp, G., & Woods, M. L. (2008). Preservice teachers' perceptions about assessment and its implementation. *Journal of Teaching in Physical Education, 27*(3), 327–346.

Guadagnoli, E. & Velicer, W. F. (1988) Relation of sample size to stability of component patterns. *Psychological Bulletin, 103 (2),* 265–275.

Hamilton, L.S., Stecher, B.M., Marsh, J.A., Sloan McCombs, J., Robyn, A., Russell, J., et al., (2007). *Standards based accountability under No Child Left Behind.* Santa Monica, CA: Rand Corporation.

Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *Journal of Educational Research, 93,* 276–286.

Lattin, J. M., Douglas, C. J., & Green, P. E. (2003). Analyzing multivariate data. Pacific Grove, CA: Duxbury.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29(2),* 4-16.

Linn, R. L. & Gronlund, N. E. (2000). *Measurement and Assessment in Teaching* (8th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Mandinach, E., Rivas, L., Light, D., & Heinze, C. (2006, April). *The impact of data-driven decision making tools on educational practice: A systems analysis of six school districts.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

McMillan, J. H. (2007). *Classroom assessment: Principles and practice for effective instruction* (4th ed.). Boston, MA: Allyn & Bacon.

McMillan, J.H. 2001. Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice* 20, no. 1: 20–32.

Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *PROFILE*, 14, 143–158.

National Research Council. 2001. *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment. Pelligrino, J., Chudowsky, N., and Glaser, R., editors. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107–110, § 115, Stat. 1425 (2002).

Pajares, M.F. 1992. Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research* 62, no. 3: 307–332.

Philippou, G. & Christou, C. (1997). Cypriot and Greek primary teachers' conceptions about mathematical assessment. *Educational Research and Evaluation, 3*(2), 140–159.

Recesso, A. M. & Zepeda, S. J. (2009) Evidential reasoning and decision support in assessment of teacher practice (pp. 363-381). In T. J. Kowalski & T. J. Lasley II (Eds.) *Handbook of Data-based Decision Making in Education*. New York, NY: Routledge.

Ryan, K., Gannon-Slater, N., & Culbertson, M. (2012) Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation, 33*(3), 414–430.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18,* 119–144.

Saltzgaver, D. (1983). One teacher's dominant conceptions of student assessment. *Curriculum Perspectives, 3*(2), 15–21.

Shepard, L.A. (2000). *The role of classroom assessment in teaching and learning. Educational Researcher, 29*(4), 4–14.

Smith, M. L. (1991). Put to the test: The effects of external testing on students. *Educational Researcher,20*(5), 8–12.

Smith, M. L., & Fey, P. (2000). Validity and accountability in high-stakes testing. *Journal of Teacher Education, 51*(5), 334–344.

Smith, L. F., Cowie, B., Gilmore, A. & Hill, M. (2012, August). Student and teacher educator perceptions of becoming assessment-capable teachers. Paper presented at EARLI Assessment SIG Conference, Brussels, Belgium.

Smith, T. (2004). Developing and evaluating cross-national survey instruments. In Stanley Presser, Jenninfer M. Rothgeb, Mick P. Couper, Judith L. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer (Eds) *Methods for Testing and Evaluating Survey Questionnaires*. New York: Wiley.

Stecher, B. & Chun, T. (2001). School and classroom practices during two years ofeducational reform in Washington state, (CSE Tech. Rep. No. 550). Los Angeles:University of California, National Center for Research on Evaluation, Standards, and Student Testing.

Torrance, H., & Pryor, J. (1998). *Investigating formative assessment: Teaching, learning and assessment in the classroom*. Buckingham: Open University Press.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). The psychology of survey response. Cambridge, England: Cambridge University Press.

Tverksy, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.

CONCLUSION

The paths from data to practice decisions are complex, mediated by characteristics of the data (Coburn & Talbert, 2006); teachers' knowledge, experiences and dispositions toward data, assessment, learning, content areas and their practice (Gannon-Slater, paper 2); organizational culture and capacity (Coburn & Turner, 2011); as well as the fiscal and political contexts in which teachers work (Honig, 2012). This project supports emerging shifts in DDDM research toward better capturing and understanding the complexity of teachers' data use. The first paper clarifies two different types of instructional decision making that teachers practice in their profession, only one of which is typically considered in DDDM research. Because the improvement of instructional decision making is a goal of data use (Coburn & Turner, 2011), this paper provides a wider lens through which instructional decisions might be considered and seeks to reframe the role data might play in both determination of means and ends. The second paper, the field study of high school science teachers, provides a multifaceted portrait of instructional decision making 'on the ground', offering additional evidence that teachers' do not find standardized assessment data useful for instruction and that they draw on multiple sources of evidence for decision making. This study also suggests future research should be inclusive of a broader definition of data and attend to the fact that the data teachers find useful for instructional decision making are varied with respect to myriad classroom contexts and instructional goals. Finally, with further validation of the instrument and examination of the relationships between pre-service teachers' conceptions of assessment and other relevant philosophies or beliefs, the third paper provides a potentially useful cognitive framework through which we can understand pre-service teachers' development of competencies in assessment, including their understandings of how data are useful for instruction. Collectively, these papers support reframing conceptions

of assessment and data in DDDM research to facilitate better understandings of how teachers use data for instructional decision making. Additionally, these papers can be connected in at least two key ways to provide further implications for research.

Understanding Teachers' Conceptions of Assessment

Coburn and Turner's (2011) conceptual framework on data use references studies on teachers attention to and interpretation of data, explaining that new information is always understood through the lens of what they already know and believe, which, in turn, influences how data is encoded, organized, and interpreted (Greeno, Collins, & Resnick, 1996; Spillane & Miele, 2007). Research on DDDM suggests that local and national political contexts (e.g., accountability pressures, teacher evaluation systems, and support for innovative practices) that surround data use initiatives are incredibly influential in how teachers use data in their instructional practices (Coburn & Talbert, 2006; Ingram et al., 2004; Jennings, 2012; Young, 2006). However, the nature of those relationships are not well understood. Teachers' beliefs about the purposes of assessment could serve as a cognitive framework through which local or national policy messages about assessment and data use are filtered in making decisions about assessment and data use. Evidence of conditional conceptions of assessment might help explain reported deficiencies in teachers data use practices (e.g., Hoover & Abrams, 2013; Ingram et al., 2004; Schneider & Gowan, 2013). For example, we might imagine that teachers who believe that assessment s*hould* be used for improvement but do not perceive that to be the case in their school or district assessment policies, might overlook potentially useful data for their instruction. Or those same teachers might persist in practicing assessment in ways that reconcile with their beliefs about the *way assessment should be* but appear to be resisting the adoption of assessment and data use initiatives in their organizations that they perceive to contradict those beliefs.

Furthermore, there is research that shows individuals attend more closely to test scores or assessment information that confirms pre-existing beliefs and discount data that challenges these beliefs (Coburn, 2001; Coburn et al., 2009; Young & Kim, 2010). For example, teachers' preferences for more informal kinds of data in instructional decision making may be supported by prior experiences where they observed that such data supported changes in instruction and those changes were judged to be successful. Thus, informal assessments better serve the goals of assessment for improvement. Alternatively, the idea that formal assessment data are mostly for students' use in regulating their own learning might be explained, in part, by a belief that formal assessment is used primarily as a means to hold students accountable. This might explain why teachers do not associate the data from these formal assessments as measures of the effectiveness of their instruction (Ingram et al., 2004). For example, a teacher may administer a mid-unit quiz to provide information to students about their learning so far and expect students to use that information to improve performance later on without considering whether the students' performance on that quiz signals a potential problem with their instruction.

Continuing to Reframe Instructional Decision Making

While the distinction between instructional decision making as instrumental reasoning and professional judgment is important, the broader issue is that further clarification is needed of what we mean by instructional decision making in order to better understand the roles that data play in such deliberations. As one example, early research on teachers' pedagogical thinking indicated that teachers engage in five different levels or "scales" of planning: yearly, term, unit, weekly, and daily (Clark & Yinger, 1987; Shavelson & Stern, 1981). My research suggests teachers draw on different sources of data in their daily or weekly instruction (e.g., primarily oral or behavioral data) than they do in planning next units or for next year (e.g., labs, performance

assessments, or unit tests). As another example, assessment scholars often position different types of assessment for improving student learning as they occur in different phases of instruction: "on-the-fly", "planned-for-interaction" and formal and embedded in the curriculum (Shavelson, 2003). "On-the-fly" assessments tend to occur spontaneously when "teachable moments take place in the classroom" where as "planned-for-interaction" reflect deliberate intentions to discern and improve student learning and may be formal or informal assessments. Formal, curriculum-embedded assessments are administered at key points in a learning sequence intended to create "teachable moments" (Heritage, 2007; Ruiz-Primo, 2011; Shavelson, 2003).

These different frameworks suggest that instructional decision making can be differentiated with respect to the goals for instruction at a given time (e.g., continue moving forward in a given lesson versus need to reteach the last three concepts) and the length of time over which decisions are made (e.g., decisions about what to do today versus next week or next semester). Currently, research on DDDM does not systematically differentiate teachers' instructional decision making in these or other possible ways and thus does not effectively consider how different data may matter in different kinds of instructional decision making (Shepard, 2005). By purposefully clarifying and systematically examining different types of instructional decisions in DDDM research, we can better understand the kinds of supports teachers need to use various kinds of data in those circumstances.

The idea that educators should make decisions informed by empirical information is perfectly logical and not anything new (Mandinach, 2012). What is new, however, is that the use of data is now inextricably connected with accountability for the improvement of student achievement (Mandinach & Honey, 2008). The basic model of accountability in the United States today targets the *school* as the organization for monitoring, intervention and change

(O'Day, 2002). But to the extent that needed change involves the behaviors of members within that organization, change must occur ultimately at the individual level – with teachers. This project reflects a modest attempt to clarify a possible cognitive lens through which pre-service and in-service teachers might understand the value of data for instruction and, articulate the kinds of instructional decisions we expect to be impacted by the use of data. The use of data to inform instruction is a practice that holds much promise. As the DDDM literature continues to grow, at least three key steps to realizing that promise are that we broaden our notions of assessment and data in the study of teachers' data use, clarify our understandings of teachers' instructional decisions and, closely examine how the current assessment climate shapes pre-service and in-service teachers' conceptions of assessment and data.

REFERENCES

Clark, C. M., & Yinger, R. J. (1977). Research on Teacher Thinking. *Curriculum inquiry*, *7*(4), 279–304.

Coburn, C. E. (2001). Collective sense-making about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23*(2), 145–170.

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., & Talbert, J. E. (2006). Conceptions of Evidence Use in School Districts : Mapping the Terrain. *American Journal of Education*, *112*(4), 469–495.

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

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.) *Handbook of Educational Psychology* (pp. 15–46). New York, NY: Macmillan.

Heritage, M. (2007). Formative Assessment : What Do Teachers Need to Know and Do? *Phi Delta Kappan*, *89*(2).

Henig, J. R. (2012). The politics of data use. *Teachers College Record*, *114*(11), 1-17.

Hoover, N. R. & Abrams, L. M. (2013) Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26(3),* 219-231.

Ingram, D., Louis, K. S., & Schroeder, R. G. (2004). Accountability policies and teacher decision-making: Barriers to the use of data to improve practice. *Teachers College Record, 106*(6), 1258-1287.

Jennings, J. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record Volume 114*(11), 1–23

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. & Honey, M. (2008), *Data-driven school improvement: Linking data and learning*. New York: Teachers College Press

O'Day, J. (2002). Complexity, Accountability, and School Improvement. *Harvard Educational Review*, *72*(3), 293-330.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The Role of instructional dialogues in assessing Students for science learning. Special Issue in Assessment for Learning. *Studies of Educational Evaluation, 37*(1), 15–24.

Schneider, M. C. & Gowan, P. (2013). Investigating teachers' skills in interpreting evidence of student learning. *Applied Measurement in Education, 26(3),* 191-204.

Shavelson, R. (2003). *On the integration of formative assessment in teaching and learning with implications for teacher education*. Stanford, CA, and Manoa, HI: Stanford Education Assessment Laboratory and University of Hawaii Curriculum Research and Development Group.

Shavelson, R. & Stern, P. (1981). Research on teachers' pedagogical thoughts, judgments, decisions, and behavior. *Review of educational research*, *51*(4), 455-498.

Shepard, L. A. (2005, October). *Formative Assessment: Caveat Emptor*. Paper presented at the ETS Invitational Conference 2005, New York.

Spillane, J. P. & Miele, D. B. (2007). Evidence in practice: A framing of the terrain (pp. 46–73). In P. A. Moss (Ed.), *Evidence and decision making*. 106th Yearbook of the National Society for the Study of Education. Malden, MA: Blackwell.

Young, V. M. (2006). Teachers' Use of Data: Loose Coupling, Agenda Setting, and Team Norms. *American Journal of Education, 112(4),* 521–548.

Young, V. M., & Kim, D. H. (2010). Using assessments for instructional improvement: a literature review. *Education Policy Analysis Archives, 18* (19). Retrieved from http://epaa.asu.edu/ojs/article/view/809.

APPENDIX A

OBSERVATION PROTOCOL

Teacher _____   Date _____

Period _____   Grade/Level _____

 Class _____

INFORMATION FROM TEACHER
1. Topic of lesson:
2. Purpose/Objectives of the lesson:
3. Intended outcomes:
4. Placement of class or lesson within the unit of study:
5. **Collect any handouts/worksheets or so on.**

Identify the major instructional format of class (if more than 1, be sure to note how long for each): (lecture, discussion, review, lab, seat work, group work, individual work etc – see Science Classroom Observation Protocol):

Notes about the classroom (number of students, materials being used, how they are seated at start of class, how they talk to each other and the teachers)

Activity/Tasks: Content, nature of activity, what students doing, what teachers' doing, interactions, duration (note number of activities/tasks cued by teachers it happens with ROMAN NUMERALS; what cues)

Evidence Decision-making: what are they assessing; how are they assessing; what cues do they seem to attend to? Do they change the course of instruction – describe specifically. What cues? What do they remark about student learning to the students?

INTERVIEW PROTOCOL

Do you have any questions for me before we begin? I'm happy to answer any questions you have about the study at the end, too. Ok, we will go ahead and start.

1) Tell me about what you are currently working on in your courses. What unit are your studdents in? How far along are they? And how is it going?

Daily Instructional decision-making

1. So, you are in these units and my guess is that you need to make some decisions along the way on what to teach and how to teach it – some of those decisions may show up in class while others occur in the course of planning. I want to start by talking to you about the instructional decisions you make WHILE You are teaching.
   a. CLASS: Does the way you teach in a 50 minutes class change or depend on the *CLASS* teaching? Probe: maturity, class size, student ability, student learning style.
   b. CONTENT: How is what you are teaching in 50 minutes related to the *CONTENT* – play into your instructional decision-making on a daily basis? PROBE: How does it matter, in that 50 minutes, if you are teaching harder or easier material? More procedural – problem solving things - or conceptual knowledge? More critical thinking, deep reasoning skills versus facts, rote memorization?
   c. PEDAGOGY: How does it matter, in that 50 minutes, if you are teaching in a particular way? In my observations, I witnessed a multitude of teaching strategies: direct instruction, labs, 'homework days', more discussion oriented days and a combination of many of those things. What do you have to keep in mind during class when you are presenting materials in a particular way?
   d. CHANGE: Can you give me an example of a time when your decision of what to teach or how to teach something changed from the beginning of the day to the end of the day? Or even *in the middle of class?*
   e. EVALUATION : How do you know if it's going well?
      i. What kinds of information do you collect or things do you look for when you want to know if a lesson in a particular day is going well?
      ii. What do you find most helpful or useful in making that judgment?
      iii. *How do you know if it's not going well?*
   f. ANYTHING ELSE *Is there anything else about your daily, in class instructional decision-*making that comes to mind that I didn't ask about? Affects your decisions on a day-to-day basis?

Member check

PLANNING

2. So, continuing with this idea that teaching is a series of decisions of what to teach and how to teach, I want to talk more about your instructional decision-making that occurs over time as you are planning lessons or units or looking ahead even further. The questions I ask are similar to before but I really want to focus again on the instructional decision-making over time versus what's going in those 50 minutes. Feel free to give me as many examples that pop into your head as I ask.
   a. How far out in advance to you typically plan?

b. How is planning influenced by: class size, maturity, student ability, student learning styles, content, preferred pedagogy

c. EVALUATION Similar to before - *how do you know if it's going well? How do you know if students are getting it?*

   i. What do you find most helpful or useful in making that judgment?

   ii. How do you know if it's not going well? Based on that information, how do you decide what do next? Can you give me an example?

   iii. How do you know if it's not going well?

d. ANYTHING ELSE: *Is there anything else about your daily, in class instructional decision*-making that comes to mind that I didn't ask about?

Member Check

INFLUENCES

a. Data collection strategies: what kinds of informal and formal data influence your instruction?

   a. What other data do you collect or attend to?

   b. What happens when data conflict?

b. Student Accountability: So one thing I've heard through out my study is this particular emphasis on holding students accountable for their learning – that "teaching is a 2-way street" and "if they don't' do their job, I can't do mine." Some of what I heard is a reality – you simply cannot grade 125 homeworks every night but there is a particular philosophy in play. I'm curious about your take on this. how do you design your instruction to hold them accountable?

   a. What do you think a student's responsibility is in your classroom?: What do you expect from students?

   b. What are your goals in holding student accountable? What do you hope to see come of it – anything else besides learning the material?

   c. So, how does student accountability play a role in your instructional decision-making? How do you design your instruction to hold them accountable?

   d. Following along those lines – how much influence do you as a teacher think you have on student learning?

c. MOTIVATION: Along those same lines, I watch and listen to a lot of teachers talk about the importance of motivating students. Do you think student motivation plays a role in your instructional decision-making? Can you give me an example?

d. GOALS: So is it fair to say, that when thinking about your teaching, you aren't just thinking about the content but you have other goals in mind? Are there other goals you keep in mind as you plan instruction? Eg: independent learners, learning skills, etc. taking the AP test etc.

e. Professional Development; Your schooling/training?

f. Colleagues

MEMBER CHECK

Support for Instructional Decision-making

1) Transition: Thanks a lot for all this information. Just one more topic and I'm done. I want to talk a little bit about the kinds of support you might get with your instruction.

   a. To start off, I'm interested in these early start days that happen twice a month. The PLC's. What do you guys do during that time? Who are you in there with?

       i.   My understanding is that the primary focus these days are the common assessments. Let's talk about that – what are you talking about in reference to the common assessments?

      ii.   What other kinds of issues or topics do you talk about?

     iii.   What kind of information do teachers share about their students?

     iv.   Do teachers talk about their decisions of what to teach and how to teach?

      v.   What is your general sense of this collaboration of the PLC– is it helpful? Useful? In what ways is it not helpful or useful?

  b. Do you guys ever observe each other teach What's that like?

  c. Are there any others ways that you collaborate as a department or across content areas?

3. What kinds of support – training, professional development or otherwise do you get to help you with your instruction? Is it useful?

Member Check

4. In the course of this interview, have any other thoughts about your instructional decision-making come to mind that I didn't ask you?

  a. What else have you taught in the past?

  b. Confirm years of experience, level of education.

  c. I want to end with the question: What do you like best about teaching chemistry? What makes it hard to do?

END: Well, that's all the questions I have for you today. Do you have any questions for me? Thank you! This has been tremendously helpful. As I analyze all this information you gave me, would you be OK if I followed up with you by email or phone if I had any questions?

APPENDIX C

PRE-SERVICE TEACHERS' CONCEPTIONS OF ASSESSMENT, TEACHING AND LEARNING SURVEY

INTRODUCTION

The purpose of this questionnaire is to gather information about your personal opinions on assessment, teaching, and learning. **There are no right or wrong answers** and your responses will be kept fully anonymous. The entire survey should take you about 10 minutes. Thank you so much for your participation!

What Does Assessment Mean to You?
The first part of the survey asks about your beliefs and opinions about *assessment*, whatever that term means to you. Assessment is a hot topic in education today and we want you to draw on your own knowledge and experience with assessment to answer the following statements. Remember, there are no right or wrong answers!

A1: Think about the word **assessment**. What **words** would you use to describe assessment? List as many as you can think of.

A2: Think about the word **assessment** again. This time, what **feelings or emotions** come to mind? List as many as you can think of.

SECTION I: 'THIS IS THE WAY I THINK ASSESSMENT IS'
☺ **PLEASE READ THESE DIRECTIONS CAREFULLY BEFORE YOU GO ON.** ☺
The next part of the survey will ask you to respond to a set of statements about assessment, whatever that term means to you. As you read these statements, please indicate the extent to which you agree or disagree based on your thoughts and beliefs about how assessment **happens or is used today. Try your hardest not to respond** based on how you think assessment *should happen or be used* and focus **only** on what your opinions and beliefs about how *assessment is currently used today.*

Based on how I think assessment <u>is</u> used today…	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
1. Students use feedback from assessments to improve their learning.	1	2	3	4	5	6
2. Assessment helps teachers identify the particular learning needs of any student.	1	2	3	4	5	6
3. Students use assessments to evaluate their own learning.	1	2	3	4	5	6
4. Assessment shows whether students can analyze and think critically.	1	2	3	4	5	6
5. Students use assessments to help them prepare for tests.	1	2	3	4	5	6
6. Assessment is integrated with teaching practice.	1	2	3	4	5	6
7. Teachers use assessment information to modify on-going teaching of students.	1	2	3	4	5	6
8. Assessment measures higher-order thinking skills.	1	2	3	4	5	6
9. Assessment allows different students to get different instruction that fits their needs.	1	2	3	4	5	6
10. Assessment identifies how students think.	1	2	3	4	5	6
11. Assessment challenges students to do their best.	1	2	3	4	5	6
12. Assessment helps students identify their strengths and weaknesses.	1	2	3	4	5	6
13. Assessment provides information on how well schools are doing.	1	2	3	4	5	6
14. Assessment determines if schools are being effective.	1	2	3	4	5	6
15. Assessment is assigning a grade to student work.	1	2	3	4	5	6
16. Assessments are useful to determine if students pass or fail a grade.	1	2	3	4	5	6

Based on how I think assessment is used today…	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
17. Assessment is a good way to evaluate a school.	1	2	3	4	5	6
18. Assessment selects students for future education or employment.	1	2	3	4	5	6
19. Assessment measures the worth or quality of schools.	1	2	3	4	5	6
20. Assessment keeps schools honest.	1	2	3	4	5	6
21. Assessment is a good way to evaluate a teacher	1	2	3	4	5	6
22. Assessment measures the worth of a teacher.	1	2	3	4	5	6
23. Assessment results reflect the quality of teaching.	1	2	3	4	5	6
24. Assessment determines if teachers are being effective.	1	2	3	4	5	6
25. Assessment determines how much students have learned.	1	2	3	4	5	6
26. Assessment is comparing a student's performance to a set of goals, standards, or criteria.	1	2	3	4	5	6
27. Assessment is used to determine if students have met state standards.	1	2	3	4	5	6
28. Assessment is useful when reporting a student's progress and achievement to parents.	1	2	3	4	5	6
29. Assessment is of little use to teachers on a day-to-day basis.	1	2	3	4	5	6
30. Assessment interrupts students' learning.	1	2	3	4	5	6
31. Assessment is a stressful activity for students.	1	2	3	4	5	6
32. Assessment takes away time that should be spent on teaching.	1	2	3	4	5	6
33. Assessment is unfair to most students.	1	2	3	4	5	6
34. Assessment has little impact on teaching.	1	2	3	4	5	6
35. Teachers pay attention to assessment only when the stakes are high.	1	2	3	4	5	6
36. Teachers ignore assessment information even if they collect it.	1	2	3	4	5	6

Section I.A. Possible Influences on Your Answers from Before
Directions: Think about your responses to the previous statements about the way assessment is used. Below is a list of possible factors that you might have been thinking about when you answered those statements. Thinking about your answers **overall**, rate how influential each factor was in determining your answers to those questions.

Possible Influences on how you think assessment is currently used today	Not at all Influential	Slightly Influential	Moderately Influential	Very Influential	Extremely Influential
37. Your own experiences with assessment	0	1	2	3	4
38. Family members' experiences with assessment	0	1	2	3	4
39. Friends' experiences with assessment	0	1	2	3	4
40. Common Core Standards	0	1	2	3	4
41. Standardized testing	0	1	2	3	4
42. Assessments other than standardized testing	0	1	2	3	4
43. Education policies (e.g., NCLB,)	0	1	2	3	4
44. Your teacher education curriculum	0	1	2	3	4
45. Observations in current or past placements	0	1	2	3	4
46. What you hear from other P-12 teachers	0	1	2	3	4
47. Student teaching (leave this question blank if you are not student teaching)	0	1	2	3	4
48. Other (please specify)_____[48_text]	0	1	2	3	4

Section III: Assessment Types
Directions: In this section there are list of different assessments a teacher might give to his/her students to collect information about their student learning. Please rate **how useful each assessment is to teachers in providing information about student learning.**

How useful are these in providing information about student learning?	Not at all Useful	Slightly Useful	Moderately Useful	Very Useful	Extremely Useful
49. Fill-in-the-blank items with a word bank	1	2	3	4	5
50. Fill-in-the-blank items without a word bank	1	2	3	4	5
51. True/false items	1	2	3	4	5
52. Multiple-choice items	1	2	3	4	5
53. Short answer or essay items	1	2	3	4	5
54. In-class assignments (e.g., practice problems, reading responses)	1	2	3	4	5
55. Homework assignments (e.g., practice problems, reading responses)	1	2	3	4	5
56. Individual projects	1	2	3	4	5
57. Group projects	1	2	3	4	5
58. Portfolios	1	2	3	4	5
59. State standardized tests (e.g., ISAT, PSAE)	1	2	3	4	5
60. National standardized tests (e.g., ACT, ITBS)	1	2	3	4	5
61. Student presentations/performances	1	2	3	4	5
62. Observing students in class	1	2	3	4	5
63. Oral questioning	1	2	3	4	5
64. Quick checks (e.g., bell ringers, exit slips)	1	2	3	4	5

A3: Can you think of any other types of assessments that teachers might give to his/her students to collect information about their learning that are missing from this list? Please list them here.

Section II: Assessment knowledge
A4 In your education program, have you received any training in assessment? YES OR NO (PIPING)

A4_Y If yes, please describe where you received such training (courses, workshops etc) and what you learned about (topics, skills, issues etc).

A5. **Directions:** The following are a list of topics common in assessment training. Even if you have not had any formal assessment training (e.g., classes, workshops), please rate the extent to which you feel you have learned about each topic.

Please indicate how much you have learned about…	Nothing	Some but not very much	A great deal
a. Writing your own assessments	0	1	2
b. Developing your own rubrics or grading keys	0	1	2
c. Administering assessments	0	1	2
d. Test preparation strategies	0	1	2
e. Analyzing assessment data that is not from standardized tests	0	1	2
f. Analyzing and interpreting standardized test data	0	1	2
g. Using assessment data to inform instruction	0	1	2
h. Communicating assessment information to different audiences (e.g., parents, administrators, students, teachers)	0	1	2

Section IV: Your Personal Beliefs and Opinions about Teaching and Learning.
Directions: In this section there are several statements about teaching and learning. Please respond to these questions using your **own personal opinions and beliefs** about teaching and learning. Be honest and frank in determining the extent to which you agree or disagree with each statement. Remember, there are no right or wrong answers!

		Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
65.	Since teachers know a lot more than students, they shouldn't let students muddle around when they can just explain answers directly.	1	2	3	4	5	6
66.	A quiet classroom is generally needed for effective learning.	1	2	3	4	5	6
67.	Students are not ready for meaningful learning until they have acquired the basic skills.	1	2	3	4	5	6
68.	It is better when the teacher-not the students-decides what activities are to be done.	1	2	3	4	5	6
69.	Working individually on textbook problems is a good way for students to practice what they learned in class.	1	2	3	4	5	6
70.	Students will take more initiative to learn when they feel free to move around the room during class.	1	2	3	4	5	6
71.	Instruction should be built around problems with clear, correct answers.	1	2	3	4	5	6
72.	Instruction should be built around ideas that most students can grasp quickly.	1	2	3	4	5	6
73.	How much students learn depends on how much background information they have - that is why teaching facts is so necessary.	1	2	3	4	5	6
74.	It is important for students to share their work with others, because other students can learn from what they do.	1	2	3	4	5	6

	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
75. Students should help establish the criteria that will be used to assess their work.	1	2	3	4	5	6

Directions for numbers 76-79: For each of the following pairs of statements, check the box that best shows **how closely your own beliefs** are to each of the statements in a given pair. The closer your beliefs to a particular statement, the closer the box you check. *Please* check *only one for each set.*

76. The teacher's main role is a facilitator that provides opportunities and resources for students to discover and construct concepts for themselves.	□	□	□	□	□	□	□	The teacher's job is to explain to students how to do the work and assign specific practice. Students do not really learn a subject unless you go over the material in a structured way.
77. The most important part of instruction is content of curriculum as that reflects some consensus on what children need to know and be able to do.	□	□	□	□	□	□	□	The most important part of instruction is that it encourages "sense-making" or critical thinking among students. Content is secondary.
78. It is useful for students to become familiar with many different ideas and skills even if their understanding is limited. Later, in college, perhaps, they will learn these things in more detail.	□	□	□	□	□	□	□	It is better for students to master a few complex ideas and skills well, and to learn what deep understanding is all about, even if the breadth of their knowledge is limited until they are older.
79. It is critical for students to become interested in doing academic work - interest and effort are more important than working on a particular subject-matter.	□	□	□	□	□	□	□	While student motivation is certainly useful, it should not drive what students study. It is more important that students learn the history, science, math and language skills in their textbooks.

Directions for 80-89: Here are several statements about teaching and how you might see yourself as a teacher. Even though you are not a teacher yet, please be honest and frank in determining the extent to which you agree or disagree with each statement.

	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
80. The amount a student can learn is primarily related to family background.	1	2	3	4	5	6
81. If students aren't disciplined at home, they aren't likely to accept any discipline.	1	2	3	4	5	6
82. When I really try, I can get through to the most difficult students.	1	2	3	4	5	6
83. A teacher is very limited in what s/he can achieve because a student's home environment is a large influence on a student's achievement.	1	2	3	4	5	6
84. If parents would do more for their children, teachers could do more.	1	2	3	4	5	6
85. If a student did not remember information I gave in a previous lesson, I would know how to increase his/her retention in the next lesson.	1	2	3	4	5	6
86. If a student in my class becomes disruptive and noisy, I feel assured that I know some techniques to redirect him/her quickly.	1	2	3	4	5	6
87. If one of my students couldn't do a class assignment, I would be able to accurately assess whether the assignment was at the correct level of difficulty.	1	2	3	4	5	6

	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
88. If I really try hard, I can get through to even the most difficult or unmotivated students.	1	2	3	4	5	6
89. When it comes right down to it, a teacher really can't do much because most of a student's motivation and performance depends on his or her home environment.	1	2	3	4	5	6

SECTION V: "THIS IS THE WAY I THINK ASSESSMENT SHOULD BE"
STOP!
☺ **PLEASE READ THESE DIRECTIONS CAREFULLY BEFORE YOU GO ON.** ☺

The next part of the survey will ask you to respond to the same set of statements about assessment from the beginning of the survey. Unlike the first time, please respond to these statements based how you think assessment **should happen or should be used.**

In this section, **please indicate the extent to which which you agree or disagree with each statement about assessment based on how you think assessment _SHOULD BE USED._**

Based on how I think assessment should be….	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
90. Students should use feedback from assessments to improve their learning.	1	2	3	4	5	6
91. Assessment should help teachers identify the particular learning needs of any student.	1	2	3	4	5	6
92. Students should use assessments to evaluate their own learning.	1	2	3	4	5	6
93. Assessment should show whether students can analyze and think critically.	1	2	3	4	5	6
94. Students should use assessments to help them prepare for tests.	1	2	3	4	5	6
95. Assessment should be integrated with teaching practice.	1	2	3	4	5	6
96. Teachers should use assessment information to modify on-going teaching of students.	1	2	3	4	5	6
97. Assessment should measure higher-order thinking skills.	1	2	3	4	5	6
98. Assessment should allow different students to get different instruction that fits their needs.	1	2	3	4	5	6
99. Assessment should identify how students think.	1	2	3	4	5	6
100. Assessment should challenge students to do their best.	1	2	3	4	5	6
101. Assessment should help students identify their strengths and weaknesses.	1	2	3	4	5	6
102. Assessment should provide information on how well schools are doing.	1	2	3	4	5	6
103. Assessment should be used to determine if schools are being effective..	1	2	3	4	5	6
104. Assessment should be assigning a grade to student work.	1	2	3	4	5	6
105. Assessments should be useful to determine if students pass or fail a grade.	1	2	3	4	5	6

	Based on how I think assessment should be….	Strongly Disagree	Moderately Disagree	Mildly Disagree	Mildly Agree	Moderately Agree	Strongly Agree
106.	Assessment should be a good way to evaluate a school.	1	2	3	4	5	6
107.	Assessment should select students for future education or employment.	1	2	3	4	5	6
108.	Assessment should measure the worth or quality of schools.	1	2	3	4	5	6
109.	Assessment should keep schools honest.	1	2	3	4	5	6
110.	Assessment should measure the worth of a teacher.	1	2	3	4	5	6
111.	Assessment is a good way to evaluate a teacher.	1	2	3	4	5	6
112.	Assessment results should reflect the quality of teaching.	1	2	3	4	5	6
113.	Assessment should determine if teachers are being effective.	1	2	3	4	5	6
114.	Assessment should determine how much students have learned.	1	2	3	4	5	6
115.	Assessment should compare a student's performance to a set of goals, standards, or criteria.	1	2	3	4	5	6
116.	Assessment should be used to determine if students have met state standards.	1	2	3	4	5	6
117.	Assessment should be useful when reporting a student's progress and achievement to parents.	1	2	3	4	5	6

Section I.A. Possible Influences on Your Answers from Before
Directions: Think about how you responded to the previous statements about how assessment should be used. Below is a list of possible factors that you might have been thinking about when you answered those statements. Thinking about your answers **overall**, rate how influential each factor was in determining your answers to those questions.

Possible Influences on how you think assessment is currently used today	Not at all Influential	Slightly Influential	Moderately Influential	Very Influential	Extremely Influential
118. Your own experiences with assessment	0	1	2	3	4
119. Family members' experiences with assessment	0	1	2	3	4
120. Friends' experiences with assessment	0	1	2	3	4
121. Common Core Standards	0	1	2	3	4
122. Standardized testing	0	1	2	3	4
123. Assessments other than standardized testing	0	1	2	3	4
124. Education policies (e.g., NCLB, RTTT)	0	1	2	3	4
125. Your teacher education curriculum	0	1	2	3	4
126. Observations in current or past placements	0	1	2	3	4
127. What you hear from other P-12 teachers	0	1	2	3	4
128. Student teaching (leave this question blank if you are not student teaching)	0	1	2	3	4
129. Other (please specify) [129_text]_____	0	1	2	3	4

B1. What education program are you currently enrolled in? *(please check next to appropriate program)*
_____Agricultural Education (B1_1_AG)
_____Early Childhood Education (
_____Elementary Education
_____Foreign Language Education
_____Music Education
_____Physical Education
_____Secondary Education - English
_____Secondary Education – Math
_____Secondary Education - Science
_____Secondary Education – Social Studies
_____Visual Arts Education
_____Other *(please specify)* _____

B2. Are you pursuing a secondary endorsement?

Yes   No

B2_Y: If yes, in what area(s)? TEXTBOX.

B3. Are you currently student teaching?

□Yes   □No

[HIDE IF THEY ARE STUDENT TEACHING]

 B3_ST_N. If you are not currently student teaching, have you completed classroom observations as part of your teacher education program?

□Yes   □No

B3_Obs_Y If yes, how many semesters have you been observing in schools?

1   2   3   4   5   6   More than 6

B4. Do you have any prior experience in a classroom outside your teacher education program?

Yes   No
B4_Y: If yes, please describe: TEXTBOX.

B5. What year are you in at the University of Illinois?
_____First
_____Second
_____Third
_____Fourth
_____Fifth
_____Sixth
_____Other *(please specify)*_____

B6. What year did you graduate from high school?
_____2008
_____2009
_____2010
_____2011
_____2012
_____Other *(please specify)*_____

B7. What type of high school did you attend? (If you attended more than one type of high school, please select the type of high school that you attended the longest)
_____Public
_____Private

_____Charter
_____Magnet
_____Homeschooled
_____Other *(please specify)* _____

B8. How old are you?
_____18
_____19
_____20
_____21
_____22
_____Other *(please specify)* _____

B9. How do you identify?
_____Female
_____Male
_____Neither male or female
_____I prefer not to answer

At end of questionnaire:

Thank you so much for participating in this questionnaire! Your responses will be very helpful in understanding pre-service teachers' perceptions of assessment, teaching, and learning. Please click "Submit" below. Thank you again!