Collaborative Data Inquiry Practitioner Tookit Validation Study

April 2026

SRI • 333 Ravenswood Avenue • Menlo Park, CA 94025-3493 • 650.859.2000 • www.sri.com

Authors:
Paul Burkander, SRI
Katrina Laguarda, SRI
Kerry Schellenberger, SRI
Selin Capan, SRI
Lindsay Brown, New York University
Candice Bocala, Harvard Graduate School of Education

Acknowledgments:
We are deeply grateful to the instructional leaders, coaches, and teachers who tested the meeting observation tools; completed meeting observations, exit tickets, and end-of-year surveys; and debriefed with us about their experiences. In particular, we thank Ellen Mandinach, Maxwell Yurkofsky, and Jonathan Supovitz, who served as advisors to this study and reviewed a draft of this report. Adam Parrott-Sheffer and Carmen Williams provided feedback on multiple drafts of the Instructional Team Meeting Observation Rubric and carried out an early pilot of the meeting observation tools. Charles Harding and Roxanne Jones provided expert editing and production support. We thank them for all of their contributions.

Suggested citation:
Burkander, P., Laguarda, K., Schellenberger, K., Capan, S., Brown, L., & Bocala, C. (2026). *Collaborative Data Inquiry Practitioner Toolkit Validation Study*. SRI.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant #R305A200292 to SRI International. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Contents

List of Exhibits

Exhibit **Page**

Executive Summary

Study Overview

This report presents findings from a validation study of the Collaborative Data Inquiry Practitioner Toolkit (CDI Practitioner Toolkit). The CDI Practitioner Toolkit operationalizes research-based indicators of effective collaborative data inquiry (CDI) so leaders and coaches can diagnose strengths and target supports.

The study examined whether the components of the CDI Practitioner Toolkit—including the Instructional Team Meeting Observation Rubric (observation rubric), Instructional Team Meeting Exit Ticket (exit ticket), and Instructional Team Member End-of-Year Survey (end-of-year survey)—reliably and consistently capture core dimensions of high-quality CDI meetings and whether those measures are associated with outcomes expected to be positively influenced by productive CDI meetings.

Toolkit Background and Development

Data inquiry, also known as data-driven decision making, is a collaborative process designed to support improvements in instructional practice. The quality of such collaborative processes has been shown to be associated with student achievement and teacher improvement (Goddard et al., 2007; Ronfeldt et al., 2015). Data inquiry models are typically built around continuous cycles of analysis, planning, action, and reflection/assessment. In CDI teams, educators use multiple forms of data (e.g., student assessment scores, student work, qualitative data from student interviews) to collaborate on their instruction, using the data to connect their teaching to student learning (Bernhardt, 2009; Boudett & Bocala, 2025; Boudett et al., 2013; Cosner, 2011; Kerr et al., 2006; Marsh et al., 2006; Stecker et al., 2008).

Where data-driven decision making and data inquiry have been implemented, instructional leaders have long observed that these processes are more productive in some teams than in others (Coburn & Turner, 2011; Datnow et al., 2013; Marsh, 2012). It is therefore important for districts to have a tool for assessing the quality of CDI meetings so they can identify and better support less productive teams. The CDI Practitioner Toolkit is intended to address that need by measuring CDI teams' engagement with data inquiry practices and providing actionable feedback to team members, coaches, school leaders, and district leaders to improve CDI meeting quality. Given the complex set of tasks associated with CDI, instructional teams typically find it challenging to describe all the features of a well-run CDI process. The toolkit operationalizes the findings from the literature to provide specific indicators of team structures and culture, collaborative inquiry, and team member investment. The observation

rubric also provides guidance about what each construct looks like developmentally, which is a useful framework to guide team member self-assessment and reflection as well as coaching.

To create the CDI Practitioner Toolkit's observation rubric and exit ticket, the research team reviewed and summarized previous literature on data-driven decision making, data use, and continuous improvement. We also drew on the theoretical frameworks and practice guides developed by the Data Wise Project (Boudett & Bocala, 2025; Boudett et al., 2005, 2013). In addition to three CDI experts, four Data Wise–certified coach practitioners and project leaders served as advisors to the current study, although the CDI Practitioner Toolkit is intended for use with any team engaged in CDI, not only those trained by Data Wise. Drawing on all of these sources, we identified eight essential constructs of CDI measured by up to six indicators each.

This study builds on early work assessing the feasibility and validity of the CDI Practitioner Toolkit. Early work employed a structured content review panel of academic and practitioner experts, a round of cognitive interviews with instructional leaders, and two rounds of pilot testing in two districts. This study furthers the validation of the CDI Practitioner Toolkit by examining (1) the scale reliability using exploratory factor analyses and the reliability of measures, (2) the internal consistency of toolkit measures using correlations of within-construct measures across toolkit components, and (3) the concurrent and predictive validity of the toolkit using bivariate and multivariate analyses.

Research Questions

The study focused on the following three research questions:

1. Does the CDI Practitioner Toolkit reliably assess CDI meeting quality?
2. To what extent does the CDI Practitioner Toolkit have internal consistency, that is, to what degree are construct-level measures correlated across toolkit components?
3. To what extent does the CDI Practitioner Toolkit demonstrate evidence of concurrent and predictive validity?

Results

Exploratory factor analyses suggested that all construct measures from the observation rubric were highly correlated with each other and loaded onto a single factor measuring observed meeting quality (see Appendix for factor loadings). Exit ticket and end-of-year survey items loaded onto two factors each. Established educator outcome scales and newly created scales showed strong reliability (e.g., Cronbach's alpha ≥ .74 across subscales).

Correlations within the same construct across CDI Practitioner Toolkit components were often similar in magnitude to correlations across different constructs, suggesting that the toolkit components may be measuring different dimensions of CDI participation. This is expected as a

result of differences in the design of and perspectives captured by the toolkit components. For example, observers and team members may have different perspectives on meeting quality, and the observation rubric and exit ticket assess a single meeting whereas the end-of-year survey asks respondents to reflect on all meetings held throughout the school year.

To assess the concurrent and predictive validity of the CDI Practitioner Toolkit,[1] we measured the degree to which factors from the toolkit components are correlated with student achievement and attendance and with the following self-reported end-of-year team outcomes measured on the end-of-year survey:

- Data proficiency, measured using a study-developed scale for team members' feelings of proficiency with tasks such as refining instructional approaches using data.
- Teacher self-efficacy, measured using the Teacher Sense of Efficacy Scale short form (Tschannen-Moran & Hoy, 2001), which includes questions such as "How much can you do to control disruptive behavior in the classroom?"
- Collective efficacy, measured using the Collective Teacher Efficacy Scale short form (Goddard, 2002), which measures the frequency with which statements are true, such as "Teachers in this school are able to get through to difficult students."
- Burnout, measured using the Maslach Burnout Inventory (Maslach et al., 2016), which measures three distinct dimensions of burnout: emotional exhaustion (feeling emotionally overextended), depersonalization (indifference toward students), and reduced personal accomplishment (feeling of competence and achievement). Higher values on the emotional exhaustion and depersonalization scales indicate more burnout, and lower values on the personal accomplishment scale indicate less burnout.

Bivariate analyses showed modest associations between observation rubric constructs and team outcomes, moderate associations for exit ticket constructs, and generally stronger associations for end-of-year survey constructs.

Multivariate models used a two-stage approach to determine the degree to which the toolkit components predict team and student outcomes. First, we used the least absolute shrinkage and selection operator (LASSO) to identify the control variables and the factors across toolkit components most predictive of each outcome. LASSO is similar to ordinary least squares (OLS) regression but includes a penalty term for any nonzero coefficient, forcing the model to select only the most predictive covariates. Second, for each team outcome, we estimated an OLS regression model that included all controls that were selected for at least two team outcomes and the toolkit component factors that were identified by LASSO for any team outcome. We then tested the joint null hypothesis of the coefficients on toolkit components all being zero. Similarly, for each student outcome, we estimated an OLS regression model that

[1] We consider these predictive analyses for the observation rubric and exit ticket, and concurrent analyses for the end-of-year survey.

included all controls identified for any student outcome and the toolkit components that were identified by LASSO for any student outcome.

Each toolkit component factor was selected for at least one team outcome, so all factors were included in the OLS regressions for each team outcome. The joint null hypothesis that all toolkit component factors have zero coefficients was rejected for self-reported data proficiency, teacher self-efficacy, and the personal accomplishment measure of burnout (Exhibit ES-1), indicating a statistically significant relationship between these outcomes and the toolkit components. For student outcomes, the retrospective meeting quality factor from the end-of-year survey was identified as predictive of math achievement, and the satisfaction with meetings factor from the end-of-year survey was predictive of attendance. Both end-of-year factors were therefore included in the OLS regressions for each student outcome. The joint null hypothesis that these factors both have zero coefficients was not rejected for any student outcome.

Exhibit ES-1. Concurrent and Predictive Validity of the CDI Practitioner Toolkit, Team Outcomes

Outcome	Observation Construct Factor	Exit Factor 1	Exit Factor 2	End-of-Year Factor 1	End-of-Year Factor 2
Any team outcomes	✓	✓	✓	✓	✓
Self-reported data proficiency	*	*	*	✓*	✓*
Teacher self-efficacy	*	*	✓*	✓*	✓*
Collective self-efficacy	✓	✓	✓	✓	
Burnout: Emotional exhaustion	✓			✓	
Burnout: Depersonalization			✓		
Burnout: Personal accomplishment	✓*	*	*	*	✓*
Any student outcomes				✓	✓
GPA					
Attendance					✓
Math achievement				✓	
ELA achievement					

Note. Blank cells indicate the CDI Practitioner Toolkit components were not identified by LASSO as predictive and were not statistically signficiant if included in the model. The observation construct factor and exit ticket factors offer evidence of predictive validity; the end-of-year survey factors offer evidence of concurrent validity because both the team outcomes and the end-of-year factors were collected in the same survey administration.

✓= Toolkit components were identified by LASSO as predictive of the outcome.

* = Toolkit components are jointly statistically significant ($p < .05$).

Conclusion

Together, these findings indicate that the CDI Practitioner Toolkit reliably measures dimensions of high-quality CDI meetings. Although the toolkit distinguishes theoretically distinct constructs, such as data use and application of learning, exploratory factor analyses suggest that each toolkit component—for example, the exit ticket alone—captures one or two unique dimensions. Future research should explore whether some items can be removed from each toolkit component without diminishing the concurrent and predictive validity of toolkit measures. Correlations among the theoretical constructs across toolkit components suggest that the different components do not consistently assess the same distinct constructs, consistent with differences in the design of and perspectives captured by the components. Bivariate analyses indicate moderate correlations between constructs and the team and student outcomes theorized to be positively associated with high-quality CDI meetings, with correlations generally being stronger for team member outcomes and measures based on the end-of-year survey. This finding suggests that the toolkit measures important dimensions of CDI meeting quality. Multivariate analyses indicate that each toolkit component contributes unique explanatory power with concurrent and future team member outcomes, but only the end-of-year survey was predictive of student outcomes conditional on baseline characteristics.

1. Introduction

Collaborative Data Inquiry

In K–12 schools, various models of data-driven decision making, data inquiry cycles, and continuous improvement have been widely adopted, often in the context of grade-level or departmental instructional team meetings (Coburn & Turner, 2011; Datnow & Hubbard, 2015; Vescio et al., 2008). The quality of such collaborative processes has been shown to be associated with student achievement and teacher improvement (Goddard et al., 2007; Ronfeldt et al., 2015). Data inquiry is a collaborative process designed to support improvements in instructional practice. Data inquiry models are typically built around continuous cycles of analysis, planning, action, and reflection/assessment. In data inquiry meetings, educators use multiple forms of data (e.g., student assessment scores, student work, qualitative data from student interviews) to collaborate on their instruction, using the data to connect their teaching to student learning (Bernhardt, 2009; Boudett et al., 2013; Cosner, 2011; Kerr et al., 2006; Marsh et al., 2006; Stecker et al., 2008). We use the term *collaborative data inquiry* (CDI) to refer to the core practices common to many of these models.

Collaborative data inquiry: A process in which teams work together to analyze and interpret data, identify patterns and insights, and use this information to inform decisions aimed at improving teaching practices and student learning.

CDI is a complex endeavor for instructional teams. Simply meeting together is not enough: Teams must identify and address a shared problem of practice to ensure their work maintains focus (Horn & Little, 2010). Such a shared problem of practice is necessary to complete the kinds of intensive, complex inquiry required of these teams, who will analyze multiple forms of data including not only test scores but also student work and measures of instructional quality (Ikemoto & Marsh, 2007). Such thoughtful, data-informed planning may provide teachers with an opportunity to use data in a way that reduces the negative pressures of accountability, both through the collaborative nature of the endeavor and by deepening their understanding of students' knowledge and abilities (Mandinach, 2025; Marsh et al., 2006).

Various CDI models and approaches have been widely adopted for more than two decades (e.g., Love, 2009; Schildkamp et al., 2018), and CDI practices can claim substantial portions of available common planning or other time set aside for teachers to collaborate (Boudett et al., 2013; DuFour et al., 2010). Despite significant investments in training and meeting time, instructional leaders have long observed that CDI appears to be more productive in some

teams than in others. It is therefore important for schools and districts to have a tool for assessing the quality of CDI meetings so they can identify and better support less productive teams.

The CDI Practitioner Toolkit is intended to address that need by measuring data team member engagement with and proficiency in data inquiry practices and providing actionable feedback to team members, coaches, school leaders, and district leaders. Given the complex set of tasks associated with CDI, instructional teams typically find it challenging to describe all the features of a well-run CDI process. The toolkit operationalizes the findings from the literature to provide specific indicators of team structures and culture, collaborative inquiry, and team member investment. The observation rubric also provides guidance about what each construct looks like developmentally, which is a useful framework to guide team member self-assessment and reflection as well as coaching.

CDI Practitioner Toolkit Design and Development

The CDI Practitioner Toolkit measures the team's and individual members' engagement with data-driven practices and provides actionable feedback to team members, coaches, and school and system leaders. Toolkit components include a team meeting observation rubric, a team meeting exit ticket, a team member end-of-year survey, a user guide, a debrief guide, and a

CDI Practitioner Toolkit Instruments

The ***Instructional Team Meeting Observation Rubric*** **(observation rubric)** is designed to be used by an external observer (an instructional coach, principal, or district-level instructional leader) to record and rate team processes as they occur over the course of a single meeting. For each construct measured, the rubric includes descriptions of practice at three levels (Emerging, Developing, and Proficient) for up to six indicators related to the construct. Observers assign scores at the indicator level first and then aggregate those scores on a scale of 1–7 to rate the construct, using a holistic approach that considers both indicator-level scores and the overall quality of the evidence and behaviors observed.

The ***Instructional Team Meeting Exit Ticket*** **(exit ticket)** is administered to all team members at the end of an observed meeting. It is intended to collect feedback on the meeting from the participants' perspective, to complement the assessment of external observers. The exit ticket is designed to be short and easy to complete (requiring fewer than 5 minutes). For this reason, it measures a subset of the indicators and constructs covered by the observation rubric.

The ***Instructional Team Member End-of-Year Survey*** **(end-of-year survey)** includes questions about team members' experiences in their instructional team throughout a school year. It complements the exit ticket by capturing team members' perceptions of team meetings over time. All members of an instructional team should respond to the survey, whether or not they attended an observed meeting. The survey takes approximately 25 minutes.

data dashboard. The toolkit is designed to be used after teams have had some engagement with data-driven practices and can be used at several timepoints to monitor team progress. This report presents evidence of the reliability and validity of three instruments designed measure the quality of team interactions during the CDI process (see box).

To create the observation rubric, the research team reviewed and summarized previous literature on data-driven decision making, data use, and continuous improvement. We also drew on the theoretical frameworks and practice guides developed by the Data Wise Project (Boudett et al., 2005, 2013), although the CDI Practitioner Toolkit is intended for use with any team engaged in CDI, not only those trained by Data Wise. Drawing on all of these sources, we identified eight essential constructs of CDI organized under three domains. Exhibit 1.1 shows the nesting of constructs within these domains and of indicators within the constructs.

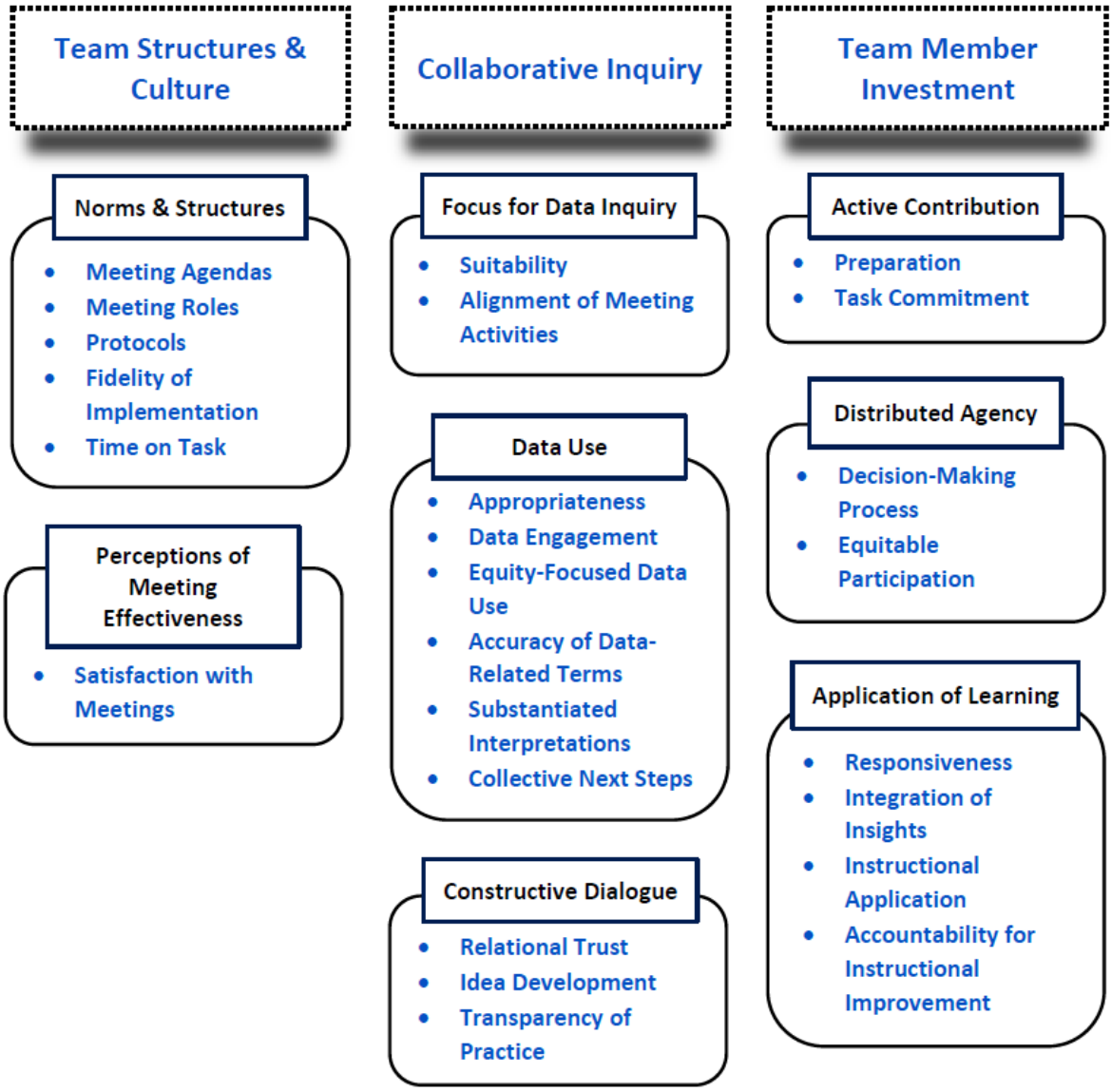
Domain 1, Team Structures and Culture, includes the frameworks and social norms that govern how members of a team function together. Team structures define the roles, responsibilities, and hierarchy within a team, determining how tasks are organized, coordinated, and executed (Ermeling, 2010; Kennedy et al., 2011; Nelson & Slavit, 2007, 2008). Team culture encompasses the shared values, beliefs, behaviors, and norms within the team, influencing how team members communicate and make decisions (Boudett et al., 2020; Nelson & Slavit, 2008; Schildkamp & Poortman, 2015). The domain of Team Structures and Culture has two key constructs: Norms and Structures, and Perceptions of Meeting Effectiveness.[2]

Domain 2, Collaborative Inquiry, captures team-based problem-solving centered on student data and instructional strategies (Cochran-Smith & Lytle, 2009). This process encompasses identifying shared challenges (Horn & Little, 2010); using multiple forms of data, including observations of teacher practice (Boudett et al., 2005, 2013; Mandinach & Gummer, 2016b), engaging in constructive critical dialogue, and implementing intellectually rich instructional changes aimed at addressing identified issues (Nelson et al., 2012). The domain of Collaborative Inquiry has three key constructs: Focus for Data Inquiry, Data Use, and Constructive Dialogue.

Domain 3, Team Member Investment, considers individual contributions of team members, distinct from collective actions. Engaged team members often bring new ideas, ask challenging questions that push thinking, and model reflective practices that can influence the team's culture and dynamics (Little, 2003; Nelson & Slavit, 2008; Nelson et al., 2012; Vescio et al., 2008). Additionally, engaged team members apply their learning from the collaborative session to their classroom practices (Jimerson & Wayman, 2015; Mandinach & Gummer, 2016a). The domain of Team Member Investment has three key constructs: Active Contribution, Distributed Agency, and Application of Learning.

[2] Perceptions of Meeting Effectiveness/Satisfaction with Meetings emerged as a construct/indicator during analysis of end-of-year survey data and was not identified a priori from the literature.

Exhibit 1.1. Collaborative Data Inquiry Constructs by Domain



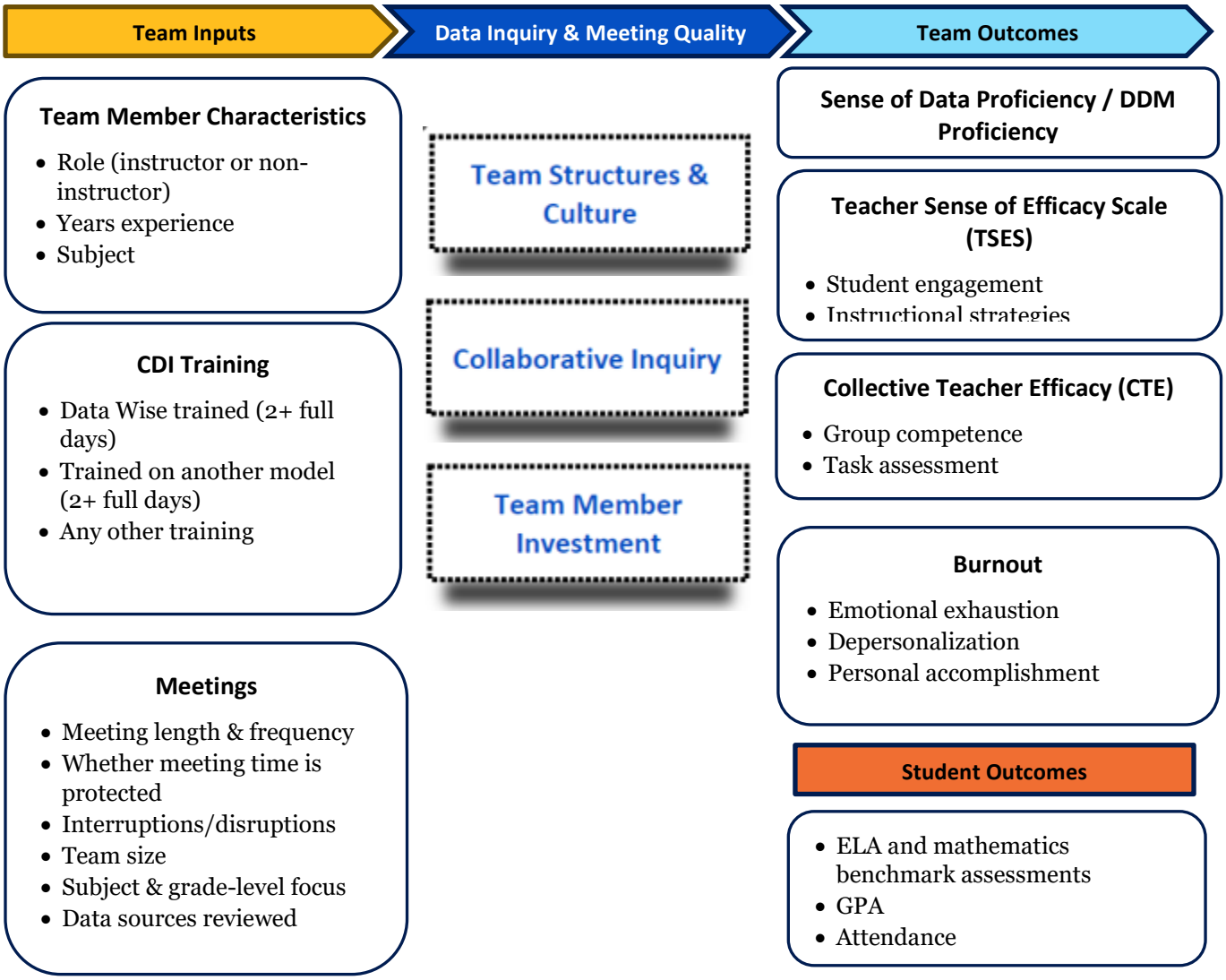
Although construct map shown in Exhibit 1.1 served as the blueprint for the design of all toolkit instruments, the observation rubric, exit ticket, and end-of-year survey diverge in their coverage of the indicators shown under each construct. The observation rubric includes measures of almost all of the indicators and constructs shown in Exhibit 1.1 (with three exceptions, noted below). The exit ticket and end-of-year survey, because they were designed to be low burden and completed quickly, cover just a subset of those indicators. For example, the Data Use construct is measured via six indicators on the observation rubric (Appropriateness, Data Engagement, etc.). The exit ticket has items aligned to two of those six indicators, and the end-of-year survey has items aligned to a slightly different subset of Data Use indicators. In addition, because team members themselves complete the exit ticket and end-of-year survey, those instruments include items that measure two indicators under

Application of Learning and one indicator under Perceptions of Meeting Effectiveness that would be difficult, if not impossible, for an external visitor to observe directly. This divergence in constructs and indicators measured has implications for some of the analyses described later.

CDI Theory of Action

Exhibit 1.2 shows the theorized inputs contributing to CDI meeting quality and the theorized outcomes from high-quality CDI meetings. Meeting quality is hypothesized to depend on team member characteristics such as their experience and role and their prior CDI training in Data Wise or other models, as well as on meeting characteristics such as length of meetings and data reviewed in meetings. High-quality data inquiry meetings are hypothesized to increase team members' data proficiency and their sense of self-efficacy as they gain more experience

Exhibit 1.2. CDI Theory of Action



working with data. This in turn can lead to increased teacher collective efficacy, which has been shown to be predicted by the quality of professional learning communities (Voelkel & Chrispeels, 2017), suggesting that high-quality CDI can improve teacher collective efficacy. Teacher collective efficacy also mediated levels of job stress related to student behavior (Klassen, 2010). Therefore, effective educator meeting practices can build collective self-efficacy, which promises to mitigate teacher stress (leading to lower burnout and turnover) and improve student outcomes, including student attendance and achievement as measured by grade point average (GPA) and English language arts (ELA) and math assessments.

Motivation for the Study

This report describes a series of analyses undertaken to establish the reliability and validity of the three components of the CDI Practitioner Toolkit, designed to help instructional leaders to assess how well teams are engaging in CDI. The study aimed to test whether measures of CDI meeting quality collected via observations and surveys would yield useful and consequential insights into how instructional teams were using their meeting time. It also set out to generate evidence that would help potential toolkit users assess the value of using all three toolkit components, as opposed to any one in isolation.

The study set out to address three research questions:

1. Does the CDI Practitioner Toolkit reliably assess CDI meeting quality?
2. To what extent does the CDI Practitioner Toolkit have internal consistency, that is, to what degree are construct-level measures correlated across toolkit components?
3. To what extent does the CDI Practitioner Toolkit demonstrate evidence of concurrent and predictive validity?

This study builds on early work collecting multiple types of validity evidence as defined by *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 2014). To collect evidence based on test content, a panel of academic and practitioner experts conducted a structured content review and established that the CDI Practitioner Toolkit is capturing critical aspects of the constructs being measured. To collect evidence based on response processes, the research team conducted cognitive interviews with 15 instructional leaders who used the toolkit to rate a meeting the leaders had recently observed, and conducted two rounds of pilot testing in two districts. These research activities provided evidence that users were responding as expected and in alignment with the intent of the measure (Brown et al., 2026).

The current study collected validation evidence based on the toolkit's internal structure and on the relationship of toolkit measures with other variables (AERA et al., 2014). To examine the internal structure of the toolkit components (observation rubric, exit ticket, and end-of-year survey), we conducted exploratory factor analyses and assessed internal consistency as the

degree to which measures of the same latent constructs are correlated across components. To collect evidence of the toolkit's relationships with instructionally relevant outcomes, we examined concurrent and predictive validity through the correlation of toolkit construct measures with established measures of team outcomes and with student outcomes and the degree to which toolkit measures predicted these outcomes in multivariate analyses.

2. Data Sources and Methods

Setting

Two mid-sized urban districts in the northeastern United States participated in the validation study. We limited eligibility for the study to schools serving one or more grades in K–8 so we could use standardized benchmark assessments of ELA and math as a study outcome. The first district included 27 regular schools serving one or more grades in K–8. The second district had 25 regular schools serving grades K–6 or above, and six serving grades 6 or 7 through 8. Both districts served predominantly Black and Latino students as well as high proportions of students from low-income families (Exhibit 2.1).

Exhibit 2.1. Demographic Characteristics of Study Districts, 2024–25

Characteristic	District 1	District 2
Schools serving study grades (K–8)	27	31
Student enrollment, grades K–8	12,226	14,594
Race/ethnicity (percent)		
African American	32%	47%
Hispanic	51%	36%
White	10%	10%
Asian, Native Hawaiian/Pacific Islander or Other	6%	7%
Free or reduced-price lunch (percent)	75%	82%

Source. Common Core of Data, 2024–25.

Data Sources

Primary data collection occurred in two rounds, beginning in early February 2025 and ending in late June 2025. In District 1, a group of district-level instructional leaders at the director and assistant superintendent levels conducted team observations in groups of 2–5. District 2 observers included two district-level instructional leaders, two school-based instructional coaches, and three principals.

We held a total of three training sessions with observers via Zoom that each lasted 1 hour. Prior to the training, observers received the complete CDI Practitioner Toolkit, including the user guide. During the training, the research team covered the goals of the toolkit, gave an overview of the toolkit elements, reviewed the structure and scoring guidelines for the observation rubric, and provided a step-by-step guide to the meeting observation process. Participants also had an

opportunity to practice scoring some of the indicators using example vignettes. After the training, the observers received a copy of their training recording and written steps for observations.

Observers attended an entire team meeting (typically 20–40 minutes in length) and rated the meeting immediately after it had concluded. They provided scores for the meetings they observed via a Qualtrics link. (In District 1, district leaders who observed in groups submitted scores individually.) In total, we obtained 39 sets of scores for 23 instructional teams in District 1 and 28 sets of scores for 16 instructional teams in District 2. Some teams were observed once because of conflicts with the observers' availability and the timing of the team meetings. Out of the 23 District 1 teams, seven had one observation and 16 had two. In District 2, four teams had one observation and 12 had two. Most of the analyses described in this report used the scores from the second observation (collected between April and June), where available, because observers gained facility in using the rubric with practice and because we expected instructional teams to develop over time. Where observations from the second round were not available, we used scores from the first round, to maximize the number of teams included in the analysis. We used the scores submitted during both rounds of observation by teams of observers in District 1 to test inter-rater reliability.

At the end of the observed meeting, observers also administered a team member exit ticket by adding a Qualtrics link to the meeting agenda or by sharing the link via QR code. Instructional team members completed the exit tickets during the last 5 minutes of each team meeting. As with observation scores, we used exit ticket responses from the second round of observations where possible and used responses from the first round when exit tickets were not administered during the second round.

From late April to late June, we also administered the end-of-year survey to members of instructional teams who were observed in either round. The end-of-year survey included measures of meeting quality, as described above, as well as measures of team member outcomes. Team members responded to the survey via a Qualtrics link. We received 56 responses (49% response rate) from District 1 and 62 responses (67% response rate) from District 2. Every instructional team had a response from at least one team member, and 35 teams had responses from at least two team members. When only some team members responded, those responses were assumed to be representative of all members of the team. We also collected administrative data on student outcomes,[3] including GPA, attendance, and ELA and math benchmark assessments.

[3] We additionally sought to collect team member retention as a team outcome. However, only one of the two districts was able to provide retention data, and there was not sufficient variation in retention across teams in that district to justify including those data in correlational or predictive analyses.

Predictors of Interest

Predictors of interest were drawn from the CDI Practitioner Toolkit components, that is, from the observation rubric, exit ticket, and end-of-year survey. For factor and reliability analyses, these were analyzed at the individual level. For correlational and predictive analyses, these were aggregated to the team level.

Observation Rubric. As described above, the observation rubric captures features of CDI across seven constructs: Norms and Structures, Focus for Data Inquiry, Data Use, Constructive Dialogue, Active Contribution, Distributed Agency, and Application of Learning. Observers score multiple indicators within each construct and then assign a construct-level rating on a 7-point scale using a holistic synthesis of indicator evidence and overall quality of practice. The validation analyses reported below used the construct-level scores from the observation rubric because these capture observers' synthesis of indicator evidence.

Exit Ticket. Exit ticket items ask team members to rate the meeting on a 7-point scale aligned to the CDI Practitioner Toolkit's Collaborative Inquiry constructs (Focus for Data Inquiry, Data Use, Constructive Dialogue) and to key features of team functioning and engagement (e.g., Norms and Structures, Active Contribution).

End-of-Year Survey. End-of-year survey items capture team members' retrospective views on team functioning over the school year. The survey includes items aligned to the CDI Practitioner Toolkit's eight CDI constructs and asks respondents about the frequency and consistency of CDI practices across the school year (e.g., whether teams routinely sustain a clear focus, whether data are used to diagnose and act on instructional problems, and whether team norms support equitable participation and shared responsibility). In addition to these predictors, the end-of-year survey also captures team inputs (such as frequency of meetings), team member characteristics, and team outcomes.

Outcomes of Interest

Team Member Outcomes. Team member outcomes were measured through the end-of-year survey and reflect the theorized benefits of high-quality CDI participation for educators. These outcomes include teacher self-efficacy, collective teacher efficacy, and burnout measured using established scales (Exhibit 2.2). In addition, the survey includes measures of team members' perceived capacity for data use (developed by the research team), which we used to create a self-reported data proficiency scale. Team outcomes are the average of these measures across team members with non-missing data.

Exhibit 2.2. Team Member Outcomes

Measure	Source	Analysis Notes
Self-Reported Data Proficiency	Original	One 11-item scale developed for the study and confirmed via factor analysis
Teacher Sense of Efficacy Scale (TSES)	Tschannen-Moran & Hoy, 2001	One 12-item scale (TSES) and three subscales with four items each: Efficacy in Student Engagement, Instructional Strategies, and Classroom Management
Collective Teacher Efficacy (CTE)	Goddard, 2002	A single 12-item scale and two subscales: Group Competence and Task Assessment
Burnout	Maslach et al., 2018	Three scales that cannot be combined: Emotional Exhaustion (nine items), Depersonalization (five items), and Personal Accomplishment (eight items)

Student Outcomes. Student outcomes were drawn from district administrative records and capture academic performance and engagement outcomes plausibly related to CDI-driven instructional improvement. These outcomes include ELA and math achievement as assessed in spring 2025 using state standardized tests, students' average daily attendance, and GPA. Student outcomes are linked to the instructional teams and aggregated to the team level.

Control Variables

Multivariate models controlled for a district indicator, team inputs, and beginning-of-year student achievement to account for the fact that CDI Practitioner Toolkit measures may be correlated with team and student outcomes because of preexisting differences across teams. For example, scales may be positively correlated with end-of-year student achievement because schools with higher student achievement at baseline are more likely to have higher achievement at the end of the year, and such high-performing schools may have more productive data team meetings because they have fewer urgent issues.

Team Inputs. Measures of team characteristics were drawn from the end-of-year survey and administrative data and were aggregated to the team level for use in analysis. These include:

- Team composition and characteristics, such as the share of team members who are primarily teachers, and average years' experience
- Prior Data Wise or other data training
- Meeting characteristics, including length and frequency of meetings

Student Achievement. To account for differences in students across districts and schools, we controlled for beginning-of-year ELA and math achievement on benchmark assessments. This measure was first standardized within district and grade to account for different assessment standards and procedures. Then, for each team, we used the average student achievement among students in any class where an instructional team member was the teacher of record.

The final set of control variables was determined using LASSO. Final OLS models for team outcomes included those controls that were identified by LASSO as predictive for at least two team outcomes, and likewise for student outcomes. Exhibit 2.3 shows the covariates selected by outcome and those included in final models for team outcomes. The final models for team outcomes controlled for meeting frequency, percentage of team members who were teachers, percentage receiving Data Wise training, percentage receiving other training, and average beginning-of-year ELA and math scores of students of team members.

Exhibit 2.3. Control Variables Selected for Team Outcomes

Control Variable	Teacher Efficacy	Collective Teacher Efficacy	Burnout: Emotional Exhaustion	Burnout: Personal Accomplishment	Burnout: Depersonalization	Self-Reported Data Proficiency	In Final Model
District 2 indicator							
Indicator for Round 1 observation					✓		
Meeting frequency		✓			✓		✓
Meeting time protected		✓					
Team size	✓						
Average years of experience							
Teacher (percent)		✓	✓		✓		✓
Math	✓						
ELA							
Science							
Social Studies	✓						
Other		✓					
Received Data Wise training		✓			✓		✓
Received other CDI training	✓					✓	✓
Beginning-of-year math		✓			✓		✓
Beginning-of-year ELA	✓	✓			✓	✓	✓

Note. Years of experience was only provided by one district. Values for that district were replaced with zero for the purpose of these analyses. For team members who were not teachers, subject indicators were set to zero.

Exhibit 2.4 shows the control variables selected for student outcomes. Models examining student outcomes controlled for meeting frequency and average beginning-of-year ELA and math achievement of students of team members.

Exhibit 2.4. Covariates Selected for Student Outcomes

Control Variable	Math Achievement	ELA Achievement	GPA	Attendance	In Final Model
District 2 indicator				✓	
Meeting frequency		✓		✓	✓
Meeting time protected					
Team size					
Average years of experience		✓			
Teacher		✓			
Math	✓				
ELA				✓	
Science					
Social Studies					
Other				✓	
Received Data Wise training				✓	
Received other CDI training					
Beginning-of-year math	✓	✓		✓	✓
Beginning-of-year ELA		✓		✓	✓

Note. Years of experience was only provide by one district. Values for that district were replaced with zero for the purpose of these analyses. For team members who were not teachers, subject indicators were set to zero.

Analytic Samples

The study included data from 23 teams in as many schools for District 1, and 17 teams across 12 schools in District 2 (Exhibit 2.5). The analytic samples for each analysis depended on the set of teams that were observed, team members who submitted an exit ticket, and team members who responded to the end-of-year survey. In District 1, 87% of districts were observed in the first round, and 83% in the second. In District 2, 88% were observed in the first round, and 87% in the second. End-of-year survey response rates were lower in District 1 (56%) than in District 2 (73%).

Exhibit 2.5. Validity Study Data Collection Samples

Participants and respondents	District 1	District 2
Schools	23	12
Teams	23	17
Team members	114	92
Observations		
First round	20	15
Section round	19	13
Exit tickets completed (response rate)		
First round	122 (100%)	66 (77%)
Section round	95 (100%)	37 (61%)
End-of-year surveys (response rate)	64 (56%)	67 (73%)

Note. Response rates for the first and second rounds of exit tickets were calculated with respect to teams included in the observation, because eam members were asked to complete an exit ticket only for observed meetings.

Differences in Samples by Analysis

Because of differences in data availability across the two participating districts, the analytic samples differ across analyses. Both districts provided student achievement data for ELA and math and data from each of the CDI Practitioner Toolkit components. District 2 was not able to provide information needed to calculate GPA; therefore, that district is excluded from analyses related to that outcome. Additionally, the set of student achievement scores associated with each team depended on the subjects taught by that team—for example, teams with no math teachers did not have any math scores associated with them.

Sample Descriptives

Exhibit 2.6 shows the team characteristics by district, as reported on the end-of-year survey. Team members in District 1 who responded to the survey were less likely to have teaching as a primary role (37% compared with 68% of survey respondents in District 2) and were more likely to be coaches (55% compared with 10% in District 2). In District 1 at the time of the study, one grade-level instructional team from each school was participating in Data Wise training. To maximize the benefit of this training, all principals and school-based coaches participated in the meetings of this instructional team. As a result, on at least some teams, there were as many coaches participating in meetings as teachers. In addition, survey response rates may have been lower among teachers, compared with coaches. Among team members who responded to the survey and were teachers, most in District 1 (57%) were math teachers. In District 2, teachers were more likely to teach multiple subjects, with many indicating they taught ELA, math, science, or social studies.

Exhibit 2.6 Average Team Member Characteristics, by District

Characteristic	District 1	District 2
Experience (years)	12.97	n/a
Primary professional role		
Teacher	37%	68%
Coach	55%	10%
Administrator	2%	3%
Other nonteaching role	7%	10%
Primary subjects taught		
Math	57%	63%
ELA	32%	71%
Science	22%	52%
Social Studies	22%	51%
Other	5%	12%
None of the above	24%	18%
Survey respondents	64	67

Note. District 2 was not able to provide experience. Percentages for subjects taught do not add to 100% because teachers could report teaching multiple subjects.

Source. End-of-year survey.

Team members in District 1 had more data training and more experience with Data Wise in particular (Exhibit 2.7). In District 1, 89% of teachers had more than two full days of data training, compared with 75% in District 2. In District 1, those receiving full training all received Data Wise training. In District 2, 75% received full training, with almost a third of those receiving some training other than Data Wise.

Exhibit 2.7. Prior Data Training

Training	District 1	District 2
Data training days over last 3 years		
1 full day or less	6%	20%
2 full days	4%	5%
More than 2 full days	89%	75%
Data training or professional programs		
2+ days Data Wise training	89%	52%
2+ days of other training	0%	23%

In both districts, almost no teams held CDI meetings less often than scheduled, and teams in District 1 met more often than those in District 2 (Exhibit 2.8). Across teams, an average of just 4% of team members in both districts reported meeting less than scheduled. Across teams in District 1, 50% of team members met three times a month or more, and 75% met at least twice a month. Across teams in District 2, 55% of team members reported meeting twice a month or less.

Exhibit 2.8. CDI Meeting Frequency, Planned and Actual

Frequency	District 1	District 2
Met less than scheduled	4%	4%
How often?		
More than 4 times a month	5%	7%
4 times a month	20%	22%
3 times a month	25%	11%
2 times a month	24%	17%
Once a month	12%	29%
Less than once a month	10%	9%

Analytic Approach

To evaluate the validity of the CDI Practitioner Toolkit, we used a three-staged analytic strategy. First, to address Research Question 1 (Does the CDI Practitioner Toolkit reliably assess CDI meeting quality?), we conducted exploratory factor analyses and examined the reliability of the resulting measures. These analyses assess whether measures for a given theoretical construct are more correlated with each other than with measures from other constructs. If indicators within a construct are strongly correlated with each other, and not with indicators from other constructs, this would suggest the construct-specific measures are capturing distinct latent characteristics. These analyses were all conducted at the individual respondent (team member or observer) level.

Second, to address Research Question 2 (To what extent does the CDI Practitioner Toolkit have internal consistency?), we explored internal consistency of the toolkit as the degree to which measures from the same constructs across toolkit components are correlated with each other. If, for example, Application of Learning as assessed by the observation rubric is more strongly correlated with Application of Learning as assessed by the end-of-year survey than it is to other construct-specific measures on the end-of-year survey, this would suggest that Application of Learning is a stable dimension of meeting quality that is captured by both the observation rubric and end-of-year survey. To account for the nesting of responses from individuals within teams, these analyses were conducted using data averaged to the team level.

Third, to address Research Question 3 (To what extent does the CDI Practitioner Toolkit demonstrate evidence of concurrent and predictive validity?), we examined the degree to which bivariate and multivariate analyses indicate relationships between toolkit components and team and student outcomes theorized to be positively affected by high-quality CDI participation. Although we can never observe the degree to which the toolkit components capture the true latent characteristics of meeting quality, a strong relationship between toolkit measures and outcomes hypothesized to be affected by high-quality CDI meetings would suggest the toolkit is measuring dimensions of meeting quality. To account for the nesting of responses from individuals within teams, we conducted these analyses using data averaged to the team level. These methods are described in more detail below.

Exploratory Factor Analyses and Scale Reliability

We conducted exploratory factor analyses and scale reliability analyses at the respondent level (team members for exit ticket and end-of-year survey items; observers for observation rubric measures), using all respondents with non-missing data for the items included in each analysis. For each of the CDI Practitioner Toolkit components (observation rubric, exit ticket, and end-of-year survey), we conducted exploratory factor analyses (reported in the Appendix). Exploratory factor analyses were supplemented with a systematic review by multiple members of the toolkit development and research team. Six reviewers classified each item in each toolkit component into one of the theoretical constructs, and reviewers discussed and reconciled differences in opinions.

For outcomes and covariates drawn from preexisting published scales, we calculated scale scores following published scoring guidance. We report internal consistency using Cronbach's alpha and McDonald's omega for each established scale, as complementary indicators of reliability. For item sets that were not part of preexisting published scales, after selecting a final factor structure based on results from the exploratory factor analysis, we created scale scores and evaluated internal consistency using Cronbach's alpha and McDonald's omega.

Consistency of Construct-Level Measures Across Toolkit Components

We examined internal consistency through pairwise correlations across construct-level measures drawn from different CDI Practitioner Toolkit components. We compared the correlation among items measuring the same construct across toolkit components—for example, Application of Learning on the observation rubric and exit ticket—to the correlation of items measuring different constructs and across toolkit components—for example, Application of Learning on the observation rubric compared with Data Use on the exit ticket—to assess the degree to which the toolkit components measure distinct and consistent dimensions of CDI participation.

Concurrent and Predictive Validity Analyses

Concurrent and predictive validity analyses examined the degree to which measures of CDI meeting quality from each of the CDI Practitioner Toolkit components are related to team and student outcomes theorized to be positively affected by CDI participation. We examined these relationships using both bivariate and multivariate models.

As an initial step, we estimated pairwise correlations between each toolkit construct and each teacher and student outcome. We used these correlations to provide a simple descriptive summary of associations between constructs measured by toolkit components and the outcomes.

For each team and student outcome, we then estimated a multivariate model. Because analytic sample sizes were relatively small—ranging from 18 to 39 teams across outcomes—we were not able to include all construct-level measures from each toolkit component and each of the controls in the same model. We addressed this in two ways. First, rather than including each distinct theoretical construct measure from the toolkit components in the model, we included the more parsimonious factors identified by the exploratory factor analysis (see Appendix). Using these factors in the analyses is preferred because highly correlated items inflate standard errors in regression models, and using the parsimonious factors conserves degrees of freedom while capturing distinct sources of variation in the toolkit components.

Second, we used LASSO as a principled variable-selection step to identify the set of control variables and toolkit factors to include in the multivariate model. LASSO models are similar to OLS regressions but include a penalty term for any nonzero coefficient. This forces the model to select only the most predictive covariates. For each team outcome, we then estimated a regression model including the set of toolkit factors identified by LASSO for any team outcome and the set of controls identified for at least two outcomes. Similarly, for each student outcome, we estimated a regression model including the set of toolkit factors identified for any student outcome and the control variables identified for at least two student outcomes. To assess the association between the toolkit components and each outcome, we tested the joint hypothesis of all coefficients on toolkit component factors being zero for each outcome.

We conducted the above process of LASSO selection and model estimation using factors from all toolkit components, and then again restricting the included factors to those from each specific toolkit component—for example, just the two exit ticket factors. Comparing results across these models allowed us to assess how each toolkit component contributes to the predictions.

3. Results

Factor Analyses and Scale Reliability

Contrary to theoretical expectations (see Exhibit 1.2), which posit there are several dimensions of CDI meeting quality, factor analyses indicated that each CDI Practitioner Toolkit component captured only one or two distinct dimensions of meeting quality. Exploratory factor analyses (see Appendix for factor loading) indicated that construct-level measures from the observation rubric were highly correlated with each other and loaded onto one factor (Exhibit 3.1). The 14 items on the exit ticket loaded onto two factors, with 12 items on one factor capturing team member assessment of meeting quality and two on the other factor assessing appropriateness of data use. The 29 items on the end-of-year survey loaded onto two factors, with 26 items on one factor capturing team members' retrospective assessment of meeting quality and three on the other capturing team members' satisfaction with meetings. Reliability for each of these measures was very high, with Cronbach's alpha and McDonald's omega being .82 or greater for all factors.

Exhibit 3.1. Summary Statistics of Newly Created Measures

Measure	Number of Items	Mean Response (*SD*)	Cronbach's Alpha	McDonald's Omega
Self-reported data proficiency	11	4.31 (0.54)	.92	.93
Observed meeting quality (all constructs)	7	5.17 (1.35)	.95	.95
Team member assessment of meeting quality (multiple constructs)	12	6.38 (0.65)	.93	.93
Data use, appropriateness	2	5.92 (1.20)	.86	.
Retrospective meeting quality (multiple constructs)	26	4.30 (0.66)	.96	.96
Satisfaction with meetings	3	4.03 (1.07)	.82	.82

Note. Because McDonald's omega cannot be calculated with fewer than three items, it is missing for exit factor 2.

All existing scales, which were used to measure team outcomes, exhibited high reliability (Exhibit 3.2). The lowest value of Cronbach's alpha was .74 for the depersonalization subscale of Burnout, which is well above the minimum reliability standard set by What Works Clearinghouse (WWC, 2022) of .60. All other scales had a Cronbach's alpha of .80 or higher. Similarly, McDonald's omega, which provides a more accurate measure of reliability for scales with five or more items, was high for all established scales, ranging from .78 to .94.

Exhibit 3.2. Summary Statistics of Team Member Outcomes

Measure	Number of Items	Response Options	Mean Response (*SD*)	Cronbach's Alpha	McDonald's Omega
Teacher Sense of Efficacy Scale	12	Five options from "nothing" to "a great deal"	4.22 (0.49)	.90	.89
Student engagement	4	Five options from "nothing" to "a great deal"	4.11 (0.56)	.80	.81
Instructional strategies	4	Five options from "nothing" to "a great deal"	4.39 (0.51)	.80	.80
Classroom management	4	Five options from "nothing" to "a great deal"	4.16 (0.67)	.89	.89
Collective Teacher Efficacy	12	Five options from "almost never true" to "almost always true"	3.52 (0.58)	.84	.84
Burnout					
Emotional exhaustion	8	Seven options from "never" to "every day"	3.27 (1.46)	.93	.94
Depersonalization	5	Seven options from "never" to "every day"	1.57 (0.84)	.74	.78
Personal accomplishment	8	Seven options from "never" to "every day"	5.96 (0.80)	.83	.84

Internal Consistency

To explore the internal consistency of the CDI Practitioner Toolkit components, we examined the correlation of scales across instruments. If the same constructs across different toolkit components are measuring the same dimensions of CDI participation, and if the constructs are measuring independent dimensions, we would expect the correlations within the same construct to be larger than the correlations across different constructs.

Exhibit 3.3 shows that the correlations across the toolkit components within constructs are generally similar in magnitude to the correlations across constructs. For example, the correlation in the Distributed Agency construct between the observation rubric and exit ticket is .10, which is the smallest observed correlation between Distributed Agency as measured on the exit ticket and all other constructs as measured on the observation rubric (.10–.42). For the exit ticket and observation rubric, one exception is Data Use: The correlation of that construct between the observation rubric and exit ticket (.55) is the largest of all pairwise correlations of constructs across these two toolkit components.

Exhibit 3.3. Correlations Between Toolkit Component Construct-Level Measures

Measure	Application of Learning	Constructive Dialogue	Data Use	Distributed Agency	Focus for Data Inquiry	Norms and Structures
End-of-year survey			**Observation rubric**			
Application of Learning	**.12**	.14	.23	-.03	.16	.12
Constructive Dialogue	.03	**.23**	.39	.10	.19	.03
Data Use	.14	.20	**.41**	.10	.24	.14
Distributed Agency	.10	.17	.42	**.10**	.30	.10
Focus for Data Inquiry	-.05	.09	.13	-.11	**.13**	-.05
Norms and Structures	.12	.14	.23	-.03	.16	**.12**
Exit ticket			**Observation rubric**			
Application of Learning	**.24**	.48	.38	.26	.31	n/a
Constructive Dialogue	.08	**.41**	.40	.26	.28	n/a
Data Use	.37	.42	**.55**	.37	.48	n/a
Distributed Agency	.09	.23	.39	**.18**	.24	n/a
Focus for Data Inquiry	.09	.31	.34	.14	**.26**	n/a
Exit ticket			**End-of-year survey**			
Application of Learning	**.45**	.47	.36	.43	.11	n/a
Constructive Dialogue	.23	**.32**	.29	.28	.03	n/a
Data Use	.12	.16	**.25**	.16	.06	n/a
Distributed Agency	.19	.36	.25	**.28**	.01	n/a
Focus for Data Inquiry	.20	.33	.27	.28	**-.03**	n/a

Note. The exit ticket and end-of-year survey did not include measures of the Norms and Structures construct.

These findings indicate that the scales may not be measuring the same dimensions of CDI participation across toolkit components. There are several potential explanations for differences across toolkit components:

- The exit ticket has fewer items for each construct, so it may provide noisier measures.
- End-of-year survey response rates were relatively low in District 1 (56%). It may be that survey respondents were not representative of all team members, leading to a disconnect between the end-of-year survey and both the observation rubric and exit tickets.

- The end-of-year survey asks respondents to reflect on all CDI meetings held throughout the year, whereas the observation rubric and exit ticket capture the quality of a single meeting.
- Team members and observers may have different perspectives of meeting quality, leading to a disconnect between the observation rubric and both the exit ticket and end-of-year survey. This suggests that district leaders should be concerned if either the observation rubric or the exit ticket indicate CDI meeting quality issues.

Concurrent and Predictive Validity

We assessed concurrent and predictive validity using bivariate analyses, which do not statistically adjust for any baseline measures, and multivariate analyses, which account for differences in student and team characteristics at baseline.

Bivariate Analyses

We first present concurrent and predictive validity results from bivariate analyses exploring the correlation between CDI Practitioner Toolkit components and team and student outcomes. Correlations between constructs from the observation rubric and team outcomes were generally moderate (Exhibit 3.4). Higher values on the Emotional Exhaustion and Depersonalization dimensions of Burnout indicate more burnout, while higher values on the Personal Accomplishment scale indicate less burnout. The negative correlations between toolkit measures and both Emotional Exhaustion and Depersonalization, and the positive correlations with Personal Accomplishment, indicate that higher values on the toolkit measures are correlated with less burnout. Some observation constructs were weakly correlated with team-level outcomes. For example, the correlation between Application of Learning (the extent to which individual team members apply insights from the CDI team meeting to their classroom practice) and four team-member outcomes was .19 or less. About half of the correlations were .29 or greater in magnitude, with the strongest correlation (.54) being between Focus for Data Inquiry and the Emotional Exhaustion subscale of Burnout, which is a measure of feeling successful in one's work.

Exhibit 3.4. Correlations Between Observation Rubric Constructs and Team Outcomes

Measure	Self-Reported Data Proficiency	Teacher Sense of Efficacy Scale	Collective Teacher Efficacy	Burnout: Emotional Exhaustion	Burnout: Depersonalization	Burnout: Personal Accomplishment
Active Contribution	.29	.39	.20	-.43	-.25	.33
Application of Learning	.09	.33	.15	-.19	.15	.35
Constructive Dialogue	.11	.27	.29	-.37	-.30	.34

Measure	Self-Reported Data Proficiency	Teacher Sense of Efficacy Scale	Collective Teacher Efficacy	Burnout: Emotional Exhaustion	Burnout: Depersonalization	Burnout: Personal Accomplishment
Data Use	.35	.46	.35	-.43	-.24	.42
Distributed Agency	.20	.29	.32	-.40	-.29	.26
Focus for Data Inquiry	.14	.30	.41	-.54	-.34	.52
Norms and Structures	.19	.27	-.07	-.25	-.07	.09

Similarly, correlations between exit ticket scales and team outcomes were generally moderate (Exhibit 3.5). Most correlations were .22 or greater in absolute value. The correlation between Data Use and the Teacher Sense of Efficacy Scale was particularly strong at .47; teams focusing on data use had higher self-efficacy. Two correlations were weak at less than .10: the correlation between Distributed Agency and Collective Teacher Efficacy (-.04), and the correlation between Constructive Dialogue and Collective Teacher Efficacy (.09).

Exhibit 3.5. Correlations Between Exit Ticket Construct Measures and Team-Level Outcomes

Measure	Self-Reported Data Proficiency	Teacher Sense of Efficacy Scale	Collective Teacher Efficacy	Burnout: Emotional Exhaustion	Burnout: Depersonalization	Burnout: Personal Accomplishment
Application of Learning	0.22	0.31	0.16	-0.31	-0.28	0.14
Constructive Dialogue	0.19	0.18	0.09	-0.34	-0.25	0.20
Data Use	0.11	0.47	0.15	-0.38	-0.27	0.30
Distributed Agency	0.20	0.11	-0.04	-0.29	-0.22	0.09
Focus for Data Inquiry	0.25	0.22	0.13	-0.30	-0.28	0.10

The correlations between end-of-year survey constructs and team outcomes tended to be stronger than those for the observation rubric or exit ticket (Exhibit 3.6). This may be because team outcomes were assessed on the end-of-year survey, so the end-of-year survey is better aligned with team outcomes in terms of timing and respondents. Most correlations were greater than .31 in absolute value. The correlations between both Data Use and Constructive Dialogue and self-reported data proficiency were particularly strong, exceeding .60 in both cases. Teacher Sense of Efficacy Scale was relatively weakly correlated with Focus for Data Inquiry, suggesting that CDI meetings with strong focus for data inquiry may not have an effect on teachers' sense of self-efficacy.

Exhibit 3.6. Correlations Between End-of-Year Survey Constructs and Team-Level Outcomes

Measure	Self-Reported Data Proficiency	Teacher Sense of Efficacy Scale	Collective Teacher Efficacy	Burnout: Emotional Exhaustion	Burnout: Depersonalization	Burnout: Personal Accomplishment
Active Contribution	.53	.56	.46	-.41	-.23	.21
Application of Learning	.46	.37	.49	-.26	-.19	.13
Constructive Dialogue	.63	.29	.41	-.36	-.27	.10
Data Use	.66	.46	.53	-.44	-.25	.20
Distributed Agency	.61	.32	.55	-.42	-.26	.19
Focus for Data Inquiry	.27	.07	.35	-.22	-.13	.16
Norms and Structures	.27	.15	.54	-.38	-.31	.16
Satisfaction with Meetings	-.15	-.20	.08	-.17	-.35	-.04

Correlations between observation rubric constructs and student achievement were relatively weak (Exhibit 3.7). Half of the correlations were less than .13 in absolute value, with several being less than .10 in absolute value. Some construct-level measures were moderately correlated with student outcomes. Norms and Structures was moderately correlated with both attendance and GPA (.24 and .27, respectively). Application of Learning was moderately correlated with GPA (.29). Active Contribution was moderately correlated with math achievement (.25), and both Focus for Data Inquiry and Data Use were moderately correlated with ELA achievement (.24 and .27, respectively).

Exhibit 3.7. Correlations Between Observation Rubric Constructs and Student Outcomes

Measure	Attendance	GPA	Math Achievement	ELA Achievement
Active Contribution	.15	.13	.25	.18
Application of Learning	.00	.29	.06	-.02
Constructive Dialogue	.05	-.16	.04	.08
Data Use	-.01	-.03	.12	.27
Distributed Agency	.00	-.23	.14	.20
Focus for Data Inquiry	.00	.01	.18	.24
Norms and Structures	.24	.27	.11	.01

Similarly, exit ticket constructs were generally weakly correlated with student outcomes, with some moderate correlations (Exhibit 3.8). Half of all correlations were less than .16 in absolute value. While Focus for Data Inquiry as measured by the observation rubric had no correlation with attendance or GPA, Focus for Data Inquiry as assessed on the exit ticket was moderately correlated with attendance and GPA (.31 and .43, respectively).

Exhibit 3.8. Correlations Between Exit Ticket Constructs and Student Outcomes

Measure	Attendance	GPA	Math Achievement	ELA Achievement
Application of Learning	.18	.11	.22	.28
Constructive Dialogue	.10	-.09	.07	.24
Data Use	.03	-.20	.06	.09
Distributed Agency	-.01	.15	.02	.19
Focus for Data Inquiry	.31	.43	.16	.30

Likewise, end-of-year survey constructs were generally weakly correlated with student outcomes, with some moderate to large correlations (Exhibit 3.9). Attendance was generally weakly correlated with end-of-year survey constructs, with the exception of Norms and Structures (-.17). GPA was moderately correlated with Constructive Dialogue, Application of Learning, and Distributed Agency (.19, .25, and .26, respectively). Math and ELA achievement were generally more strongly correlated with exit ticket constructs than attendance and GPA. The correlations between math achievement and Application of Learning, Data Use, and Distributed Learning were relatively strong (.45, .42, and .46, respectively). ELA achievement was moderately correlated with Application of Learning, Constructive Dialogue, and Data Use (.35, .30, and .38, respectively).

Exhibit 3.9. Correlations Between End-of-Year Survey Constructs and Student Outcomes

Measure	Attendance	GPA	Math Achievement	ELA Achievement
Active Contribution	-.05	.16	.22	.14
Application of Learning	.06	.25	.45	.35
Constructive Dialogue	.04	.19	.35	.30
Data Use	.11	.08	.42	.38
Distributed Agency	.05	.26	.46	.38
Focus for Data Inquiry	-.11	.13	-.12	-.08
Norms and Structures	-.17	.12	.18	.14
Satisfaction with Meetings	-.12	-.13	-.22	-.20

Multivariate Analyses

The above bivariate analyses provide some evidence for the validity of the CDI Practitioner Toolkit components, but these analyses may not capture causal relationships because of confounding factors or reverse causation. For example, it is possible that teams who exhibited less burnout throughout the year had more productive CDI team meetings because they were less burned out, leading to a spurious correlation between burnout and measures of CDI team meeting quality. Multivariate analyses accounted for possible confounding factors by controlling for baseline student and team characteristics, such as meeting frequency and average ELA and math achievement of students of team members.

Because the exploratory factor analyses reported above identified a parsimonious set of factors from each toolkit component, the multivariate analyses described in this section used those factors, rather than measures of the theoretical constructs reported above.

For all team member outcomes, all factors from each toolkit component were selected as predictive for at least one outcome by the LASSO procedure (Exhibit 3.10). As described above, we used LASSO to address the challenge that there were few teams included in the analysis relative to the number of potential control variables, and we included in final OLS regressions all factors identified for at least one teacher outcome. Even with a rich set of control variables, LASSO consistently selected toolkit components for team member outcomes, suggesting they are predictive of those outcomes.

For student outcomes, only factors from the end-of-year survey were selected as predictive (see Exhibit 3.10), with end-of-year factor 1 selected as predictive of math, and end-of-year factor 2 selected as predictive of attendance. This may be because it takes more than a year for the relationship between quality of data meeting and student outcomes to emerge.

Exhibit 3.10. Concurrent and Predictive Validity of the CDI Practitioner Toolkit, Team Outcomes

Outcome	Observed Meeting Quality (All Constructs)	Team Member Assessment of Meeting Quality	Data Use, Appropriateness	Retrospective Meeting Quality (Multiple Constructs)	Team Member Satisfaction with Meetings
Any team outcomes	✓	✓	✓	✓	✓
Self-reported data proficiency				✓	✓
Teacher self-efficacy			✓	✓	✓
Collective self-efficacy	✓	✓	✓	✓	
Burnout: Emotional Exhaustion	✓			✓	

Outcome	Observed Meeting Quality (All Constructs)	Team Member Assessment of Meeting Quality	Data Use, Appropriateness	Retrospective Meeting Quality (Multiple Constructs)	Team Member Satisfaction with Meetings
Burnout: Depersonalization			✓		
Burnout: Personal Accomplishment	✓				✓
Any student outcomes				✓	✓
GPA					
Attendance					✓
Math achievement				✓	
ELA achievement					

Note. Blank cells indicate the CDI Practitioner Toolkit components were not identified by LASSO as predictive. The measures reported in the first three columns (observed meeting quality; team member assessment of meeting quality; and data use, appropriateness) offer evidence of predictive validity. The end-of-year measures reported in the final two columns offer evidence of concurrent validity because both the team outcomes and the end-of-year factors were collected in the same survey administration.

✓= Toolkit components were identified by LASSO as predictive of the outcome.

To further assess the strength of the associations between toolkit components and team member outcomes, we estimated OLS regressions using the LASSO-identified toolkit factors and control variables. We then tested the joint null hypothesis that all included toolkit factors have no association with the outcome. We repeated the LASSO selection exercise and OLS regression modeling using all toolkit factors as well as restricting the factors to those from just one toolkit component. (See Appendix for the factors selected by LASSO when only one toolkit component was included.)

When all toolkit components were included in this regression (see Appendix for summary results for models using select toolkit components), we find found a statistically significant relationship between toolkit components and self-reported data proficiency, teacher sense of self-efficacy, and burnout from depersonalization (Exhibit 3.11). Among individual toolkit component factors, several had a statistically significant relationship with the outcome, with the coefficient having the expected sign. Specifically, increases in retrospective meeting quality were positively and significantly associated with higher self-reported data proficiency; higher date use was positively and significantly associated with teacher self-efficacy, and team member satisfaction with meetings was negatively and significantly associated with the depersonalization dimension of burnout (for which higher scores indicate more burnout).

Exhibit 3.11. Regression Results for Team Outcomes

Outcome	Self-Reported Data Proficiency	Teacher Self-Efficacy	Collective Self-Efficacy	Burnout: Emotional Exhaustion	Burnout: Depersonalization	Burnout: Personal Accomplishment
Toolkit measures	†	†		†		
Observed meeting quality (all constructs)	0.03 (0.05)	0.00 (0.05)	0.04 (0.06)	-0.19 (0.17)	0.03 (0.06)	0.13 (0.1)
Team member assessment of meeting quality	-0.12 (0.13)	-0.01 (0.15)	-0.25 (0.16)	0.02 (0.45)	-0.02 (0.17)	0.08 (0.28)
Data use, appropriateness	-0.03 (0.07)	0.22** (0.08)	0.05 (0.09)	-0.02 (0.24)	-0.05 (0.09)	0.21 (0.15)
Team member satisfaction with meetings	-0.03 (0.06)	-0.07 (0.07)	0.07 (0.08)	-0.11 (0.22)	-0.25*** (0.08)	-0.15 (0.14)
Retrospective meeting quality (multiple constructs)	0.34** (0.15)	0.19 (0.18)	0.41** (0.19)	-0.16 (0.55)	0.23 (0.2)	-0.11 (0.34)
First observation indicator	-0.13 (0.16)	-0.06 (0.19)	0.11 (0.2)	0.35 (0.56)	0.27 (0.21)	0.36 (0.36)
Meeting frequency	-0.05 (0.06)	-0.02 (0.07)	0.13 (0.08)	-0.05 (0.22)	-0.06 (0.08)	0.32** (0.14)
Teacher (percent)	0.01 (0.2)	0.24 (0.26)	-0.43 (0.25)	1.81** (0.71)	0.9*** (0.26)	-0.28 (0.45)
Received Data Wise training	-0.11 (0.25)	-0.10 (0.29)	-0.50 (0.32)	0.23 (0.9)	0.63* (0.33)	-0.04 (0.56)
Received other CDI training	-0.87 (0.52)	-1.44** (0.59)	0.47 (0.66)	-0.87 (1.85)	0.75 (0.69)	0.62 (1.16)
Team size	-0.04 (0.03)	-0.02 (0.04)	0.00 (0.04)	-0.14 (0.12)	0.01 (0.05)	-0.04 (0.08)
Beginning-of-year ELA	0.45* (0.23)	0.31 (0.26)	0.17 (0.29)	-0.84 (0.81)	-0.44 (0.3)	-0.67 (0.51)
Beginning-of-year math	-0.17 (0.18)	-0.16 (0.21)	0.37 (0.23)	-0.13 (0.65)	-0.08 (0.24)	0.41 (0.41)
Number of teams	37	35	37	37	37	37
R^2	0.60	0.61	0.66	0.57	0.70	0.38
MSE	0.07	0.09	0.11	0.87	0.12	0.34
AIC	16.53	23.63	33.52	110.08	36.82	75.78

Note. The toolkit measures reported in the first three rows (observed meeting quality; team member assessment of meeting quality; and data use, appropriateness) offer evidence of predictive validity. The end-of-year measures reported in next two rows offer evidence of concurrent validity because both the team outcomes and the end-of-year factors were collected in the same survey administration.

† indicates rejection of null hypothesis that all toolkit components are jointly insignificant at the 5% level.

*** $p < .01$; ** $p < .05$; * $p < 0.10$.

By contrast, for student outcomes, only end-of-year survey items were selected as predictive, and there was no statistically significant relationship between selected toolkit factors and each

outcome (Exhibit 3.12). Coefficients on toolkit component factors tended to be small; for example, for attendance, the coefficients on retrospective meeting quality and team member satisfaction with meetings are both about -0.01.

Exhibit 3.12. Regression Results for Student Outcomes

Outcome	GPA	Attendance	Math Achievement	ELA Achievement
Toolkit measures				
Retrospective meeting quality (multiple constructs)	0.06 (0.39)	-0.01 (0.03)	0.03 (0.07)	0.00 (0.05)
Team member satisfaction with meetings	-0.08 (0.13)	-0.01 (0.01)	0.00 (0.03)	-0.02 (0.02)
Meeting frequency	0.00 (0.13)	0.02 (0.01)	0.02 (0.03)	0.07*** (0.02)
Teacher (percent)	-0.06 (0.52)	-0.12*** (0.04)	-0.15 (0.09)	-0.09 (0.07)
Beginning of year ELA	-0.71 (0.71)	0.05 (0.05)	-0.02 (0.12)	0.68*** (0.09)
Beginning of year math	0.80 (0.69)	0.00 (0.04)	0.83*** (0.1)	0.07 (0.08)
Number of teams	17	37	37	37
R^2	0.21	0.40	0.83	0.82
MSE	0.11	0.00	0.02	0.01
AIC	15.89	-87.36	-28.69	-46.90

Note. For each outcome, we could not reject the null hypothesis that all toolkit components are jointly insignificant at the 5% level.
*** $p < .01$; ** $p < .05$; * $p < .10$.

4. Discussion

Together, these findings indicate that the CDI Practitioner Toolkit reliably measures dimensions of high-quality CDI meetings. Although the toolkit distinguishes theoretically distinct constructs, such as Data Use and Application of Learning, exploratory factor analyses suggest that each component captures one or two unique dimensions. This finding is expected because of differences in design of the toolkit components—for example, the observation rubric and exit ticket assess a single CDI meeting, whereas the end-of-year survey asks respondents to reflect on all meetings held throughout the school year. Nonetheless, this finding suggests that future research should explore whether shorter versions of each instrument would be as valid.

Bivariate analyses indicate moderate correlations with some theoretical constructs and teacher and student outcomes theorized to be positively associated with high-quality CDI meetings. Correlations were generally stronger for team member outcomes. This finding suggests that the CDI Practitioner Toolkit is measuring important dimensions of CDI meeting quality but that student outcomes may be too distal to be strongly related with meeting quality.

Multivariate analyses indicate that each toolkit component contributes unique explanatory power to predicting team member outcomes. Only the end-of-year survey was selected as predictive of student outcomes conditional on baseline characteristics, including beginning-of-year achievement, although end-of-year survey factors were not statistically significantly associated with student outcomes. This finding may be due to it taking more than a year of CDI meetings before the quality of those meetings is meaningfully associated with student achievement.

Limitations

Several factors limit the degree to which we can fully assess the validity of the CDI Practitioner Toolkit.

Sample Size. The study included 39 teams across two districts. Multivariate analyses at the team level were therefore limited in the number of control variables that could be included in each model. Similarly, analyses were limited in their ability to detect statistically significant relationships.

Outcomes. Although the study included several outcomes expected to be positively associated with high-quality CDI participation, the study did not have access to more proximal outcomes, such as the degree to which teachers adapted their practices in classrooms. Associations with more distal outcomes, such as GPA and student achievement, may take more than a year to manifest.

Aggregation Bias. Aggregating data from lower to higher units of analysis (e.g., from student to schools) can make it harder to detect associations between variables because the aggregation reduces variation. Because the identifying variation was at the team level, internal and predictive validity analyses used team-level data, limiting the ability to detect associations.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Bernhardt, V. L. (2009). Data use: Data-driven decision making takes a big-picture view of the needs of teachers and students. *Journal of Staff Development*, *1*(30), 24–27.

Boudett, K. P., & Bocala, C. (2025). *Data Wise, Third Edition: Educators collaborating so each learner thrives*. Harvard Education Press.

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*. Harvard Education Press.

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

Boudett, K. P., City, E. A., & Murnane, R. J. (Eds.). (2020). *Data wise, revised and expanded edition: A step-by-step guide to using assessment results to improve teaching and learning*. Harvard Education Press.

Brown, L., Bocala, C., Laguarda, K., Burkander, P., Capan, S., & Schellenberger, K. (2026, April 8–12). *Developing and validating a practitioner toolkit to support collaborative data inquiry* [Paper presentation]. Annual Meeting of the American Educational Research Association, Los Angeles, CA, United States.

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

Cochran-Smith, M., & Lytle, S. L. (2009). *Inquiry as stance: Practitioner research for the next generation*. Teachers College Press.

Cosner, S. (2011). Teacher learning, instructional considerations and principal communication: Lessons from a longitudinal study of collaborative data use by teachers. *Educational Management Administration & Leadership*, *39*(5), 568–589.

Datnow, A., & Hubbard, L. (2015). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change*, *16*(1), 7–28.

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

DuFour, R., DuFour, R., Eaker, R., & Many, T. (2010). *Learning by doing: A handbook for professional learning communities at work* (2nd ed.). Solution Tree Press.

Ermeling, B. A. (2010). Tracing the effects of teacher inquiry on classroom practice. *Teaching and Teacher Education*, *26*(3), 377–388.

Goddard, R. (2002.) A theoretical and empirical analysis of the measurement of collective efficacy: the development of a short form. *Educational and Psychological Measurement*, *62*(1), 97-110.

Goddard, Y. L., Goddard, R. D., & Tschannen-Moran, M. (2007). A theoretical and empirical investigation of teacher collaboration for school improvement and student achievement in public elementary schools. *Teachers College Record, 109*(4), 877–896.

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. *Yearbook of the National Society for the Study of Education*, *106*(1), 105–131.

Jimerson, J. B., & Wayman, J. C. (2015). Professional learning for using data: Examining teacher needs & supports. *Teachers College Record*, *117*(4), 1–36.

Kennedy, A., Deuel, A., Nelson, T. H., & Slavit, D. (2011). Requiring collaboration or distributing leadership? *Phi Delta Kappan*, *92*(8), 20–24.

Kerr, K. A., 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*(4), 496–520.

Klassen, R. M. (2010). Teacher stress: The mediating role of collective efficacy beliefs. *The Journal of educational research*, *103*(5), 342-350.

Little, J. W. (2003). Inside teacher community: Representations of classroom practice. *Teachers College Record*, *105*(6), 913–945.

Love, N. (2009). *Using data to improve learning for all: A collaborative inquiry approach*. Corwin Press.

Mandinach, E. B. (2025). *Data ethics: How educators can use data effectively and responsibly*. Bloomsbury Publishing.

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

Mandinach, E. B., & Gummer, E. S. (2016b). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, *60*, 366–376.

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*. RAND Corporation.

Maslach, C., Jackson, S. E., & Leiter, M. P. (2016). *Maslach Burnout Inventory: Manual 4th edition*. Mind Garden.

Nelson, T. H., & Slavit, D. (2007). Collaborative inquiry among science and mathematics teachers in the USA: Professional learning experiences through cross-grade, cross-discipline dialogue. *Professional Development in Education*, *33*(1), 23–39.

Nelson T. H., & Slavit D. (2008). Supported teacher collaborative inquiry. *Teacher Education Quarterly*, *35*(1), 99–116.

Nelson, T. H., Slavit, D., & Deuel, A. (2012). Two dimensions of an inquiry stance toward student-learning data. *Teachers College Record*, *114*(8), 1–42.

Ronfeldt, M., Farmer, S. O., McQueen, K., & Grissom, J. A. (2015). Teacher Collaboration in Instructional Teams and Student Achievement. *American Educational Research Journal*, *52*(3), 475–514. https://doi.org/10.3102/0002831215585562

Schildkamp, K., Handelzalts, A., Poortman, C. L., Leusink, H., Meerdink, M., Smit, M., Ebbeler, J., & Hubers, M. D. (2018). *The Data Team Procedure: A systematic approach to school improvement*. Springer.

Schildkamp, K., & Poortman, C. (2015). Factors influencing the functioning of data teams. *Teachers College Record*, *117*(4), 1–42.

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing School Failure: Alternative Education for Children and Youth*, *52*(2), 48–58.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*(7), 783–805.

Vescio, V., Ross, D., & Adams, A. (2008). A review of research on the impact of professional learning communities on teaching practice and student learning. *Teaching and Teacher Education*, *24*(1), 80–91.

Voelkel, R. H., Jr. & Chrispeels, J. H. (2017). Understanding the link between professional learning communities and teacher collective efficacy. School Effectiveness and School Improvement, 28(4), 505-526.

What Works Clearinghouse. (2022). *Procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Appendix. Supplemental Results

Factor Analyses

Exhibit A-1 shows results from the exploratory factor analysis (EFA) of the observation rubric constructs. All constructs from the observation rubric were highly correlated with each other, and the EFA identified only one factor with an eigenvalue greater than 1.

Exhibit A-1. Exploratory Factor Analysis Results, Observation Rubric

Measure	Factor 1
Active Contribution	0.91
Application of Learning	0.90
Constructive Dialogue	0.89
Data Use	0.89
Distributed Agency	0.85
Focus for Data Inquiry	0.94
Norms and Structures	0.86

Exhibit A-2 shows results from the EFA of the exit ticket constructs. The EFA identified two factors with an eigenvalue greater than 1.

Exhibit A-2. Exploratory Factor Analysis Results, Exit Ticket

Measure	Factor 1	Factor 2
I learn things from the data and feedback shared during team meetings.	0.89	0.05
I have a voice in my team's decision-making process.	0.85	-0.05
I feel comfortable expressing my views in team meetings, even when the topic is challenging.	0.82	0.02
I believe that as a group, this team is improving teaching and learning for students.	0.81	0.07
I feel that everyone on our team is working toward the same goal.	0.75	0.02
I have a clear understanding of the next steps I will take after today's meeting.	0.69	0.19
I learn things during team meetings that I would not have thought about on my own.	0.68	0.05
I have made changes to my professional practice as a result of the discussions I have had with this team.	0.65	0.08
Our team has agreed on the next steps we will take based on our review, discussion, or analysis of data today.	0.63	0.30

Measure	Factor 1	Factor 2
I plan to make changes in my professional practice as a result of today's discussion.	0.48	0.22
In today's meeting, our team had access to data that supported productive discussion about important instructional issues	0.35	0.75
Today we looked at data which gave me useful insights into my students' learnings and/or into my own instructional practices.	0.32	0.82

Exhibit A-3 shows results from the EFA of the end-of-year survey constructs. The EFA identified four factors with Eigenvalues greater than 1.

Exhibit A-3. Exploratory Factor Analysis Results, End-of-Year Survey

Measure	Factor 1	Factor 2	Factor 3	Factor 4
The data we discussed gave me useful insights into my students' learning needs.	0.89	-0.03	-0.12	-0.13
The data we discussed gave me useful insights into my teaching practice.	0.87	-0.05	-0.04	0.01
Our discussions helped us develop a shared understanding of what we need to do	0.85	0.09	0.13	0.08
We used data to better understand the strengths and assets of our students.	0.86	0.04	0.01	-0.13
There was a meaningful connection to what we discussed at our last meeting.	0.82	0.00	-0.02	-0.03
Our objectives were important for us to discuss together.	0.81	0.13	-0.05	0.26
Everyone in the meetings contributed to meeting our objectives.	0.81	0.18	0.14	-0.06
We thoroughly understood the data we reviewed before discussing what actions to take	0.81	-0.06	-0.12	-0.04
I left the meeting knowing what my next steps were.	0.81	0.01	-0.06	-0.12
We reviewed data to help us achieve our objectives.	0.79	-0.08	-0.14	-0.07
We reviewed our progress from our last meeting.	0.79	-0.01	-0.09	-0.18
We respect one another's professional competence.	0.75	0.07	0.55	0.01
Meeting with the team was a good use of my time.	0.75	0.17	0.02	0.33
My relationships with my team members are open and honest.	0.75	0.10	0.41	-0.09
Our meetings had clear expectations or norms.	0.75	0.17	-0.03	0.12
We made factual observations about the data before we interpreted the data.	0.73	-0.09	-0.08	-0.19
We provide strong social support for one another.	0.71	-0.03	0.58	0.05
I understood what the data we discussed measured.	0.71	-0.04	-0.15	-0.22
I did something different in my classroom because of our meeting.	0.70	-0.23	-0.15	0.16

Measure	Factor 1	Factor 2	Factor 3	Factor 4
We had all the resources/knowledge needed to make progress on our objectives.	0.70	-0.01	-0.11	0.44
We have a great deal of cooperative effort among staff members.	0.69	0.05	0.52	-0.04
I was fully engaged in the discussions.	0.66	0.05	0.04	-0.16
I was clear on our meeting objectives at the start of the meeting.	0.64	0.07	-0.03	0.36
Most team members arrived well-prepared.	0.62	0.14	0.08	0.31
I learned things that I would not have thought about on my own.	0.62	-0.17	-0.16	-0.07
We had access to knowledgeable instructional leaders/high quality instructional materials that we needed	0.61	0.06	0.07	0.43
We reviewed disaggregated data of different student subgroups.	0.54	-0.35	-0.20	-0.04
We discussed the biases we may have with interpreting data.	0.52	-0.25	-0.04	-0.16
We had enough time to meet our objectives.	0.51	0.13	0.03	0.41
When we disagreed, we did so respectfully.	0.48	0.09	0.18	0.05
We should have discussed more important topics instead of what we talked about. (reverse-coded)	0.33	0.35	0.02	-0.21
I left the meeting without learning something new. (reverse-coded)	0.24	0.80	0.07	0.05
I left feeling frustrated with our lack of progress. (reverse-coded)	0.21	0.70	-0.03	0.07
I did not understand the connection between the data we discussed and our meeting objectives. (reverse coded)	0.08	0.71	-0.04	-0.05

Multivariate Analyses

Below we report the R^2 from these regressions, which measures the share of variation in outcomes explained by the control variables and toolkit components, to assess the relative contribution to predictive power from the toolkit components. For all team outcomes, the explanatory power of the CDI Practitioner Toolkit is highest when all toolkit components are included (Exhibit A-4).

Exhibit A-4. Statistical Significance of Toolkit Components and Model Performance, Team Member Outcomes (R^2)

Model	Self-Reported Data Proficiency	Teacher Sense of Efficacy Scale	Collective Teacher Efficacy	Burnout: Emotional Exhaustion	Burnout: Depersonalization	Burnout: Personal Accomplishment
All measures	0.57*	0.60*	0.63	0.55	0.70*	0.37
Observation rubric only	0.37	0.29	0.32	0.52*	0.50	0.17
Exit ticket only	0.45	0.54*	0.43	0.48	0.51	0.24
End-of-year survey only	0.53*	0.33	0.43*	0.46	0.60*	0.07

* Indicates that the joint test of significance across toolkit components was significiant at the 5% level.

By contrast, for student outcomes, there was no statistically significant relationship between selected toolkit factors and each outcome. Because toolkit factors were only selected when factors from the end-of-year survey were included, for models using only factors from the observation rubric or only from the exit ticket, we report the R^2, but the test of statistical significance is not applicable. Even without including toolkit factors, baseline covariates explain a large share of variation in ELA and math assessments, and adding toolkit factors does not significantly increase the explanatory power of the models. When toolkit factors are selected, they are not statistically significantly associated with any student outcomes (Exhibit A-5).

Exhibit A-5. Model Performance, Student Outcomes (R^2)

Model	Math Achievement	ELA Achievement	GPA	Attendance
All measures	0.83	0.82	0.21	0.40
Observation rubric only	0.83 n.a.	0.82 n.a.	0.17 n.a.	0.38 n.a.
Exit ticket only	0.83 n.a.	0.84 n.a.	0.17 n.a.	0.53 n.a.
End-of-year survey only	0.83	0.82	0.21	0.40

Note. n.a. indicates that no toolkit factors were selected, and therefore there is no test of statistical signficance.

SRI Education, a division of SRI, is helping federal and state agencies, school districts, major foundations, nonprofit organizations, and international and commercial clients tackle some of the most complex issues in education to help students succeed. Our mission is **to reduce barriers and optimize outcomes for all children, youth, and families.** We do this by conducting high-quality research, supporting use of data and evidence, helping to strengthen state and local systems, and developing tools that improve teaching and accelerate and deepen learning. Our work covers a range of topics, including early learning and development, student behavior and well-being, teaching quality, digital learning, STEM and computer science, literacy and language arts, and college and career pathways.

Silicon Valley
(SRI headquarters)
333 Ravenswood Avenue
Menlo Park, CA 94025
1.650.859.2000

Washington, DC
1100 Wilson Boulevard
Suite 2700
Arlington, VA 22209
1.703.524.2053

education@sri.com

www.sri.com/education-learning/