

How are we approaching data-informed practice? Development of the *Survey of Data Use and Professional Learning*

Jo Beth Jimerson¹

Received: 25 September 2014 / Accepted: 19 May 2015 / Published online: 2 June 2015
© Springer Science+Business Media New York 2015

Abstract As in international schooling contexts, talk about data-driven practice has become ubiquitous in schooling dialogues in the USA, and with the pending reauthorization of the No Child Left Behind Act (the main driver of increased data use in American schools), educators in the USA should expect even greater calls for formalized data use. Yet, the field lacks readily accessible tools that allow school district leaders and evaluators to examine educator perceptions related to data-informed practice. This paper outlines the process used in the development, testing, and validation of one instrument that district leaders and evaluators may use to learn more about the ways in which data are used and perceived in their respective contexts. Potential applications as well as limitations of the instrument are outlined.

Keywords Educational data use · Data-driven decision making · Educational data use · Continuous improvement · School improvement · Survey construction · Professional learning for data use

In communities across the USA and much of the world, policymakers, parents, business leaders, and educators are focused on the challenge of improving schools. Whether trying to “turn around” a school deemed “failing” or simply exploring ways to make an already-successful school better, these actors work to determine what is and is not working to effect improvement. One strategy employed in these endeavors is data-informed decision making (DIDM, also referred to as data-driven decision making, or DDDM) (Mandinach and Jackson 2012; Marsh et al. 2006; Militello et al. 2013).

Manuscript submitted to *Educational Assessment, Evaluation and Accountability*.

✉ Jo Beth Jimerson
j.jimerson@tcu.edu

¹ Texas Christian University, TCU Box 297900, Fort Worth, TX 76129, USA

Indeed, school and system leaders frequently claim to be “data-driven” or “evidence-based” in their decisions (Coburn et al. 2009).

Taking action in accord with evidence is easier said than done, however. While data-driven may be popular jargon at present, educational data use can be a nuanced and complex proposition (Hamilton et al. 2009; Means et al. 2011). Teachers and school leaders must work to fit evidence collected by way of standardized tests and classroom or campus-based assessments with their existing knowledge and lived experience (Brown and Rogers 2014; Saunders 2014), and they sometimes must do so in contexts that are highly politicized (Coburn et al. 2009; Daly 2009). Brown (2014) points out that evidence use is neither a simple nor linear process; similarly, educational data use cannot be reduced to a technocratic solution aimed at discovering the “one best way” to structure teaching and learning (see Cho and Wayman 2014, for more on the fallacy of expecting data use to unfold in highly predictable ways). These factors complicate how we think about and engage in data use in schooling contexts.

1 Conceptualizing “data-informed practice”

Drawing from work that emphasizes the importance of inquiry (Jimerson and McGhee 2013; Schildkamp and Teddlie 2008), the use of multiple and broad measures (Earl and Katz 2006; Mandinach and Jackson 2012; Wayman 2013), and the necessity of evidence being actionable at the classroom and/or system level (Mandinach and Jackson 2012; Schildkamp and Kuiper 2010), data-informed practice may be envisioned as an ongoing, inquiry-based process that incorporates multiple pieces of evidence (e.g., classroom assessments, standardized test data, interest inventories, and parent information, among others) to an end of identifying obstacles to student and/or organizational success and subsequently implementing strategies for better serving the academic, social, and emotional needs of individual students and groups of students.

Data-informed practice, then, must be about more than analyzing test scores and identifying “bubble students” or assigning children to educational triage in preparation for the next exam (e.g., Booher-Jennings 2005; Daly 2009; Marsh et al. 2006). It requires conditions wherein educators have the time and resources needed to engage in collaboration around multiple data to identify the needs of the whole student and to adjust instruction accordingly (Katz and Dack 2014; Kerr et al. 2006; Schildkamp and Kuiper 2010; Schildkamp and Teddlie 2008; Verhaeghe et al. 2010). And, it requires that those involved be committed to an ongoing give-and-take as they engage in “the process of acquiring, using, critiquing, and creating professional knowledge-in-action” (Saunders 2014, p. 14) through a lens informed by both a broader evidence base (that is, the established research on an issue) and the evidence base created by the “lived experience” of observing and assessing students in particular contexts on a regular basis.

2 Gauging the intersection of data and practice

Research on the elements that underpin educational data use has proliferated in the past decade, so the field (researchers, evaluators, and school district leaders) has a well-

grounded understanding that, for example, time, collaboration, and accessible data systems facilitate the ability of teachers to use data to inform practice (see Cho and Wayman 2014; Hamilton et al. 2009; Ikemoto and Marsh 2007; Louis et al. 2010; Mandinach 2012; Park and Datnow 2009; Schildkamp and Kuiper 2010; Supovitz 2012). However, just knowing where an organization wants to *be* in terms of practice and attitudes is only part of the story; to effectively plan, an organization must also take measure of where the members of an organization *are*. Point C (the end of the journey) cannot be efficiently reached without taking stock of point A (the starting point) or point B (the current position).

To this end, some researchers have developed and implemented broad-scale surveys to aid in collecting evidence about DIDM practices. Some of these are via well-funded efforts such as those used in the US Department of Education-identified priority areas and produce useful information for how we as a field understand practice across contexts (see Means et al. 2009, 2011). Others are developed in conjunction with large-scale efforts related to achievement outcomes as well as practices (the excellent work done in association with The Wallace Foundation in Louis et al. (2010), for example).

However, most of these instruments place emphasis on what educators *do* with data—i.e., who uses data and when, to what ends data are used, and how frequently educators engage in these uses. Some (e.g., Means et al. 2009) press on how educators make use of data systems. Few (Wayman et al. 2009 being an exception) attempt to provide a method for collecting information related to the more subjective issues of teacher attitudes related to educational data use.

Because of the dearth of data use-related surveys available to the field, when practitioner-evaluators or district leaders want to know more about these issues, they must either cobble together items from existing instruments (which can be time-consuming in terms of search, particularly for resource-strapped entities) or create their own instruments that serve their more immediate purposes. This becomes problematic, because as Desimone and LeFloch (2004) have noted, good survey instruments in the world of education are lacking. The process involved in developing and refining questions, checking for item stability and clarity, and piloting an instrument can be time-consuming and costly. Still, as educational leaders and policymakers press forward with efforts to infuse practice with constructive data-informed decision making, it is critical that we develop and share instruments that can help us learn more about these issues in credible ways.

By outlining the process by which one such instrument—the *Survey of Data Use & Professional Learning (S-DUPL)*—was developed, I hope to support district leaders as they engage in evaluation and subsequent improvement of approaches to data-informed practice. This paper aims to accomplish this in two ways. First, by making the development of the instrument transparent from initial item construction through piloting, I outline the process in such a way that those largely unfamiliar with survey construction and improvement may gain insight into the process so that they are better situated to develop instruments appropriate to their localized needs. Second, I provide information related to issues of validity and stability for the items and scales included in the instrument so that evaluators or school leaders who want to learn more about the ways in which data are used and perceived in a localized context may have a ready bank of items from which to draw.

3 Background and approach

The impetus for the creation of the *S-DUPL* arose from dual desires to learn more about teacher perceptions of needs specific to data use-related skills and knowledge and to learn whether teachers considered those needs well-met by the professional learning structures in place in their respective districts. An accompanying hope was that the instrument could be used by evaluators and school leaders not only as a means to collect such data but also as a “learning artifact.” That is, the deployment and analysis of the instrument could serve to seed dialogue around the intended purposes and existing supports for educational data use in a given context. Although prior research informed some specifics in terms of the concepts to be included (e.g., Hamilton et al. 2009; Ikemoto and Marsh 2007; Wayman and Jimerson 2014), no existing instrument provided the focus necessary to measure teacher perceptions in these areas. Therefore, the creation of an instrument tailored for these purposes seemed a timely and worthwhile endeavor.

Unfortunately, the educational landscape is fraught with poorly conceived survey instruments, as Desimone and LeFloch (2004) point out. Without proper consideration for established guidelines specific to survey construction, the conclusions of even the most well-intended evaluations may be inaccurate due to the deployment of an unreliable instrument (Desimone and LeFloch 2004; Johnson and Christensen 2013). In such cases, efforts to make evidence-based decisions may unwittingly be derailed by hastily designed or poorly constructed surveys. Many of the issues researchers attempt to explore via survey instruments are often nuanced and complex; without well-designed, thoughtful instruments for data collection, we risk drawing inaccurate conclusions because, in effect, what we *think* we are asking may not align with study participants’ *interpretation* of the survey questions.

Desimone and LeFloch (2004) recognized this and issued a call to action for educational researchers to be more diligent in constructing useful, reliable survey instruments that provide the kinds of information sought. Beyond attending to the basic tenets of item construction (Creswell 2003; Groves et al. 2009), Desimone and LeFloch (2004) suggested that educational researchers could improve upon the instruments used in data collection by making use of cognitive interviewing processes. Cognitive interviewing allows the researcher to better understand the thought process of a participant as the individual moves through a survey instrument in a “think-aloud” mode. By incorporating such a process into the construction of surveys used for educational research, researchers could design stronger data collection instruments, and the field (researchers, practitioners, and policymakers alike) could benefit from more accurate and useful information.

4 Process

The *Survey of Data Use & Professional Learning* was developed and revised through five iterative stages: (1) grounding of initial item construction in research related to educator data use, (2) analysis and revision of items through expert review and cognitive interviews, (3) small-scale pilot of instrument, (4) pilot and subsequent revision of instrument in a small district setting, and (5) application of test-retest

protocol to larger sample population to recheck item and scale reliabilities. In the sections that follow, I outline the evolution of the instrument through these phases.

4.1 Stage 1: grounding of the instrument and initial item construction

Initial construction of the instrument began with a review of the literature around educational data use. This review revealed that while educational data use has been an increasing expectation for decades now—catalyzed in particular in the USA by the No Child Left Behind Act of 2001—teachers and school leaders still find constructive data use challenging for a variety of reasons (Hamilton et al. 2009; Kerr et al. 2006; Schildkamp and Kuiper 2010). To start, many teachers enter the field with a less-than-solid grasp of how to appropriately assess students and how to use data in ways that inform daily practice (Mandinach and Gummer 2013). Data literacy is frequently under-addressed by teacher preparation programs (Mandinach and Gummer 2013; Means et al. 2011). Once teachers complete preparation and enter the field as professionals, they may or may not be supported effectively in learning how to use data to inform practice (Anderson et al. 2010; Jimerson and Wayman 2015; Means et al. 2009).

Of particular importance to the development of the *S-DUPL*, research suggests that school leaders must be diligent in attending to several factors that influence whether and how teachers engage in data use. Four factors appeared most prominent across studies: (1) data use-related skills and knowledge, (2) trust, (3) vision and common language, and (4) time. As these factors were examined, draft items were constructed (and revised) to press on each general theme; thus, the review of research and drafting of items occurred in iterative fashion.

Data use-related skills and knowledge Of no great surprise, teacher abilities to engage in data-informed practice depend on capacity in terms of data use-related skills and knowledge (Means et al. 2011; National Forum on Education Statistics [NFES] 2012). That is, teachers must be able to access the multiple types of data—often from multiple data systems—that they need to inform questions (Cho and Wayman 2014; Means et al. 2011; Wayman and Stringfield 2006). Teachers must be able to bring appropriate data to bear on the questions being asked—i.e., to be adept at inquiry (Earl and Katz 2006; Jimerson and McGhee 2013; Katz and Dack 2014; National Forum on Education Statistics [NFES] 2012). They must be able to accurately interpret the data they collect (Mandinach and Jackson 2012; Means et al. 2011). Perhaps most importantly, they must be able to imbue data with meaning—to interpret in such a way that it becomes actionable (Hamilton et al. 2009; Mandinach and Gummer 2013; Schildkamp and Kuiper 2010).

Clearly, effective data use depends on actual capacity in a number of interrelated skills; further, Means et al. (2011) asserted that teacher confidence in these critical skills affects engagement in data-informed practice. In response to this research, several items were drafted to measure confidence in data use-related skills (see Table 3¹). To further press on this issue—and to get at how well teachers perceived current district- or campus-based professional learning supported development in these areas—items

¹ Because the tables noted in this section also include statistics related to pilot testing, these may all be found in the later section entitled, “Stages 3, 4, and 5: Piloting, Revision, and Statistical Analyses of Items”

pertaining to professional learning supports for data use were also drafted (see Table 4). In time, these items developed into the “Confidence” and “Effectiveness of Data-Related Professional Learning” scales, respectively.

Trust Constructive data use does not simply develop in the absence of a healthy, inquiry-oriented culture. Rather, constructive data use is frequently characterized as an extension of a professional culture wherein educators feel a sense of internal accountability to support excellent teaching among one another (Copland 2003; Elmore 2004; Louis 2007). A sense of trust and interdependence among leaders and colleagues is critical if a goal is to encourage informed risk-taking rather than reinforce the perceived safety of instructional rigidity (Copland 2003; Daly 2009; Louis et al. 2010; Marsh et al. 2010; Wayman and Stringfield 2006). This is in part because, in such cultures, educators are “willing to constructively challenge each other to provide evidence for claims made during an inquiry process” (Ikemoto and Marsh 2007, p. 124) and can do so without risking employment or reputation. In contrast, a lack of trust among leaders and teachers in a school—or a context in which data are used to shame or punish teachers in an attempt to catalyze assumed capacity—impedes the sharing of data and strategies so critical to improving practice (Ingram et al. 2004; Louis 2007; Mandinach and Jackson 2012).

In short, effective data use rises or falls not just on *whether* teachers and leaders engage in data use, but on *how* they do so. To this end, items were drafted to assess educator perceptions of the culture in which data use is happening in a given context. The goal was to ascertain the comfort level educators reported when working with their colleagues around data or data-informed problems. Also, because “data use” and “accountability” can also be inextricably linked concepts for many educators (see Booher-Jennings 2005; Daly 2009; Valli and Buese 2007), items were also drafted to assess whether educators reported anxieties around data use related to misuse or abuse of data. These items, over time, coalesced into the “anxiety” and “collaboration” scales, respectively).

Vision and common language Work beyond the field of education has long asserted that the success of a team is dependent on building not only an esprit de corps among team members but also of ensuring that these team members are working toward a common goal in coherent fashion (Senge 2006). Similarly, without a shared vision or common language, attempts at data use may be haphazard and characterized by low buy-in (Mandinach and Jackson 2012; Wayman and Cho 2009). Efforts to hone and clarify what is considered data use, the intended purposes and potential benefits of using data, and how data use fits with continuous improvement practices can contribute to a positive and constructive culture for inquiry (Park and Datnow 2009; Wayman et al. 2012a, 2012b, 2012c; Wayman and Stringfield 2006). Ikemoto and Marsh (2007) noted that leaders who promote “norms of openness and collaboration” greatly enable data use, making idea-sharing across hierarchies possible (p. 124). Wayman and colleagues have written about the importance of building a shared language around data use from the “ground up”—that is, involving teachers from the outset in defining terms and honing understandings to catalyze increased clarity around expectations and processes (Wayman et al. 2007; Wayman and Stringfield 2006; Wayman et al. 2012b).

With this work in mind, items were developed to assess agreement among educators in language and in purpose. First, items were developed that aimed at parsing whether educators themselves used terms such as “data,” “evidence,” or “information” interchangeably; research often distinguishes these terms, but the language of practice can be more fluid. Also, data may carry accountability-laden baggage that evidence or information does not (see Jimerson and McGhee 2013); to help educators take stock of the language being used to describe informed inquiry processes, and what these terms mean to their colleagues, the items in Table 5 were drafted. Additional items were created to assess the degree to which educators perceived data use as ultimately benefiting students or teachers (see Table 6). These items developed into the “Construal of Data” and “Beneficence of Data” scales, respectively.

Time A final factor that influences whether and how teachers engage in educational data use is time. That is, the busyness of the school day and the management of all that teachers already have on their plates (see Olsen and Sexton 2009) can preclude thoughtful data use if leaders do not schedule blocks of time for dedicated, reflective, collaborative data use (Ikemoto and Marsh 2007; Wayman and Cho 2009; Wayman et al. 2012c). This interaction with colleagues around data is critical to sense-making and to educators’ ability to fit new evidence with tacit or professional knowledge, yet the lack of adequate time for reflection and collaboration emerges as a common theme in educational data use (Hamilton et al. 2009; Kerr et al. 2006; Wayman et al. 2012c). Supovitz and Klein (2003) examined data use at various levels in school contexts and found that a sense of “inquiry” or a positive data-using culture developed when “...uses of data were regular parts of school life” (p. 2). Data use that is sporadic cannot feed a rich culture of inquiry; Valli and Buese (2007) noted that the lack of time in the day to engage in thoughtful deliberation around student needs, combined with pressures to comply with curricular and institutional directives, created a situation in which teachers felt that their relationships with students suffered as a focus on data-related activities increased.

The practice of foregrounding spreadsheets or data reports and backgrounding students whose data are represented in those reports is counter to much research that suggests good data use should be integrated into a “teacher-as-learner” model of data use (Hamilton et al. 2009; Mandinach and Jackson 2012; Wayman and Stringfield 2006). In such a model, teachers engage in data-informed inquiry to assess the needs of the whole child (as opposed to the child as test taker) and then work collaboratively and creatively to address those needs. Collaboration is essential as it increases the accuracy of interpretations of data (Means et al. 2011) and provides the opportunity for synergistic problem-solving (Fullan 2007; Kerr et al. 2006; Mandinach and Jackson 2012; Means et al. 2011; Schildkamp and Kuiper 2010). However, this model requires leaders to put in place supports to provide teachers regular, structured time devoted to collaborative data use, instead of simply layering data use atop other responsibilities and expecting teachers to “find the time” to use data well (Mandinach and Jackson 2012; Wayman and Cho 2009; Wayman et al. 2012c).

Because the lack of structures useful in guiding collaborative time has been identified as a barrier to effective data use (Ikemoto and Marsh 2007; Kerr et al. 2006), items were developed to assess the degree to which teachers perceived their schools as characterized by a culture of collaboration (see Table 8). In addition to the item specific

to time in the “Culture of Collaboration” scale, items specific to the presence or absence of Professional Learning Communities (see DuFour et al. 2005) and the degree to which PLC-oriented dialogues incorporated data use were drafted (see Table 10).

4.2 Stage 2: instrument strengthening and revision

Beyond being rooted in empirical research, development of the *S-DUPL* had to meet standards for validity and item stability to be of use to evaluators and practitioner-leaders. Groves et al. (2009) caution that a foundational aspect of any survey is whether the instrument accurately captures the information sought—that is, whether the instrument can be said to be a valid and reliable measure of particular concepts. To this end, upon completion of initial drafting of items, and prior to a pilot deployment, the instrument was honed and analyzed through a series of expert reviews and cognitive interviews (Desimone and LeFloch 2004; Groves et al. 2009).

A panel of 12 reviewers was purposively sampled from among the types of persons likely to respond to the instrument. Thus, four central office administrators, four campus principals, and four teachers were asked to participate in structured review process. Table 1 provides some demographic information about the reviewers, including years of experience in the current position and total years of experience in EC-12 education. Additionally, the 12 reviewers represented eight unique school districts. This was a desirable in order to preclude a cluster of participants from reading a question in the same way due to a particular district culture or an emphasis on a particular language in any given district. Also, diversity in reviewers was desirable because the instrument was not being developed for any one particular context; thus, the

Table 1 Characteristics of expert reviewers

Reviewer	Title/role	Years of experience	Gender
1	Principal of a mid-size urban elementary	7 (current) 18 (total)	Female
2	Early childhood teacher at a small, specialized elementary campus	4 (current) 8 (total)	Female
3	Director of Accountability and Assessment of a large suburban district	6 (current) 19 (total)	Female
4	Director of Assessment at a mid-size suburban elementary	2 (current) 27 (total)	Female
5	Principal at a small, rural middle school	3 (current) 15 (total)	Male
6	Fourth grade teacher at a large urban elementary school	1 (current) 5 (total)	Female
7	Assistant superintendent of a mid-size suburban district	6 (current) 27 (total)	Female
8	English teacher at a large suburban middle school	4 (current) 8 (total)	Female
9	Principal of a mid-size urban elementary	4 (current) 16 (total)	Male
10	Social studies teacher at a large suburban high school	4 (current) 5 (total)	Male
11	Assistant Superintendent of a mid-size suburban district	7 (current) 40 (total)	Female
12	Principal of a large, urban high school	3 (current) 19 (total)	Female

Mean experience in current position: 4.6 years

Mean total experience, EC-12 education: 18.8 years

items needed to be clear across schooling contexts as well as across levels within a district.

The interviewing and review process lasted approximately 1 h for each participant: During this time, each participant was provided a print copy of the draft instrument and asked to perform a “talk aloud” form of a cognitive interview. That is, the participant read each question aloud and was encouraged to verbalize the thoughts, questions, or frustrations that resulted in making sense of each question and the available answer choices. In this way, unclear or poorly worded questions can be identified by the researcher/evaluator, who takes notes and probes for understanding when points of confusion arise. To better facilitate the process, I not only took notes on an alternate version of the instrument and probed for understanding when issues arose, but I also audio recorded each session to support further analysis. At the conclusion of each talk aloud, the participants were asked to comment on any issues related to the instrument; in this way, issues that might have escaped first notice in item drafting might be addressed during the revision process. This provided a layer of expert review on top of the cognitive interviews.

The cognitive interviewing and expert review process unfolded over a 5-week period and allowed for in-depth analysis of items that led to decisions specific to revisions and deletions. For example, early iterations of the instrument had two sections of items focused on how frequently educators used particular types of data (e.g., interest inventories, Diagnostic Reading Assessment, benchmark exams, curriculum-based assessments (CBAs), and STAAR² exams) and the degree to which respondents perceived these data as useful. However, after multiple reviewers across district contexts reviewed the survey, it became clear that these sections were too context-specific to be of use in a broad instrument. Specifically, the terms “benchmarks” and “CBAs” were used in ways that varied greatly by district context. Some reported that benchmarks were written by district personnel, whereas CBAs were more akin to unit tests, often part of an adopted curriculum or text. For others, this pattern was reversed: Benchmarks were defined as short-cycle measures, often developed by grade-level teams or content area departments, while CBAs signified assessments required by the district and analyzed on a broader scale. Some noted that benchmarks meant tests used to help predict success on the STAAR, while others said that CBAs or “STAAR blueprints” were used for this purpose.

Due to issues like these, some items were determined to be of little use in broad applications of this particular instrument and were eventually removed. In the event the survey is piloted in a particular district context, it may be useful for the evaluator/researcher to develop a list of data and meanings for those data specific to context and then inquire about those context-specific types of data to inform “frequency of use” or “perceived usefulness” analyses. While definitions for some types of data (i.e., state required exams) were consistent across contexts, others were not; without this consistency, the instrument would not have enabled valid and reliable comparisons of the frequency or perceived usefulness of multiple types of data across districts.

² The State of Texas Assessment of Academic Readiness is the current state-mandated standardized exam for students in grades 3–10; subjects tested vary by grade level.

In another vein, the process helped reshape items. For example, a reviewer who was a principal worked through items that initially used the phrasing, “teaching”:

Reviewer 1: “I use a variety of information to inform my teaching.” Yeah, um... once again, that’s a hard question for administrators or central office staff. [Pause] Maybe “my practice”? [...]

In response to this item creating a stumbling block for multiple reviewers who were campus administrators or central office personnel, the item wording was revised to, “I use a variety of information to inform my teaching and/or daily practice.” The latter wording brought clarity to later reviewers and enabled the item to be used with teachers and administrators alike.

As another example, one item was initially worded, “Data use is all about accountability.” This gave rise to the following dialogue with reviewer 3—a Director of Accountability and Assessment:

Reviewer 3: “Data use is all about accountability.” [Pause] I’d say “Agree.” I was hinging on “strongly agree” because unfortunately the way our system is set up, it has been about TAKS scores, but I see data use beyond our traditional accountability. I mean—I see “accountability” beyond state [test scores]. And it should be “strongly agree”—it is all about accountability. When you have those kids, every day, for the entire year—did you teach those kids? That’s accountability.

Researcher: This is one I’ve considered rewording. In a prior study there were teachers who used terms like, “it’s a game”—“this data use is just about an accountability game.” But I hesitated to put that term in because I worried it would be too leading. [...] If this said “about the accountability game” would you have answered that differently?

Reviewer 3: I would have. If you asked if data use was all about playing the “accountability game” I would say “strongly disagree” because immediately when I hear “playing the accountability game” I know what they’re getting at. [...] The way I read “accountability”—that’s why I wanted you to define “accountability” because the way I read it was, it’s accountability for teachers—not just scores. Like attendance—if my kids aren’t coming to school and I haven’t called home, I play a role in that—I shouldn’t leave it to the attendance person or the AP or the Principal. I should be calling. [...] So you might want to put “accountability rating” or “the rating game...”

This dialogue led to the realization that for some teachers and leaders, accountability was used synonymously with “responsibility.” This led to rephrasing the item—which was created in part to assess whether data was linked closely with the formal accountability system—to “Data use is all about accountability ratings,” which carried the intended meaning more effectively.

Through this process of engaging with experts, consulting session notes, and reviewing recordings, the instrument was revised every three to four reviewers. In this manner, the face validity of the instrument was continually and iteratively addressed

and improved. By the end of the process, an improved instrument was prepared for piloting via Qualtrics electronic survey software.

4.3 Stages 3, 4, and 5: piloting, revision, and statistical analyses of items

The last three stages of the process occurred sequentially, but for purposes of data representation and discussion are presented together, by scale. In this section, I describe six scales and some individual items. Item and scale consistency were assessed via two methods: Cronbach's alpha and bivariate correlation. These measures are reported for the initial small-scale pilot (P1) and for the final revision (P3). As participants in the districtwide pilot (P2) took the survey once, no test-retest measure is reported. Where results are reported as "N/A," the item did not appear on the instrument in substantially similar format at that time.

Participant pools and procedures Table 2 presents data specific to each participant pool, including the total number of participants completing the survey. Participants in pilots 1 and 3 were invited to take the survey online; convenience sampling was used to select persons for pool 1, and persons responding to a general email invitation issued via a principal's network were invited to participate in pool 3. All professional educators (i.e., teachers, administrators, and degreed support staff) in the school district involved in the districtwide pilot were invited to participate.

Participants in pilots 1 and 3 were asked to complete the survey via Qualtrics software twice, 10–14 days apart. For pilot 3, through the support of grant funds, those who agreed to complete both administrations were provided with a small incentive for participating. Both administrations were scheduled to occur during an off-contract window for participants to minimize the potential of a treatment or other professional activities (for example, a professional development event or conference) affecting responses. Item stability was then analyzed via SPSS with bivariate correlation procedures (for items) and Cronbach's alpha (for scale reliability analyses). To provide a

Table 2 Descriptions of participant pools for pilots 1, 2, and 3

	Pilot 1	Pilot 2	Pilot 3
Instrument response rate	30/34 (88.2 %)	154/184 (83.7 %)	111/120 (92.5 %)
Mean experience in EC-12 education	18.2 years	13.1 years	13.1 years
Mean experience in current position	6.4 years	6.3 years	5.1 years
Positions currently occupied			
Classroom teachers	63.3 %	72.8 %	73.0 %
Campus administrators	23.3 %	6.9 %	17.1 %
Central office leaders	6.7 %	9.4 %	1.8 %
Educators in support roles (i.e., diagnosticians, instructional or data coaches, specialists)	3.3 %	4.3 %	5.4 %
Counselors, librarians	3.3 %	3.7 %	2.7 %
Other	n/a	2.9 %	n/a

conservative analysis, absolute, rather than consistent, agreement was chosen as the standard for bivariate correlation within SPSS. Participants in the districtwide pilot (pilot 2) were invited to participate via email in August 2012 and were allowed a 2-week window to complete the instrument. Two reminder emails during that window assisted in increasing the response rate. Cronbach's alpha was again used in analysis of scales

Scale 1: confidence The Confidence scale consisted of six items that assessed reported confidence in skills and abilities pertinent to data use (see Table 3). Analyses across the three participant pools suggest that the items in this scale do correlate from T1 to T2 at statistically significant rates—that is, T2 responses can typically be predicted from T1 responses. However, for research purposes, the r^2 value would be lower than optimal for use as stand-alone items. To get a sense of reliability for the scale as a whole, I further analyzed data from pilot 3, the most developed iteration of the instrument, by correlating scale data from the T1 and T2 administrations; this revealed that the scale as a whole produced stable results over time ($r^2=.77$, $n=109$). Thus, evaluators who wish to get a broader sense of educator confidence in skills and abilities pertinent to data use may find the scale, used in conjunction with open-ended queries or other qualitative procedures, useful.

Scale 2: effectiveness of data-related professional learning The “Effectiveness” scale consisted of five items that assessed perceptions regarding the utility of current

Table 3 Analyses of the “Confidence” scale and of individual items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. I can easily access and navigate district data systems. ^a	$r^2=.54^{**}$ ($n=28$)		$r^2=.53^{**}$ ($n=109$)
2. I am comfortable collaborating around data with colleagues.	$r^2=.71^{**}$ ($n=28$)		$r^2=.62^{**}$ ($n=109$)
3. I can formulate worthwhile questions to guide my data use.	$r^2=.79^{**}$ ($n=28$)		$r^2=.57^{**}$ ($n=109$)
4. When I examine data reports, I am confident that my interpretations are accurate.	$r^2=.85^{**}$ ($n=28$)		$r^2=.52^{**}$ ($n=109$)
5. Once I analyze data and draw conclusions, I know what action steps to take next.	$r^2=.59^{**}$ ($n=28$)		$r^2=.58^{**}$ ($n=109$)
6. I am able to support others in learning how to effectively use data.	$r^2=.81^{**}$ ($n=27$)		$r^2=.68^{**}$ ($n=109$)
Scale reliability (α)	$\alpha_{(T1)}=.87$ ($n=27$) $\alpha_{(T2)}=.91$ ($n=29$)	$\alpha=.88$ ($n=150$)	$\alpha_{(T1)}=.87$ ($n=110$) $\alpha_{(T2)}=.90$ ($n=109$)

^a Unless otherwise noted, all items were set on a scale that included the answer choices, “strongly disagree,” “disagree,” “neither agree nor disagree,” “agree,” and “strongly agree”

*Correlation significant at the $p<.05$ level

**Correlation significant at the $p<.01$ level

professional learning supports related to data use (see Table 4). Analyses across the three participant pools suggest that the items in this scale do correlate from T1 to T2 at statistically significant rates—that is, T2 responses can typically be predicted from T1 responses. However, for research purposes, the r^2 value would be lower than optimal for use as stand-alone items, and responses to item 4, “When I’ve participated in professional learning related to data use, I leave knowing how I can apply it in my practice,” should be interpreted with great caution, if at all, apart from other data points. The scale as a whole was stable from T1 to T2 ($r^2=.77$, $n=109$).

Again, evaluators or leaders might use this scale to highlight areas of strength or concern which could then be explored in more depth. However, as individual items lack a high degree of stability over time, other data points (interviews, focus groups, observations, etc.) should be used to complement and triangulate these data.

Scale 3: construal of data The “Construal of Data” scale consisted of three items that assessed the essence of various elements used to inform instruction (see Table 5). The items in this scale were initially drafted to press on a suspicion that educators in varying roles might prefer different terms for data. That is, administrators might be more comfortable with data, as a linguistic marker for what was used to direct planning and decision-making, but teachers might prefer evidence or information as these terms avoided accountability system-laden connections. These suspicions did not materialize (see Jimerson and McGhee 2013); however, the items formed a workable scale that fit with other evidence from the district-wide pilot to suggest the scale effectively

Table 4 Analyses of the “Effectiveness” scale and of individual items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. My district’s professional learning has been adequate for helping me learn to use data systems.	$r^2=.57^{**}$ ($n=28$)		$r^2=.61^{**}$ ($n=109$)
2. My district’s professional learning has helped me learn to analyze and interpret data.	$r^2=.54^{**}$ ($n=28$)		$r^2=.58^{**}$ ($n=109$)
3. The professional learning in this district has enabled me to use data effectively to inform what I do on an everyday basis.	$r^2=.56^{**}$ ($n=28$)		$r^2=.68^{**}$ ($n=109$)
4. When I’ve participated in professional learning related to data use, I leave knowing how I can apply it in my daily practice.	$r^2=.59^{**}$ ($n=28$)		$r^2=.47^{**}$ ($n=109$)
5. The professional learning I am provided in this district helps me use data in ways that benefit my teaching and students’ learning.	$r^2=.30$ ($n=28$)		$r^2=.62^{**}$ ($n=109$)
Scale reliability (α)	$\alpha_{(T1)}=.89$ ($n=28$) $\alpha_{(T2)}=.92$ ($n=29$)	$\alpha=.94$ ($n=150$)	$\alpha_{(T1)}=.89$ ($n=110$) $\alpha_{(T2)}=.94$ ($n=109$)

*Correlation significant at the $p<.05$ level

**Correlation significant at the $p<.01$ level

Table 5 Analyses of the “Construal” scale and of individual items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. I use a variety of data to inform my teaching and/or daily practice.	$r^2 = .68^{**}$ ($n=28$)		$r^2 = .63^{**}$ ($n=109$)
2. I use a variety of information to inform by teaching and/or daily practice.	$r^2 = .40^*$ ($n=28$)		$r^2 = .42^{**}$ ($n=109$)
3. I use a variety of evidence to inform my teaching and/or daily practice.	$r^2 = .44^*$ ($n=27$)		$r^2 = .54^{**}$ ($n=109$)
Scale reliability (α)	$\alpha_{(T1)} = .68$ ($n=27$) $\alpha_{(T2)} = .81$ ($n=29$)	$\alpha = .85$ ($n=150$)	$\alpha_{(T1)} = .82$ ($n=110$) $\alpha_{(T2)} = .89$ ($n=109$)

*Correlation significant at the $p < .05$ level

**Correlation significant at the $p < .01$ level

measured to what degree educators asserted using *something* apart from intuition or hunches to guide instructional planning.

Analyses across the three pools suggest that the items in this scale do correlate from T1 to T2 at statistically significant rates. Yet, the r^2 value would be lower than optimal for use as stand-alone items, and responses to any of these individual items should be interpreted with great caution, if at all, apart from other data points. As a whole, this scale proved only moderately stable over time ($r^2 = .67$, $n=109$) and should be interpreted with caution and, ideally, only in the presence of other data points.

Scale 4: beneficence of data The “Beneficence of Data” scale comprised items that assessed participants’ perceptions whether educator data use makes possible healthy, informed practice (see Table 6). Analyses across the three participant pools suggest that the items in this scale again correlate from T1 to T2 at statistically significant rates. However, for research purposes, the r^2 value would be lower than optimal for use as stand-alone items, particularly for items 1 and 2. Analyses of data from pilot 3 once again indicated that the scale as a whole did prove stable over time ($r^2 = .77$, $n=109$). This scale might prove useful to evaluators or district planners in gauging the “reputation” that data use has achieved in a given educational context.

Scale 5: data anxiety The “Data Anxiety” scale comprised three items that assessed participants’ level of concern or worry regarding possible misuse or abuse of data (see Table 7). Analyses across the three participant pools suggest that the items in this scale again correlated from T1 to T2 at statistically significant rates. However, for research purposes, the r^2 value is only marginally stable over time; these items should be interpreted only in concert with other data points for optimal decision making. Analyses for pilot 3 again indicated that the scale as a whole was stable over time ($r^2 = .72$, $n=109$). This scale (along with item 1 in Table 12) may help incoming administrators gauge the degree to which faculty embrace or resist data use, which could be useful in planning early steps toward establishing a healthy, collaborative campus culture. In effect, it may highlight contexts where new leaders may have to do work to make up for the prior bad acts of leaders who misused or abused data.

Table 6 Analyses of the “Beneficence of Data” scale and of individual items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. Data use helps me make informed decisions.	$r^2 = .48^*$ ($n=28$)		$r^2 = .56^{**}$ ($n=109$)
2. Data use helps me make ethical decisions.	$r^2 = .71^{**}$ ($n=28$)		$r^2 = .54^{**}$ ($n=109$)
3. Data use benefits educators and students.	$r^2 = .64^{**}$ ($n=28$)		$r^2 = .70^{**}$ ($n=109$)
4. Data use is about continuous improvement in the classroom.	$r^2 = .71^{**}$ ($n=28$)		$r^2 = .69^{**}$ ($n=109$)
Scale reliability (α)	$\alpha_{(T1)} = .80$ ($n=27$) $\alpha_{(T2)} = .85$ ($n=29$)	$\alpha = .78$ ($n=150$)	$\alpha_{(T1)} = .77$ ($n=110$) $\alpha_{(T2)} = .81$ ($n=109$)

*Correlation significant at the $p < .05$ level

**Correlation significant at the $p < .01$ level

Scale 6: culture of collaboration The “Culture of Collaboration” scale consisted of five items that assessed participants’ level of concern or worry regarding possible misuse or abuse of data (see Table 8). Analyses across the three participant pools suggest that the items in this scale correlate from T1 to T2 at statistically significant rates. However, for research purposes, the r^2 value is only marginally stable over time, particularly for items 1, 2, 3, and 5: These should be interpreted cautiously if at all as individual items, and then only in concert with other data points. Analyses of pilot 3 data demonstrated that the scale itself was stable over time ($r^2 = .76$, $n=109$).

Nonscale items: vision/rationale block In addition to the items included in the scales, other items comprised blocks of questions but did not comprise scales. These included a block of items aimed at assessing how well a district or campus leadership team had communicated a clear vision and rationale for data use (see Table 9). These blocks of

Table 7 Analyses of the “Data Anxiety” scale and of individual items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. I worry that data will be used to shame or punish teachers at my school.	$r^2 = .69^{**}$ ($n=27$)		$r^2 = .66^{**}$ ($n=109$)
2. I am concerned that data use will increase unhealthy competition among educators at my campus.	$r^2 = .66^{**}$ ($n=27$)		$r^2 = .63^{**}$ ($n=109$)
3. I am concerned that data use will increase unhealthy competition among educators in my district.	$r^2 = .43^*$ ($n=27$)		$r^2 = .64^{**}$ ($n=109$)
Scale reliability (α)	$\alpha_{(T1)} = .81$ ($n=27$) $\alpha_{(T2)} = .84$ ($n=27$)	$\alpha = .88$ ($n=150$)	$\alpha_{(T1)} = .86$ ($n=110$) $\alpha_{(T2)} = .86$ ($n=109$)

*Correlation significant at the $p < .05$ level

**Correlation significant at the $p < .01$ level

Table 8 Analyses of the “Culture of Collaboration” scale and of individual items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. My supervisor regularly models data use.	$r^2=.16$ ($n=28$)		$r^2=.48^{**}$ ($n=109$)
2. When I collaborate with others, we use data to inform our dialogues and decision-making.	$r^2=.46^*$ ($n=28$)		$r^2=.63^{**}$ ($n=109$)
3. I get useful information when I engage in collaborative data use.	$r^2=.57^{**}$ ($n=28$)		$r^2=.57^{**}$ ($n=109$)
4. I am comfortable talking about data with my supervisor.	$r^2=.82^{**}$ ($n=27$)		$r^2=.69^{**}$ ($n=109$)
5. I have adequate time to engage in collaborative data use.	$r^2=.71^{**}$ ($n=27$)		$r^2=.62^{**}$ ($n=109$)
Scale reliability (α)	$\alpha_{(T1)}=.62$ ($n=27$) $\alpha_{(T2)}=.80$ ($n=29$)	$\alpha=.78$ ($n=150$)	$\alpha_{(T1)}=.64$ ($n=110$) $\alpha_{(T2)}=.78$ ($n=109$)

*Correlation significant at the $p<.05$ level

**Correlation significant at the $p<.01$ level

items—and some stand-alone items—are described by providing the test-retest correlations from pilot 1 and pilot 3 and by offering any triangulating data from pilot 2.

The first block of items dealt with whether a clear vision and direction for data use had been established—by either district leaders or by a campus principal—and whether the respondents perceived they had a clear idea of how and why there were supposed to be using educational data. These items were relatively stable over time in the final iteration of the instrument and could potentially be applied by evaluators or district leaders wanting to get a sense of unity pertaining to a vision or rationale for data use. However, these items should be but a starting point: As other work has suggested, educators may report having a “clear sense” of a vision for data use then vary greatly even within campuses as to what that vision *is* (Wayman et al. 2007; Wayman et al. (2012). These items may contribute to an understanding of educator confidence in terms of vision or rationale for data use, but they will present an accurate rendering only

Table 9 Analyses of individual items, “Vision/Rationale” block

Items	Pilot 1	Pilot 2	Pilot 3
1. My district has articulated a clear rationale for data use.	$r^2=.47^*$ ($n=27$)		$r^2=.61^{**}$ ($n=109$)
2. My principal has articulated a clear rationale for data use.	$r^2=.21$ ($n=27$)		$r^2=.68^{**}$ ($n=109$)
3. I know what I’m supposed to be doing with data.	$r^2=.80^{**}$ ($n=27$)		$r^2=.65^{**}$ ($n=109$)
4. I understand how data use fits with our campus and district goals.	$r^2=.71^{**}$ ($n=28$)		$r^2=.63^{**}$ ($n=109$)

*Correlation significant at the $p<.05$ level

**Correlation significant at the $p<.01$ level

when used in conjunction with open-ended questioning or other means to allow deeper exploration of the concepts.

Nonscale items: professional learning community block A second block of items aimed at exploring the presence and data-orientation of any Professional Learning Communities (PLCs). PLCs involve ongoing, inquiry-based, collaborative learning in which small groups of educators bring data and research to bear on questions of contextual significance (DuFour et al. 2005). These items were relatively strong in terms of stability (see Table 10). Due to skip questioning embedded in the electronic version of the instrument, respondents answered questions 2, 3, and 4 only if they responded to item 1 with a “yes.”

Beyond displaying high stability, these items may be useful to district leaders and evaluators. For example, if a district provides time and resources for PLC support, but 10–15 % of respondents say they “aren’t sure” of whether they participate in a PLC, and still others respond in the negative, this can be an indication of a poorly communicated or poorly implemented policy. Further, some educators may be using the term “PLC” as a synonym for “team meeting”—an inaccurate use of the term. PLCs are designed to be regular occurrences and to be steeped in inquiry and the use of data and research to inform small group learning and subsequent decision making (DuFour et al. 2005). Thus, if only 40 % of respondents indicate that the PLC meets at least monthly, or if roughly a third of respondents indicate that data are used infrequently or never, leaders may wish to further examine how the PLC format has been implemented and resourced.

Nonscale items: planning for data-related professional learning block A third block of items sought to collect data around planning for professional learning as it

Table 10 Analyses of individual items, “PLC” block

Items	Pilot 1	Pilot 2	Pilot 3
1. Do you currently participate in a “Professional Learning Community” (PLC) at your campus or within your district? ^a	$r^2 = .78^{**}$ ($n = 27$)		$r^2 = .62^{**}$ ($n = 109$)
2. Is there time built into the school schedule (apart from your regular conference period) for PLC collaboration? ^b	$r^2 = .82$ ($n = 10$)		$r^2 = .78^{**}$ ($n = 36$)
3. How often does your PLC meet? ^c	$r^2 = .81^{**}$ ($n = 10$)		$r^2 = .89^{**}$ ($n = 36$)
4. How frequently does your PLC examine data to inform discussions or decision-making? ^d	$r^2 = .65^*$ ($n = 10$)		$r^2 = .83^{**}$ ($n = 36$)

*Correlation significant at the $p < .05$ level

**Correlation significant at the $p < .01$ level

^a Possible answers included “yes,” “no,” and “I’m not sure”

^b Answer choices included “yes” or “no”

^c Answer choices included “annually or 1–2 times per year,” “2–3 times per semester,” “monthly,” “2–3 times per month,” “weekly,” “2–4 times per week,” and “daily”

^d Answer choices included “never,” “rarely,” “sometimes,” “often,” and “at every meeting”

pertained to data systems and data use. These four items (see Table 11) proved marginally stable over time. Because research indicates that a critical element of effective professional learning is that it is embedded in content-rich experiences—that is, it relates to the day-to-day jobs that teachers undertake (Guskey and Yoon 2009; Desimone et al. 2002)—district leaders may gain insight from items 3 and 4 into how teachers perceive their day-to-day work to be connected to data use-related supports. Certainly, data from these items should be triangulated with feedback gained from audits and observations of professional learning events and PLC-embedded learning.

Items 1 and 2 provide insight as to the culture underlying planning for data use-related supports. The heart of professional learning should involve a servant-leader orientation: Like good classroom instruction, professional learning should meet the educator at a point of need, challenge the learner, and provide the scaffolded supports needed to move the learner to a greater understanding or skill set. Planners should not assume that all teachers need the same training on the same data system on the same day (see Wayman et al. 2012b; Darling-Hammond et al. 2009).

Nonscale items: stand-alone items Finally, four items on the instrument were determined to be useful as stand-alone items (see Table 12). These items proved relatively stable over time in both test-retest pilots. Further, in each of these cases, the trends reflected in these items aligned with other survey and qualitative data collected via the district-wide pilot study. For example, teachers and school leaders offered numerous comments (in response to open-ended survey items as well as in interviews and focus groups) that described firsthand experiences with leaders who used data to shame or punish teachers; this aligned with approximately 40 % of teachers who reported experiencing prior data misuse/abuse.

These items can provide leaders and evaluators with insight into various aspects of the underlying culture of a campus or district. Item 1 can help new leaders understand why some teachers might be justifiably reticent to engage in data use and can also help them gauge the magnitude of that population. Items 2 and 4 can serve as indicators of how committed individuals already are to steeping day-to-day work in inquiry and data-informed practice. These items can point to existing strengths on which leaders can build, without assuming that one must “recreate the wheel.” Finally, as student-involved data use becomes more a norm of practice than an exception (see Kennedy

Table 11 Analyses of individual items, “Planning for professional learning” block

Items	Pilot 1	Pilot 2	Pilot 3
1. Planners of professional learning ask me about my needs regarding data use.	$r^2 = .64^{**}$ ($n=28$)		$r^2 = .62^{**}$ ($n=109$)
2. When I share ideas for using data, leaders listen.	$r^2 = .56^{**}$ ($n=28$)		$r^2 = .61^{**}$ ($n=109$)
3. Professional learning related to data use also relates to the content area/grade level I teach.	$r^2 = .19$ ($n=28$)		$r^2 = .59^{**}$ ($n=109$)
4. I learn something about using data or data systems in most professional learning sessions I attend.	$r^2 = .70^{**}$ ($n=27$)		$r^2 = .59^{**}$ ($n=109$)

*Correlation significant at the $p < .05$ level

**Correlation significant at the $p < .01$ level

Table 12 Analysis of stand-alone items

Items in scale	Pilot 1	Pilot 2	Pilot 3
1. I have worked in one or more schools where data were used to punish or shame teachers.	$r^2 = .76^{**}$ ($n=27$)		$r^2 = .62^{**}$ ($n=109$)
2. I generate my own data when I have questions about instruction.	$r^2 = .75^{**}$ ($n=27$)		$r^2 = .77^{**}$ ($n=109$)
3. I believe it is important to involve students in examining their own data.	<i>Item not included</i>		$r^2 = .69^{**}$ ($n=109$)
4. If no one (on my team, in my department) wanted to use data, I would still make the effort to do so.	$r^2 = .69^{**}$ ($n=27$)		$r^2 = .74^{**}$ ($n=109$)

*Correlation significant at the $p < .05$ level

**Correlation significant at the $p < .01$ level

and Datnow 2011; Marsh et al. 2014), leaders can use item 3 to assess the degree to which teachers assert this practice is essential; results from this item could therefore serve as a starting point to ongoing dialogues about how best to engage students in tracking and analyzing their own data.

5 Applications, limitations, and a look ahead

The expectations from taxpayers and policy makers that educators use data to inform practice—and to continually improve practice in equitable ways—will only continue to increase. As district/school system leaders, evaluators, and researchers work diligently to implement robust data use practices and expectations, the instrument described in this paper may prove useful in several ways. In this section, I outline potential applications of the instrument (in whole or in part), describe some limitations of the instrument and cautions for use, and discuss how the process described can be used to inform the creation, deployment, and analyses of similar instruments.

5.1 Applications in evaluation and research

The most immediate application of the instrument would be to use the items as a first step in gauging attitudes and perceptions toward data use among teachers at a campus or in system of schools. Work in the area of implementation consistently indicates that context matters: A program implemented in context A may need to be adjusted for success in context B (Datnow 2006; Honig 2006). Yet, system leaders may be tempted to overlay a process or expectation from another context without carefully considering current strengths or needs in the target context. In either situation (i.e., co-construction or implementation with an expectation of high levels of fidelity to model), the instrument could be used to provide baseline data prior to the implementation of new procedures or professional learning supports; such data could be used to inform ongoing evaluation to assess progress over time.

Use of evaluations regarding perceptual data may also help incoming leaders (at a campus or system level) gauge the tenor of the present “data use culture” so that they know what initiatives may be implemented with gusto or where more time engaging in dialogue may be warranted in order to mitigate past abuses/misuses of data. In writing about inquiry-based practice, Rallis and Rossman (2012) argue:

...the trustworthiness, or rigor, of a study should depend not just on whether the researcher got the technical matters right—whether about instrumentation or the protection of human subjects. Trustworthiness should also be judged by how well the researcher got the relational matters right. (p. 73)

Rallis and Rothman’s focus on the relational aspects of inquiry applies to school improvement efforts in two ways. First, relational matters, including respect for persons and transparency, are always essential in conducting assessments or evaluations of initiatives. However, these may be more pronounced considerations for internal evaluators or when district leaders undertake use of the instrument to explore perceptions related to data use within the district/system. Establishing that those conducting the evaluation are as interested in teacher perspectives as with outcomes may garner buy-in to the process.

Second, effective data use in and of itself requires transparency, clear direction, and a culture of trust among teachers and leaders (Earl and Katz 2006; Louis 2007; Mandinach and Jackson 2012). Appropriate use of the instrument may help leaders identify existing barriers to this trust so that data use processes can either be strengthened or, with great effort, transitioned from dysfunctional processes to healthy, productive processes that allow egos and fears to be set aside for the good of identifying and meeting student needs.

In terms of more in-depth analysis for research and evaluation purposes, the instrument provides a way for investigators to combine and overlay items to gain insight into particular questions of importance. For example, I have been able to use instrument data to examine how reported anxiety around data use relates to whether teachers report having personally experienced data misuse or abuse; unsurprisingly, those who have experience such misuse or abuse have higher anxiety when asked to engage in data-informed practice. I have worked with colleagues to explore whether persons tapped as mentors for early career teachers differ in attitudes and perceptions around data use than their non-tapped peers (and from the early career teachers themselves). Other areas for exploration include the following:

- Is reported commitment to data use related to prior reported experience with data misuse/abuse?
- To what degree do teachers feel “heard” when proposing ideas for professional learning related to data use?
- Are there differences by role (i.e., teachers, campus leaders, or district-level leaders) in whether educators attribute beneficent characteristics to data use?

By making these items widely available and by being transparent about the validity and reliability of the items, I hope to support feasible evaluation for leaders in district contexts.

5.2 Applications and implications: leading schools and school systems

Certainly, survey instruments such as the *S-DUPL* may be used to gauge measures relating to program implementation and progress, and this may inform leaders in terms of when to make adjustments to better fit context. However, as noted throughout the paper, educational data use is no “magic bullet” to increasing student achievement; as Louis et al. (2010) note, *how* data use is implemented and supported matters. And, how data use is implemented must take into account the ways in which evidence or explicit knowledge gained via the survey instrument is combined with educators’ professional judgment or “practical wisdom” in the decision-making process (see Brown 2014). To this end, the *S-DUPL* may be used not only to facilitate the collection of data that can provide clues as to the conditions under which data use is being implemented in a given context (useful as that may be) but also as an artifact itself around which system and campus leaders gather to excavate their own tacit understandings and beliefs about the role of data in informing classroom practice.

Too often, data use signals an expectation that formal, “subjective” forms of evidence should be privileged in decision-making processes, at the expense of professional judgment (Jimerson and McGhee 2013; Valli and Buese 2007). This situation is akin to the way in which Brown (2014) characterizes education policy-making and “evidence use”—that is, as a process that presumes:

...the mind of the policymaker must be “empty” of knowledge in relation to a given issue until they have been provided with evidence in relation to it. Patently, this cannot be true: policymakers will have already digested research in relation to a given issue before they are specifically required to tackle a given problem. As such, adopting a phronetic³ approach illustrates the fallacy of conceiving of evidence use as something separate from policy development: that instead, we must recognize that policy-makers and their decisions will already be (explicitly or implicitly) informed by every facet that has shaped their perspective/reality to date; including the evidence and knowledge they have already adopted. (p. 30)

Slightly reworded—doing nothing more than substituting “teacher” or “school leader” for “policymaker,” and “teaching and learning” for “policy development,”—the parallels become particularly instructive. Do system leaders assume that teachers have no knowledge supporting their judgments until data enter the picture in formal and structured ways, or do they encourage teachers to draw from both practical wisdom (informed by experience) and new evidence as they plan, instruct, and assess students? The latter would result in a situation where, teachers (rather than “policymakers”) are “...continuously incorporating the most up to date evidence into their thinking, enabling it to be intuitively conjoined with other pertinent and salient factors in order to provide a holistic and well-rounded decision in relation to given issues” (Brown, p. 19). This way of considering data-informed practice provides a rich description of what educational data use could be if reframed as process that mutually reinforces—rather than replaces—professional judgment.

³ Brown refers to an approach that honors “professional wisdom” as a *phronetic* approach, drawing upon Aristotle’s notion of “phronesis,” which Brown notes is “often translated as ‘practical wisdom.’” (2014, p. 24).

In this vein of thinking, district and campus leaders could use portions of the *S-DUPL* to inform their own evolving professional judgment—to fit new knowledge gained from the instrument with existing beliefs and knowledge in order to build their own capacity for *leading* data use efforts. Leading data use is challenging in and of itself (Earl and Katz 2006; Wayman 2013; Wayman et al. 2012c). With an approach that honors the knowledge gained via experience as well as new, contextualized data collected through instruments like the *S-DUPL*, leaders no longer need to be experts with all answers at the ready; instead, they must be committed to learning from and with those they lead. In this, data generated through the *S-DUPL* may provide an effective learning artifact around which leaders (and teachers) may dialogue.

5.3 Limitations and Considerations

The instrument or a selection of individual items or scales can be informative for the purposes of evaluation or continuous improvement processes; however, those who employ surveys must also take care to approach inquiry thoughtfully and in ways that produce the most credible, accurate results possible. Three cautions arise in considering limitations of this instrument: (1) For maximum usefulness, items and scales should be interpreted in conjunction with multiple other data points, (2) some uses of the instrument may be more at risk for inaccurate results due to respondent trust levels, and (3) for appropriate use, the instrument must be considered a useful tool for measuring issues related to the conditions contextualizing data use in a particular locale, but users must not rely on simple managerial responses to trends identified through use of the tool itself to “fix” mediocre practice.

To the first point, the scales on the instrument proved stable over time to a higher degree than most individual items. This indicates that users can gain knowledge via the scales (and, of course, through the items with higher degrees of stability). Users could accurately take stock of the magnitude of teachers with higher degrees of anxiety around data or of the general sense of whether educators in a district attributed beneficent potential to the use of data. However, one of the tenets of effective data use is that single data points should rarely trigger action or changes in course (Boudett et al. 2005; Shen and Cooley 2008). To be a “data-driven school,” then, this tenet must also be extended to the ways in which evaluators or district leaders use data to inform action.

To this end, users should bring multiple data points to bear on discussion leading to action steps. This means that users would likely need to consider multiple items, multiple scales, items plus open-ended questions or scales used in conjunction with observational data or follow-up interviews to effect the best decisions possible. The instrument can play a part in a broader evidence-gathering process but should not comprise the totality of that process. Taking pains to engage in a broad-based evidence gathering process may help prevent “cherry-picking” data—that is, simply using limited data to justify predetermined decisions (see Coburn et al. 2009).

To the second point, one challenge with any research—survey research included—is the tendency for respondent sufficing or answering in ways that are perceived as “socially desirable” (Groves et al. 2009). While the focus of this instrument is a topic many may see as nonthreatening—what educators think about data use—it clearly *can* be anxiety producing, as some respondents indicated. Further, almost a third of the

respondents in the districtwide pilot noted that they had personally witnessed data misuse or abuse. In an age when accountability data are used for high-stakes decisions (Berliner 2013; Daly 2009), there is a risk that teachers will respond in ways that exaggerate their acceptance of data use as a positive practice and understate their fears or reluctance to use data. This tendency could also skew responses to items that ask teachers to provide feedback on the quality of campus-based or district supports, including how well their leaders model constructive data use.

These tendencies underscore how essential it is for system leaders to be diligent about creating cultures of trust, where feedback is valued even when it is contrary to the dominant narrative. However, they also underscore how cautious leaders and evaluators must be in interpreting survey data and how essential it is to triangulate data produced via any instrument with observations, open dialogues, and document review, where possible.

A desire for social desirability among respondents also points to another consideration for those who would use this (or any) instrument as part of an evaluation or continuous improvement process: Respondents must feel safe if they are to provide candid answers. In each of the pilots described, respondents were assured not only of confidentiality but also of anonymity: The resulting data set included no personally identifiable information. However, respondents may still have been concerned about providing information that could be perceived as making answers identifiable (e.g., years in current position, teaching content area). That I was external to the systems in which these respondents worked may have been helpful, but it also highlights a difficulty for those who would use items or the instrument for purposes of internal evaluation. Internal evaluators or campus/district leaders will likely have to take extra precautions to assure respondents that answers cannot be tracked back to individuals. In a similar vein, when results do not confirm predetermined lines of thinking or are less than flattering, leaders will have to do the difficult work of demonstrating openness to change and dialogue rather than succumbing to reactive or defensive justifications of current practices. These are challenges with which external actors may not have to contend.

To the last caution, not all implementations of “educational data use” are equal—some result in improved outcomes, while others clearly do not (Louis 2007). This is because no evaluation or instrument will in and of itself fix a broken system. Data use, like any other strategy or tool, cannot simply be overlaid in lockstep fashion onto a dysfunctional system and result in benefit. Actors co-construct the ways in which initiatives play out in contexts (Datnow 2006; Park and Datnow 2009), and educator decision-making is continually re-informed and reinvented as practitioners fit experience, professional wisdom, and new evidence together in the decision-making process. Thus, leaders should not expect that one-time use of the instrument and technocratic responses to results will do as much to improve a system as consistent revisiting of progress and expectations in light of the portrait of data use culture painted by analysis of data from the instrument. Use cannot provide simplistic answers, but it can help gauge important issues related to the climate and culture surrounding data use in a given context, such as

- Are teachers anxious when discussing data with supervisors?

- How are leaders framing data use for teachers—what messages are teachers receiving?
- Are PLCs being afforded regular, structured time and are they basing dialogues in evidence?
- Are we spending too much time on data system access at the expense of moving data through interpretation to action?

The *S-DUPL*—like any instrument—is a means to an end, but not an end in and of itself. It cannot declare that a school or district has “arrived” at the end of the data use journey, though it can facilitate the collection of data that help measure progress along the path.

5.4 Moving forward: toward timely and useful evaluation-for-improvement

Moving forward, two considerations for use of the instrument and process described in this paper arise. First, the instrument described can assist time-pressed educational leaders in constructing credible inquiry processes to inform practice. Educational data use has become a ubiquitous concern of researchers, evaluators, system leaders, and policy makers. Schools are expected to be involved in evidence-based continuous improvement processes. Yet, the design of tools and instruments to facilitate this process can be difficult, particularly for educators who already have overly full plates and increasing demands on time. In response, some undoubtedly create their own instruments, and these may vary in quality just as much as those created by educational researchers (see Desimone and LeFloch 2004). To address this need, this paper has described an instrument that can provide evaluators and district leaders with reviewed and tested items and scales to the end of helping inform their work in context.

Second, this paper attempts to reach beyond the description of a single instrument to lay out the story behind the process of development of tools to support inquiry. As we share these narratives with one another and with those who work in school system contexts, we all benefit by reinforcing the rigor needed to create instruments capable of capturing credible, actionable data. We also benefit from reminding one another of the importance of open dialogue and of consulting multiple data points when engaging in major decision-making. Constructive and accurate evaluation is difficult enough without those who engage in research and evaluation not baring their own learning processes. When we as educators and evaluators support one another in developing better instruments, we can produce more accurate data, and the ability for practitioners to make truly evidence-based decisions improves.

6 Conclusion

A perennial concern of researchers is how to produce work that not only informs the field but also supports the improvement of teaching and learning to the benefit of all students. Educational data use promises to be a policy and practice cornerstone that will continue to gain steam in the coming years. However, “data-driven decision making” can be constructive or destructive, depending on the ways in which district leaders and policymakers shape and support data use, and what works in one context may not work

in another. An instrument such as the one described in this paper can help researchers and district leaders inform effective decision-making and can provide a roadmap to the improvement of their own data collection instruments to the end of improved teaching and learning.

Acknowledgements The author wishes to acknowledge and graciously thank Dr. Judy Groulx and the late Dr. Sherrie Reynolds, faculty mentors at TCU whose encouragement and assistance greatly strengthened this work. The author also thanks Dr. Annie Nguyen for her assistance and insights during the preparation of the manuscript.

References

- Anderson, S., Leithwood, K., & Strauss, T. (2010). Leading data use in schools: organizational conditions and practices at the school and district levels. *Leadership and Policy in Schools*, 9(3), 292–327. doi:10.1080/15700761003731492/.
- Berliner, D. (2013). Problems with value-added evaluations of teachers? Let me count the ways! *The Teacher Educator* 48(3), 2013
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268.
- Boudett, K. P., City, E. A., & Murnane, R. J. (Eds.). (2005). *Data Wise: a step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.
- Brown, C. (2014). Advancing policy makers’ expertise in evidence-use: a new approach to enhancing the role research can have in aiding educational policy development. *Journal of Educational Change*, 15(1), 19–36.
- Brown, C. & Rogers, S. (2014). Knowledge creation as an approach to facilitating evidence informed practice: examining ways to measure the success of using this method with early years practitioners in Camden (London). *Journal of Educational Change*, early online publication, DOI 10.1007/s10833-014-9238-9
- Cho, V., & Wayman, J. C. (2014). Districts’ efforts for data use and computer data systems: the role of sensemaking in system use and implementation. *Teachers College Record*, 116(2), 1–45.
- Coburn, C. E., Honig, M. I., & Stein, M. K. (2009). What’s the evidence on districts’ use of evidence? In J. D. Bransford, D. J. Stipek, N. J. Vye, L. M. Gomez, & D. Lam (Eds.), *The role of research in educational improvement* (pp. 67–87). Cambridge: Harvard Education Press.
- Copland, M. (2003). Leadership of inquiry: building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis*, 25(4), 375–395.
- Creswell, J. W. (2003). *Research design: qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: SAGE.
- Daly, A. (2009). Rigid response in an age of accountability: the potential of leadership and trust. *Educational Administration Quarterly*, 45(2), 168–216.
- Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: a status report on teacher development in the United States and abroad*. Dallas, TX: National Staff Development Council.
- Datnow, A. (2006). Connections in the policy chain: the “co-construction” of implementation in comprehensive school reform. In M. I. Honig (Ed.), *New directions in education policy implementation* (pp. 105–124). New York: SUNY Press.
- Desimone, L. M., & LeFloch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Analysis*, 26(1), 1–22.
- Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers’ instruction: results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis*, 24(2), 81–112.
- DuFour, R., Eaker, R., & DuFour, R. (Eds.). (2005). *On common ground: the power of professional learning communities* (pp. 7–29). Bloomington, IN: Solution Tree
- Earl, L. M., & Katz, S. (2006). *Leading schools in a data-rich world: harnessing data for school improvement*. Thousand Oaks, CA: Corwin.

- Elmore, R. F. (2004). *School reform from the inside out: policy, practice, and performance*. Cambridge, MA: Harvard University Press.
- Fullan, M. (2007). *The New Meaning of Educational Change* (4th ed.). New York: Teachers College Press.
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey Methodology* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.
- Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan*, 90(7), 495–500.
- 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.
- Honig, M. I. (2006). Complexity and policy implementation: challenges and opportunities for the field. In M. I. Honig (Ed.), *New directions in education policy implementation* (pp. 105–124). New York: SUNY Press.
- 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* (pp. 105–131). 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.
- Jimerson, J. B., & McGhee, M. (2013). Leading inquiry in schools: examining mental models of data-informed practice. *Current Issues in Education*, 16(1), 1–20.
Retrieved from <http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1163>.
- Jimerson, J. B., & Wayman, J. C. (2015). Professional learning for using data: examining teacher needs and supports. *Teachers College Record*, 117(4), 1–36.
- Johnson, B., & Christensen, L. (2013). *Educational research: quantitative, qualitative, and mixed approaches*. Thousand Oaks, CA: Sage Publications.
- Katz, S., & Dack, L. A. (2014). Towards a culture of inquiry for data use in schools: breaking down professional learning barriers through intentional interruption. *Studies in Educational Evaluation*, 42, 35–40.
- Kennedy, B. L., & Datnow, A. (2011). Student involvement and data-driven decision making: developing a new typology. *Youth & Society*, 43(4), 1246–1271.
- Kerr, K. I., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to promote data use for instructional improvement: actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(August), 496–520.
- Louis, K. S. (2007). Trust and improvement in schools. *Journal of Educational Change*, 8, 1–24.
- Louis, K. S., Leithwood, K., Wahlstrom, K. L., Anderson, S. E., Michlin, M., Mascal, B., Gordon, M., Strauss, T., Thomas, E., & Moore, S. (2010). *Learning from leadership: investigating the links to improved student learning (final report of research findings)*. New York, NY: The Wallace Foundation.
- Mandinach, E. B. (2012). A perfect time for data use: using data-driven decision making to inform practice. *Educational Psychologist*, 47(2), 71–85.
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30–37.
- Mandinach, E. B., & Jackson, S. S. (2012). *Transforming teaching and learning through data-driven decision making (Classroom insights from educational psychology series)*. Thousand Oaks, CA: Corwin.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data-driven decision making: evidence from Recent RAND research*. www.rand.org
- Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making: policy implementation and effects in Florida middle schools. *Educational Policy*, 24(6), 872–907.
- Marsh, J. A., Farrell, C. C., & Bertrand, M. (2014). Trickle-down accountability: how middle school teachers engage students in data use. *Educational Policy*. doi:10.1177/0895904814531653.
- Means, B., Padilla, C., DeBarger, A., & Bakia, M. (2009). *Implementing data-informed decision-making in schools—Teacher access, supports and use*. Washington, D.C.: U.S. Department of Education Office of Planning, Evaluation and Policy Development.
- Means, B., Chen, E., DeBarger, A., & Padilla, C. (2011). *Teachers' ability to use data to inform instruction: challenges and supports*. Washington, CD: U.S. Department of Education, office of Planning, Evaluation, and Policy Development.
- Militello, M., Bass, L., Kackson, K. T., & Wang, Y. (2013). How data are used and misused in schools: perceptions from teachers and principals. *Education Sciences*, 3, 98–120.

- National Forum on Education Statistics [NFES]. (2012). *Forum Guide to Taking Action with Education Data*. (NFES 2013–801). U.S. Department of Education. Washington, DC: National Center for Education Statistics
- Olsen, B., & Sexton, D. (2009). Threat rigidity, school reform, and how teachers view their work inside current education policy contexts. *American Educational Research Journal*, 46(1), 9–44.
- Park, V., & Datnow, A. (2009). Co-constructing distributed leadership: district and school connections in data-driven decision-making. *School Leadership and Management*, 29(5), 477–494.
- Rallis, S. F., & Rossman, G. B. (2012). *The research journey*. New York: Guilford Press.
- Saunders, L. (2014). “Decoro, Sprezzatura, Grazia”—A creative metaphor for teaching and teachers’ learning today? *Professional Development Today*, 16(2), 12–20.
- Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26, 482–496.
- Schildkamp, K., & Teddlie, C. (2008). School performance feedback systems in the USA and in The Netherlands: a comparison. *Educational Research and Evaluation*, 14(3), 255–282.
- Senge, P. M. (2006). *The fifth discipline: the art & practice of the learning organization*. New York: Currency Doubleday.
- Shen, J., & Cooley, V. E. (2008). Critical issues in using data for decision-making. *International Journal of Leadership in Education*, 11(3), 319–329.
- Supovitz, J. (2012). Getting at student understanding—The key to teachers’ use of data. *Teachers College Record* 114(11)
- Supovitz, J. A., & Klein, V. (2003). *Mapping a course for improved student learning: how innovative schools systematically use student performance data to guide improvement*. Philadelphia: Consortium for Policy Research in Education- University of Pennsylvania.
- Valli, L., & Buese, D. (2007). The changing roles of teachers in an era of high-stakes accountability. *American Educational Research Journal*, 44(3), 519–558.
- Verhaeghe, G., Vanhoof, J., Valcke, M., & Van Petegem, P. (2010). Using school performance feedback: perceptions of primary school principals. *School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice*, 21(2), 167–188.
- Wayman, J. C. (2013). Leading data use: pre-service course for principals and superintendents. *The Journal of Educational Research & Policy Studies*, 13(2), 6–13.
- Wayman, J. C., & Cho, V. (2009). Preparing educators to effectively use student data systems. In T. J. Kowalski & T. J. Lasley (Eds.), *Handbook on data-based decision-making in education* (pp. 89–104). New York: Routledge.
- Wayman, J. C., & Jimerson, J. B. (2014). Teacher needs for data-related professional learning. *Studies in Educational Evaluation*, 42, 25–34.
- Wayman, J. C., & Stringfield, S. (2006). Technology supported involvement of entire faculties in examination of student data for instructional improvement. *American Journal of Education*, 112, 549–571.
- Wayman, J. C., Cho, V., & Johnston, M. T. (2007). *The data-informed district: a district-wide evaluation of data use in the Natrona County School District*. Austin: The University of Texas.
- Wayman, J. C., Cho, V., & Shaw, S. (2009). *Survey of Educator Data Use*. Unpublished document
- Wayman, J. C., Cho, V., Jimerson, J. B., Spikes, D. D. (2012a). District-Wide Effects on Data Use in the Classroom. *Education Policy Analysis Archives*, 20 (25). Retrieved [date], from <http://epaa.asu.edu/ojs/article/view/979>
- Wayman, J. C., Jimerson, J. B., & Cho, V. (2012b). Organizational considerations in establishing the data-informed district. *School Effectiveness and School Improvement*, 23(2), 159–178.
- Wayman, J. C., Spring, S. D., & Lemke, M. A., Lehr, M.D. (2012c). Using data to inform practice: effective principal leadership strategies. Paper presented at the 2012 Annual Meeting of the American Educational Research Association, Vancouver, Canada