Utilizing an Inquiry-Agnostic Framework to Advance the Development of a Holistic Teacher Evaluation Tool

Michelle Tenam-Zemach; Daniel R. Conn; Nathan C. Anderson; Erica D. Alstad

E-mail: micheletz@gmail.com; daniel.conn@minotstateu.edu; nathan.c.anderson@minotstateu.edu; erica.alstad@gmail.com

Received November 2024
Accepted for publication September 2024
Published December 2024

Abstract

This paper introduces the development of the CP^{2}R Classroom Observation and Evaluation Tool, an innovative, holistic, and reflective tool designed to assess teachers through four critical domains: Capacity, Passion, Presence, and Relevance. We apply an inquiry-agnostic model, the A+ Inquiry framework, to advance the development of the tool and address the limitations of current teacher evaluation systems. The researchers conducted a mixed-methods study to determine K-12 educational leaders' perceptions of the tool. The findings suggest that educational leaders find the CP^{2}R classroom observation and evaluation tool highly relevant. Nevertheless, significant concerns exist about its feasibility within the recommended time frame for leaders to conduct classroom observations. Recommendations for improving the tool include simplifying its format, enhancing clarity, providing additional resources, and offering comprehensive training. Overall, findings indicate that the evaluation tool can potentially support teacher growth, improve teacher-supervisor relationships, and ultimately enhance student learning outcomes.

Keywords: *assessment; evaluation; inquiry; professional development*

Assessment, in its broadest sense, refers to appraising or evaluating something. Over the past two decades, assessment has shaped nearly every aspect of educational decision-making and outcomes at all levels. As Tenam-Zemach et al. (2021) argue, assessment has evolved into an industrial complex that anchors schools' ecosystems to data in potentially limiting ways. However, data remains essential for growth and development. For teachers, informative and targeted feedback from supervisors and peers, along with the professional support that follows, is critical to their development and growth (Hattie & Timperley, 2007; Kraft & Gilmour, 2022). Black and Wiliam (1998) speak more specifically to assessment in education as the general term that

refers "to all those activities undertaken by teachers - and by their students in assessing themselves - that provide information to be used as feedback to modify teaching and learning activities" (p. 140). Like their students, teachers are learners, particularly when applying feedback data from teaching to their teaching practices. By doing so, teachers can modify their instructional approaches to meet their students' needs better. Teacher attrition and retention are significant issues challenging all school systems (Ingersoll, 2001; Sutcher et al., 2016).

Consequently, providing teachers with feedback on their instruction and classroom behaviors is essential, making teacher evaluation a pivotal component in educational systems. These evaluations ensure that teachers meet specific effectiveness standards and contribute positively to student outcomes. However, traditional evaluation methods are criticized for their narrow focus and inability to provide a comprehensive picture of teacher performance (Darling-Hammond, 2013; Kraft & Gilmour, 2016). Moreover, teacher attrition and retention problems disrupt student learning and impose considerable financial burdens on educational institutions (Ingersoll, 2001; Sutcher et al., 2016). While data from teacher evaluation systems can positively contribute to teacher growth and development under certain conditions, the hierarchical, standards-based, and punitive approaches that dominate most states often frustrate teachers and administrators, leading to job dissatisfaction and attrition (Darling-Hammond, 2015; Kraft & Gilmour, 2016).

To address the limitations of current teacher evaluation systems, the authors embarked on a journey to develop a holistic classroom observation tool that includes the domains and descriptors necessary to transform teaching and learning. This paper introduces the development of an inquiry-agnostic approach to creating a comprehensive, holistic, and potentially transformative teacher classroom observation and evaluation tool. The CP^2R Classroom Observation and Evaluation Tool, developed as an evolved iteration of Anderson et al.'s (2018) CPR Success and Failure Analysis conceptual model, offers a unique and innovative approach to teacher evaluation by assessing teachers through four critical domains: Capacity, Passion, Presence, and Relevance. Additionally, we apply the A+ Inquiry framework, an inquiry-agnostic model, to highlight the process we underwent to advance the development of the teacher assessment tool.

CP^2R Revisited

The CP^2R model offers a holistic, reflective, and non-determinative approach to help someone understand their fitness to perform a particular role (Anderson et al., 2022). Each of the four domains of the model includes several subdomains that contribute to a comprehensive understanding of role suitability and alignment. While it is outside the purview of this essay to delve deeply into the CP^2R framework (see Anderson et al., 2018, 2020, 2022), we provide a brief overview to clarify why we chose the model as the basis for the classroom observation and evaluation tool.

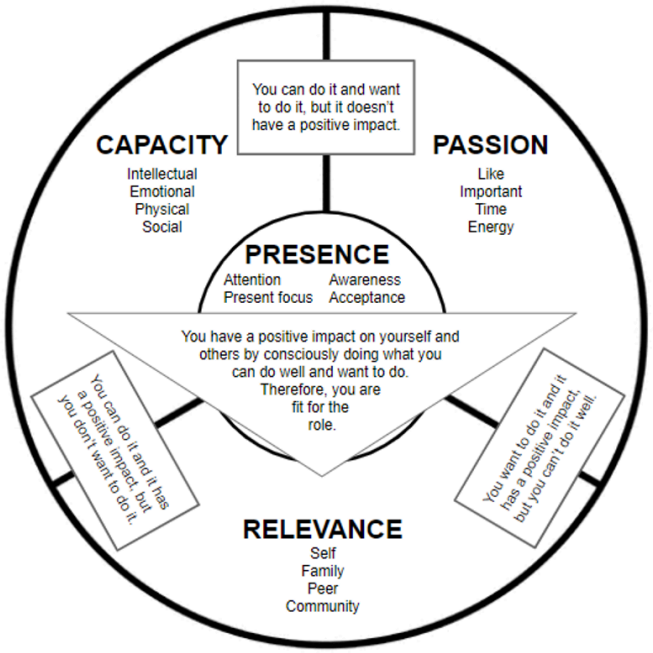
The CP^2R framework assesses a person's fitness for a role through a holistic, reflective, and purposeful approach. Each domain contains several subdomains, which help individuals evaluate their alignment with a specific role or responsibility (Figure 1). The first domain, Capacity, refers to a person's ability to perform their role effectively. The domain includes four subdomains: intellectual, social, emotional, and physical. For teachers, capacity encompasses their skills, knowledge, and competence in delivering educational content. For example, a teacher's capacity to connect socially and emotionally to their students' needs and lives is critical for successful teacher-student relationship building (Jennings & Greenberg,

2009). Evaluating teachers' capacity ensures they possess the necessary qualifications and expertise to fulfill their responsibilities (Anderson et al., 2018).

Figure 1

CP²R Framework

Note. Adapted from *CP²R overview, origin, and evolution* (p. 6), by N. C. Anderson, D. R. Conn, & M. Tenam-Zemach (2022).



Presence, the second domain, focuses on an individual's level of present focus for, attention to, awareness of, and acceptance of their role. Being present is critical for success in any role, as it enhances focus, engagement, and interpersonal interactions, all of which contribute to better performance and outcomes (Brown & Ryan, 2003; Kabat-Zinn, 2003). For teachers, presence involves being attentive to students' needs, adapting to classroom dynamics, and maintaining a positive and supportive learning environment. Being present enables individuals to navigate and respond to challenges more effectively, make informed decisions, and promote meaningful relationships, which is crucial for fostering effective teacher-student interactions (Nightingale, 2013). Developing this domain is particularly vital for teachers because they must be fully engaged and mindful in their educator role.

Passion, the third domain, evaluates a person's enthusiasm and dedication to their role. A person is considered passionate and well-suited for their role when they enjoy their work, find it essential, and invest time and energy. A passionate teacher, for instance, is likely to believe their job is essential and put time and energy into the responsibilities required to fulfill the role successfully, such as developing lesson plans, finding resources, and engaging with colleagues. This domain emphasizes the importance of intrinsic motivation in teaching (Nightingale, 2013). Teachers who see the importance of their work stay motivated and committed to continuous improvement.

Relevance, the fourth domain, assesses the impact and significance of one's work. For this domain, one considers the impact of their role on themselves (Ryan & Deci, 2000), their family, peers, and community (Greenhaus & Kossek, 2022). For teachers to view their work as relevant, they may consider how their efforts align with broader educational goals and how their work impacts their lives, their students' lives, and the community. For relevance to be a strong indicator of a person's fitness to teach, their work must be meaningful and positively influence their students' learning and lived experiences (Anderson et al., 2018).

When reflecting on the need for a teacher evaluation framework to assess teacher classroom instruction and behaviors, we turned to the holistic and comprehensive CP²R model. We understood that using the CP²R framework would encourage the development of a flexible, differentiated, and meaningful tool to support teacher growth and development. We began developing the evaluation

tool by considering the needs of teachers, supervisors, and students. We focused on each domain and its subdomains to develop criteria based on the literature, research, and our collective experiences as teachers, administrators, and education scholars.

Developing a CP²R Evaluation Tool: The Process

After an extensive, deliberative, and collaborative inquiry-driven process grounded in teacher best practices, we developed a holistic classroom observation tool for multiple, shorter walkthroughs throughout the year to address teacher evaluation challenges. The CP²R Teacher Classroom Observation and Evaluation Tool consists of several forms that comprehensively and accurately assess a teacher's instructional practices, student engagement, and overall instructional fitness. The tool also offers opportunities for teachers to participate in their professional growth and development.

One critical knowledge gap in current evaluation methods is the lack of attention to the subtleties of classroom interactions and the dynamic nature of instruction (Pianta & Hamre, 2009). Walkthroughs offer a unique opportunity to observe these nuances, providing a more comprehensive view of the teacher's approach and the learning environment. To address this, as part of the Tool's development, we incorporated a notes section into each subdomain of the Classroom Observation and Data Collection Form, enabling supervisors to document their interpretations more effectively. Recognizing the importance of teacher involvement, we also added a Goal-Setting and Action-Planning Form completed by teachers after a debriefing session with their supervisor. This addition ensures that teachers not only understand the observation criteria but are also actively engaged in their own professional growth. As a result, this tool not only collects valuable data on teaching and learning but also empowers teachers by fostering collaboration and ownership in the evaluation process.Instructional coaches and other in-building support personnel are vital in teachers' evaluation processes (Desimone, 2009; Guskey, 2002). Their ongoing guidance and encouragement are crucial for fostering a growth-oriented culture (Darling-Hammond et al., 2017). When teachers collaborate with expert instructional coaches to learn to apply new practical pedagogic approaches, students' academic outcomes increase (Kraft et al., 2018). Yet, the real power of the CP²R teacher evaluation system approach lies in the debriefing sessions with their supervisors, where teachers can reflect on the data, set meaningful goals, and take control of their professional growth (Knight, 2007). Supervisors, meanwhile, are there to support, not judge—acting as facilitators in the debrief discussions and goal-setting processes, ensuring that the teacher's voice is central to their development journey.

A+ Inquiry

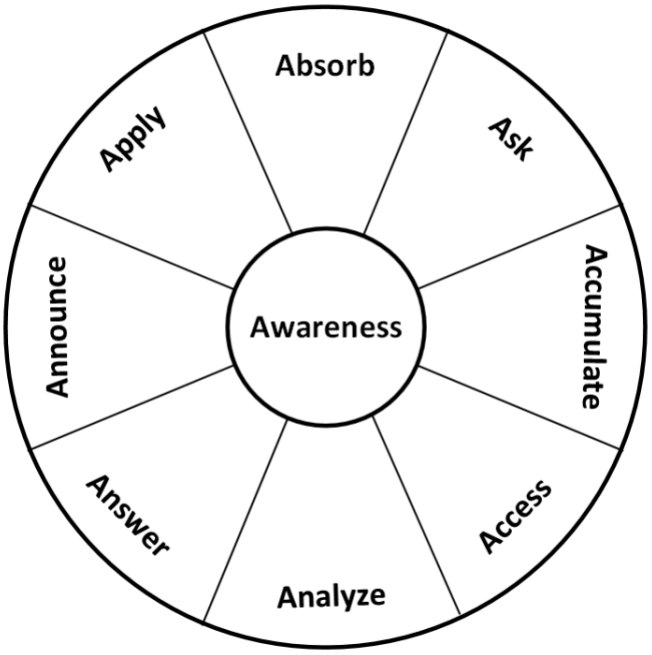
To present our process for developing the tool, we utilize the A+ Inquiry framework (see Appendix A). A+ Inquiry is a universally applicable disciplined inquiry framework synthesizing standard stages of evaluation, assessment, and research (Anderson, 2022; Anderson et al., 2014; Pontenila et al., 2025). The model may be utilized to plan or conduct a study, establish a common language for discussing inquiry processes, tell an evidence-based story, provide a rationale for action, or diagnose gaps in inquiry processes. A+ Inquiry uses alliteration (Bryant et al., 1990; Lea et al., 2008; Stoll, 1940) and visualization (Gilbert, 2008; Tufte, 1990; Ware, 2000) to promote understanding of inquiry stages and how they support one another. The framework is displayed as cyclical, with eight stages connected by a hub. Each stage and the hub begin with the letter A. The stages are Absorb, Ask, Accumulate, Access, Analyze, Answer, Announce, and Apply, which are

bound in the center by a hub of Awareness (Figure 2).

Figure 2

A+ Inquiry Diagram

Note. Adapted from "Disciplined inquiry: Using the A+ Inquiry framework as a tool for eliminating data hoarding, mindless decision-making, and other barriers to effective ESA programming," by N. C. Anderson, M. R. Brockel, and T. E. Kana (2014), *Perspectives: A Journal of Research and Opinion About Educational Service Agencies, 20*(3).



Absorb Stage

In the Absorb stage, we acknowledged what we already knew about the context and found a knowledge gap (i.e., a need to know more about something). Guiding questions in this stage included:

What is Already Known About the Context?

Traditional teacher evaluation methods, while prevalent, have significant limitations in providing a comprehensive and accurate assessment of teacher performance. Traditional teacher evaluation systems rely heavily on classroom observations, student test scores, and self-assessment surveys (Hill & Grossman, 2013). While these methods offer some insights, various biases and external factors influence and limit their scope (Milanowski, 2017). Classroom observations, though common, are often criticized for their subjectivity and variability, as different observers' biases and the sporadic nature of these observations can lead to an incomplete assessment of a teacher's performance (Darling-Hammond, 2013; Hill & Grossman, 2013). Standardized test scores as an evaluation metric also have significant drawbacks, as they are affected by myriad factors beyond a teacher's control, such as students' socio-economic background, prior knowledge, and external support systems. Consequently, student test scores may not accurately reflect a teacher's effectiveness (Braun, 2005; Conn & Tenam-Zemach, 2019; Tenam-Zemach et al., 2021). Self-assessment surveys, while providing a reflective tool for teachers, are often unreliable due to potential biases, with teachers overestimating or underestimating their skills, leading to distorted self-perceptions and evaluations (Taut & Sun, 2014).

We found a need for a more holistic, reflective, and non-determinative approach to teacher evaluation that comprehensively and accurately assesses a teacher's capacity, passion, presence, and relevance in the classroom. To further expand on the Absorb Stage, Author 1 reflects on her participation in developing the CP^2R Classroom Observation and Evaluation Tool.

Author 1's Reflection. *I have always had an affinity for working with K-12 teachers to promote their understanding of best instructional practices.*

Having worked as a teacher, faculty, and in various teacher educator roles, I firmly believe and argue that targeted, articulated, and timely assessment and professional development of teachers is critical to teacher retention and growth. The literature supports these arguments. Darling-Hammond et al. (2017) argue that professional development tailored to teachers' content areas and aligned with school curricula significantly impacts teaching practices and student learning. Professional development is more likely to be applied effectively in the classroom when relevant to the teachers' daily work and challenges.

Furthermore, targeted professional development helps teachers develop specialized knowledge and skills that improve their instructional practices. Guskey (2002) emphasizes the importance of articulated professional development programs that are coherent and connected over time. Rather than isolated, one-time workshops, ongoing professional development programs that build on previous learning experiences transform teaching practice. Such programs encourage teachers to reflect, apply new strategies, and receive feedback, all essential components of adult learning (Knowles, 1984).

As a former professor of curriculum and instruction and a curriculum evaluator, I have visited and observed hundreds of classrooms throughout the country. Typically, I utilize a classroom walkthrough tool that aligns with "best instructional practices" and meets the district's purpose for the evaluation. For example, in one district-wide evaluation, the audit team focused on IB and Gifted and Talented classrooms. Consequently, the evaluation team refined the tool to emphasize research and literature supporting best practices tailored to those subgroups.

In 2018, I began working with several colleagues on a framework to assess a person's role-fit. The CPR model offered a unique and exciting approach to guiding one's reflection on a person's alignment with a particular role. As my colleagues and I explored the model, we developed and advanced it to create the CP^2R *framework. We have applied the model in various ways and contexts. For example, we have trained university professors to use the model to assess their curriculum, improve faculty instruction, and reflect on their fitness to teach social, emotional, and cultural content (for more applications of the* CP^2R *model, see Droptheknowledge.com).*

Exasperated with the classroom observation tools I had been using, I began exploring the potential of adapting CP2R to assess teachers' classroom instruction. Despite my experience and expertise in developing and implementing observation tools—and the various iterations I have used to evaluate teachers in classrooms across the United States—I have rarely, if ever, found that the required tools provided the necessary indicators to comprehensively and accurately assess teachers' instruction and behaviors. I was often frustrated by the absence of indicators providing a more holistic evaluation of the teachers' qualities and observable actions.

Teaching is a complex endeavor. Having taught across various academic levels, I understand the precarious nature of teacher evaluation (Darling-Hammond, 2013; Papay, 2012). After completing a district-wide curriculum evaluation in the fall of 2022 and reflecting on my knowledge of teacher evaluation, professional development, and growth, I knew that my colleagues and I could utilize the CP^2R *model to develop a holistic, reflective, and impactful classroom observation tool that would provide teachers, supervisors, and districts the data needed to improve instruction, teacher-supervisor, and teacher-student relationships. I approached the colleagues with*

whom I had developed the CP²R *model and shared my idea of using it to create a teacher classroom observation tool. My colleagues and I have decades of experience and knowledge in teacher evaluation and teaching best instructional practices; thus, we felt confident that we could produce a tool that would be innovative, impactful, and comprehensive.*

What Is the Knowledge Gap That Needs to Be Filled?

We responded to the need for more holistic approaches to teacher evaluation by developing a first draft of the CP²R Classroom Observation tool. We intended to pilot the tool with current teachers and administrators; however, we needed to know whether educational leaders would consider it useful and helpful. Before proceeding with a pilot phase, we sought to learn more about educational leaders' perceptions of the tool's relevance and feasibility, as well as potential implementation challenges and recommendations for improving the tool.

Why Is It Important to Fill the Knowledge Gap?

Addressing this gap is critical for improving teacher evaluation practices and enhancing teacher development, satisfaction, and retention, ultimately leading to better student educational outcomes. To further answer this question, Author 4 shared the following professional reflection.

Author 4's Reflection. *As a former school principal and now an educational consultant, my experience in classroom teaching and leadership has highlighted the complexities of evaluating teachers and their instructional practices. While traditional teacher evaluations should capture the full scope of a teacher's abilities, relying heavily on infrequent, high-stakes observations, my experiences do not reflect these outcomes. This approach must include the nuanced interactions and evolving dynamics that occur daily in a classroom, often leading to an elaborate show instead of accurately depicting teachers' and students' daily habits and behaviors. I've found that teacher evaluations can lose meaning because of the inauthentic data within these classroom observations, becoming just another box to check for teachers and administrators rather than a starting point for teacher-led growth.*

The evaluation forms available to most school principals lack the necessary components to help supervisors understand and support teacher growth. Rather than support a holistic view of teacher performance, these forms often focus on a narrow set of criteria that fail to capture the breadth of a teacher's daily practices that multiple, shorter observations can accomplish. Essential aspects like a teacher's emotional connection to their teaching and students, their tone and facial expressions during student interactions, and their ability to be fully present in the lesson while maintaining attention to each student in the room often go unnoticed. Additionally, a teacher's willingness to listen unbiasedly and adapt their instruction based on real-time student feedback is crucial yet seldom evaluated.

When evaluations are top-down and punitive, they can erode trust, stifle collaboration, and diminish the teacher's intrinsic motivation for growth. This dynamic creates a counterproductive environment where teachers can interpret their evaluations as threatening their autonomy and efficacy rather than an opportunity for meaningful professional development. Recognizing these limitations during my time as a principal, I took the initiative to create walkthrough tools that reflected my goals as a supervisor; these tools allowed me to gather more meaningful data during my daily classroom visits and provided a more organic and holistic view of teaching and learning. They helped me better understand and support my teachers' growth. However, the district-approved evaluation tool I was required to use needed more flexibility to

incorporate this valuable data into the formal evaluation process. This disconnect reinforced the top-down approach to evaluation, making it challenging to shift teachers' mindsets away from viewing evaluations as coercive rather than growth-oriented. Even when I introduced SMART goals and encouraged staff to take ownership of their development, breaking through this long-established mindset took time and effort. The gaps created by narrow-focused, punitive evaluation tools and a top-down approach to eliciting growth in teaching and learning have created barriers within teacher-administrator relationships, making collaborative growth challenging.

Ask Stage

In the Ask stage, we formulated questions that, if answered, would help close the knowledge gap that we identified in the Absorb stage. The research questions we formulated in this stage included:

> **RQ1:** Do K-12 educational leaders perceive the tool to be relevant?
>
> **RQ2:** Do K-12 educational leaders perceive the tool's implementation as feasible?
>
> **RQ3:** What are K-12 educational leaders' perceptions of challenges with implementing the tool?
>
> **RQ4:** What are K-12 educational leaders' recommendations for improving the tool?

Accumulate Stage

In the Accumulate stage, we described and implemented methods to collect quantitative and qualitative data that we could analyze to answer the questions posed in the Ask stage. This stage is critical in ensuring that the data collected is robust, reliable, and relevant to the research questions (Creswell, 2014). The guiding questions in this stage included:

What Data Could Help Answer the Questions Posed in the Ask Stage?

Data from surveys and feedback from K-12 educational leaders and teacher education scholars regarding the relevance, feasibility, challenges, and recommendations for the CP^{2}R Classroom Observation Tool helped answer the questions posed in the Ask stage (see Appendix B). We gathered quantitative and qualitative data to provide a comprehensive understanding of the tool's effectiveness and areas for improvement (Maxwell, 2013).

Did the Data Need to Be Collected, or Had They Already Been Collected?

We chose to collect primary data through surveys and feedback forms. Collecting fresh data ensured that the feedback was current and relevant to the educational context (Fowler, 2014).

When and Where Were the Data Collected?

We collected the data through a Google Form online over a two-month period during the same academic year as the upcoming pilot phase of the CP^{2}R Classroom Observation Tool implementation. This data collection phase is crucial for testing the tool in a real-world setting and gathering initial feedback to refine it before broader implementation (Patton, 2002).

Who Were the Data Collected From?

The study included 14 survey participants who came from public and private universities and K-12 institutional settings. Of these, approximately 36% identified as school administrators, 27% as classroom teachers, and 7% as instructors (with the term "instructor" being self-defined, as the form included an "other" option). The remaining participants were in various educational roles, such

as university professors. Half of the respondents hold a doctoral degree, 36% have a master's degree, and 14% hold a bachelor's degree. Participants' experience in K-12 education varied. The average number of years participants have worked in K-12 education was 18.2.. Their time in current roles also differed, with four respondents having one year of experience and three having 12 or more years. On average, participants had spent 6.14 years in their current role. Additionally, 57% of respondents indicated responsibility for formally observing teachers in a classroom setting.

What Instrument Was Utilized to Collect the Data?

A Google Form survey with Likert-type items and open-ended questions. This mixed-methods approach allowed for the collection of structured quantitative data and rich qualitative insights (Dillman et al., 2014).

What Was the Procedure for Collecting the Data?

We sent personalized email invitations with a link to the survey to potential reviewers. We chose this method for its efficiency and ability to reach a broad audience while providing a personal touch to encourage participation (Dillman et al., 2014).

Who Was Responsible for Collecting the Data?

We were responsible for distributing the survey and collecting responses. Ensuring a clear division of responsibilities helped keep data integrity and reliability (Creswell & Poth, 2016).

Access Stage

In the Access stage, we retrieved the collected data in preparation for analysis. This stage ensured that we accurately and efficiently prepared the data for later analysis (Miles et al., 2014). The guiding questions included:

Where Were the Data Stored and Retrieved After They Had Been Collected in the Accumulate Stage?

We stored the data in a Google Sheet linked to the Google Form survey. This storage method allowed for easy access and organization of the data (Google, 2020).

What Was the Procedure for Retrieving the Data in Preparation for Analysis?

The research team accessed the Google Sheet to retrieve and prepare the data for analysis. Data cleaning and preparation were crucial steps to ensure the accuracy and usability of the data for analysis (Keller & Warrack, 2003).

Who Was Responsible for Retrieving the Data That Would Be Analyzed?

Author 3 handled data retrieval because he has extensive expertise and experience with data analytics and assessment. Having a specialized analyst ensured that the data were handled with expertise and precision (Miles et al., 2014).

Analyze Stage

In the Analyze stage, we analyzed the retrieved data. This stage involved applying analytical techniques to derive meaningful insights from the data (Creswell & Poth, 2016). The guiding questions included:

What Data Analysis Methods Were Implemented to Analyze the Data That Were Retrieved in the Access Stage?

We performed a quantitative analysis of Likert-type items and a qualitative thematic analysis of open-ended responses. The data were analyzed descriptively by calculating response frequencies by question. For example, we calculated the percentage of respondents by their level of agreement (i.e., strongly agree to strongly

disagree). Quantitative analysis offered statistical insights, while qualitative analysis offered a deeper understanding of the respondents' perspectives (Braun & Clarke, 2006).

What Tools Were Required to Analyze the Data?

We utilized statistical software for quantitative analysis and coding software for qualitative analysis, ensuring the data was analyzed systematically and accurately (Bazeley, 2013).

Who Was Responsible for Analyzing the Data?

Our team included data analysts and qualitative researchers. This multidisciplinary approach ensured comprehensive analysis from multiple perspectives (Creswell, 2014).

Answer Stage

In the Answer stage, we responded to and interpreted the questions posed in the Ask stage, which included finding potential implications and limitations (Creswell & Poth, 2016). The guiding questions included:

What Were the Answers to the Questions Posed in the Ask Stage Based on the Analysis of the Data That Was Conducted in the Analyze Stage?

Analysis results provided insights into the relevance, feasibility, challenges, and recommendations for the CP^{2}R Classroom Observation Tool. These insights were crucial for developing and refining the tool (Miles et al., 2014).

Research Question 1: Do K-12 Educational Leaders Perceive the Tool to Be Relevant? Overall, K-12 educational leaders perceive the CP^{2}R Classroom Observation Tool to be highly relevant, with a significant majority agreeing on its usefulness, ability to inform decisions, and support for improvement plans and standards.

- Usefulness as a tool for supervisors: 100% of respondents agreed or strongly agreed.
- Informing decisions for improved instruction: 84.62% of respondents agreed or strongly agreed.
- Support for improvement plans (SIP, CIP, DIP): 84.62% of respondents agreed or strongly agreed.
- Support for accreditor-required standards: 61.54% of respondents agreed or strongly agreed.
- Interest from supervisors: 76.92% of respondents agreed or strongly agreed.

While the data proves the relevance of the CP^{2}R Classroom Observation Tool, it is crucial to explore the feasibility of its implementation within the time constraints typically faced by educational leaders. The following section delves into K-12 educational leaders' beliefs regarding the practicality of utilizing the tools within a 15–20-minute observation period.

Research Question 2: Do K-12 Educational Leaders Perceive the Tool to Be Feasible to Implement? A significant proportion of K-12 educational leaders have concerns about the feasibility of implementing the tool within the given period, indicating a need for potential modifications to improve its practicality. Only 46.15% of respondents agreed or strongly agreed that leaders could implement the tool within a 15–20-minute observation.

Research Question 3: What Are K-12 Educational Leaders' Perceptions of Challenges with Implementing the Tool? Multiple respondents mentioned that the tool is time-consuming and lengthy. In response to the various survey questions, the main challenges K-12 educational leaders perceived include the

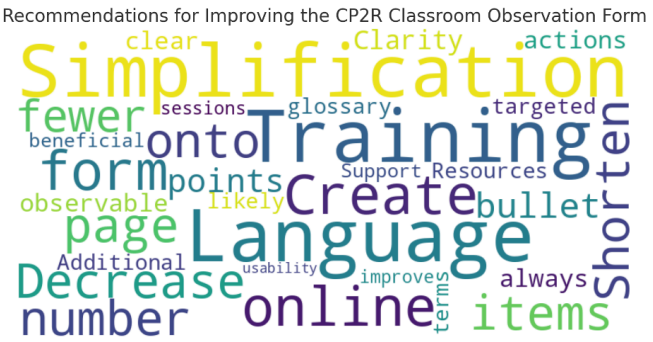
length and complexity of the tool, the time required to complete it, and the potential for variability in interpretation. For example, in responding to the question of the types of challenges administrators envision they may encounter using the tool, Respondent 1, a doctoral veteran administrator, stated, "Time needed to complete the form to observe all indicators." Respondent 2, another public school administrator with 26 years of experience in K-12 education, commented that the challenge with the tool is "time; it's a little long." Respondent 4, a director of teaching and learning at an independent school with a Master's degree and 22 years of experience in education, commented that the tool is "a lot of pages- I could see lots of flipping back and forth." Regarding the theme of complexity and length of the tool, one public school building principal of 17 years pointed out that one challenge of using the tool is "being able to observe all of the elements and complete the form in one session."

Research Question 4: What Are K-12 Educational Leaders' Recommendations for Improving the Tool? To improve the tool's usability, K-12 educational leaders recommend simplifying the tool and offering it online, enhancing clarity and language, providing additional resources like glossaries, and delivering training sessions. Respondent 4 stated that "a targeted glossary of terms would likely be beneficial," while Respondent 1 suggested we "create an online form. Decrease the number of indicators to observe in each subdomain." On the clarity and language of the tool, Respondent 3 suggested that "The language in the observable actions is not consistent…The language should be consistent for both student and teacher boxes." Respondent 12 resisted commenting on offering recommendations on the tool "without knowing what the training will entail."

Visual Representation of the Data. To visually summarize the recommendations provided by K-12 educational leaders for improving the CP^{2}R Classroom Observation Tool, the word cloud in Figure 3 highlights the fundamental suggestions and themes mentioned in their feedback. In summary, the recommendations for improving the CP^{2}R Classroom Observation Tool Include simplifying it, enhancing its clarity of language, providing additional resources like glossaries, and offering comprehensive training sessions. These improvements will make the tool more practical, user-friendly, and effective in supporting educational leaders' needs.

Figure 3

Recommendations for Improving the CP^{2}R Classroom Observation ToolWord Cloud



How Did the Answers Relate to What Was Already Known About the Context (e.g., the Existing Knowledge About the Context That Was Identified in the Absorb Stage)?

We compared the results to existing knowledge and literature on teacher evaluation methods. This comparison helped contextualize the findings and highlight the tool's unique contributions (Maxwell, 2013).

What Were the Limitations of the Answers?

Potential limitations included sample size, response bias, and the generalizability of the findings. Acknowledging these limitations was essential for providing a balanced interpretation of the results (Patton, 2002).

What Were the Implications of the Answers?

The findings informed revisions and improvements to the CP^{2}R Classroom Observation Tool and guided future research and implementation efforts. These implications ensure the tool's relevance and effectiveness (Creswell, 2014).

Announce Stage

In the Announce stage, we communicated the answers (i.e., data analysis results) and applicable potential implications and limitations to stakeholders. This stage was crucial for ensuring we effectively disseminated the findings and could inform practice and policy (Patton, 2002). The guiding questions included:

Which Stakeholders Might Benefit From Being Informed About the Results Revealed in the Answer Stage?

Educational leaders, teachers, policymakers, and researchers concerned about and affected by teacher evaluation would benefit from the results revealed in the Answer stage. These stakeholders could use the findings to enhance teacher evaluation practices (Fowler, 2014).

Which Answers, Limitations, and Implications Were Important to Communicate to Each Stakeholder?

We issued relevant findings, limitations, and actionable recommendations. Tailoring the communication to each stakeholder ensured the information was practical and applicable (Richards & Hemphill, 2018).

What Was the Procedure for Disseminating the Results to, and/or Discussing the Results with, the Stakeholders?

We are preparing a comprehensive report and presentation to share with stakeholders through meetings, conferences, and publications. Effective dissemination strategies will ensure broad reach and impact (Bazeley, 2013). In addition to publishing this journal article, we shared findings with local school districts, regional education cooperatives, and their state education department. We also issued a press release after the tool received state approval as an official teacher evaluation tool. We plan to present at relevant conferences at state and national levels and continue to seek publication opportunities.

Who Was Responsible for Disseminating the Results and Other Relevant Information to the Stakeholders?

As a team, we disseminated the results and other relevant information to local school districts. Our coordinated efforts ensured the dissemination was prompt and well-received (Patton, 2002).

Apply Stage

In the Apply stage, we made decisions and took actions based on the answers, limitations, implications of the answers, and applicable discussions among stakeholders. This stage ensured that the findings led to practical improvements and informed future practice (Creswell & Poth, 2016). The guiding questions included:

What Decisions or Actions Were Informed By the Results Revealed in the Answer Stage and, if Appropriate, Relevant Discussion That Occurred in the Announce Stage?

To ensure the CP^{2}R Classroom Observation and Evaluation Tool is still relevant and effective, revisions were made, implementation strategies were developed, and plans for future research were set (Maxwell, 2013). Although all respondents agreed the tool would be useful, only 46.15% believed it could be implemented within a 15–20-minute observation period. In response, we are recruiting more participants from seven K-12 school districts in north central North Dakota to

assess its feasibility further. When considering the respondents' recommendations for improvement to the tool, we simplified the process by adding a rating page focused solely on the observable action header for each subdomain. We developed an online version of the tool for data collection, clarified language where applicable, and are developing more resources like a glossary of terms. Moreover, we have increased the number of virtual training sessions, and we plan to record them and make them available online to participants using the tool. Additionally, we recommend that principals focus on specific domains during observations for better time management and diligence.

Why Were the Decisions or Actions Important to Implement?

To ensure the tool was effective, practical, and beneficial for teacher evaluation and development, implementing these decisions was critical for maximizing the tool's impact (Patton, 2002).

What Cautions Were Considered When Making Decisions or Taking Actions Based on the Results?

We considered potential biases, limitations, and the need for ongoing evaluation and feedback. These cautions ensured that our actions were well-informed and balanced (Miles et al., 2014).

Who Was Responsible for Making the Decisions or Taking the Actions?

In the Apply stage, we made decisions and took actions based on the answer, limitations, and implications of the answer, as well as applicable discussions among stakeholders. This stage ensured that the findings led to practical improvements and informed future practice (Creswell & Poth, 2016). The guiding questions included:

Why Were the Decisions or Actions Important to Implement?

To ensure the tool was effective, practical, and beneficial for teacher evaluation and development. Implementing these decisions was critical for maximizing the tool's impact (Patton, 2002).

What Cautions Were Considered When Making Decisions or Taking Actions Based on the Results?

We considered potential biases, limitations, and the need for ongoing evaluation and feedback. These cautions helped ensure our actions were well-informed and balanced (Miles et al., 2014).

Who Was Responsible for Making the Decisions or Taking the Actions?

We collaborated with educational leaders and stakeholders who provided feedback on the tool. In the future, we plan to work with up to 10 school districts during the 2024-2025 academic year. In preparation for the next stage of piloting the CP^2R classroom observation tool, we have been in regular conversation with a regional education cooperative, and they plan to work with each participating district to gather data to inform future iterations of the tool. This extensive collaboration would offer a diverse range of feedback and data, enhancing the validity and applicability of the CP^2R Classroom Observation and Evaluation Tool across different educational contexts. By engaging multiple districts, we aim to refine and confirm the tool through practical application, ensuring its effectiveness and utility in various school settings.

Conclusion

The CP^2R Classroom Observation and Evaluation Tool is unique in its focus on contributing to the relationship-building process between supervisors, teachers, and students. Its comprehensive focus on capacity, passion, presence, and relevance ensures that teacher evaluations are not merely a means of

evaluation but a tool for meaningful professional growth and development. The development of the CP^{2}R Classroom Observation and Evaluation Tool offers an impactful solution to the limitations of traditional teacher evaluation systems by adopting a holistic, inquiry-agnostic approach. This framework provides a comprehensive assessment that emphasizes teacher growth, student engagement, and the overall effectiveness of instructional practices. The CP^{2}R model's reflective and non-determinative approach encourages teachers to assess their alignment with educational roles, fostering a supportive and growth-oriented environment for educators.

The A+ Inquiry framework further enhances the development process of the CP^{2}R tool by offering a structured, systematic approach to evaluation. This framework emphasizes the importance of each stage of the inquiry process, ensuring that the tool is evidence-based and adaptable to the unique needs of different educational contexts. By involving teachers in the evaluation process through continuous feedback and debriefing sessions with supervisors, the CP^{2}R tool empowers educators to participate actively in their professional development. It facilitates teachers taking ownership of their professional development and potentially increases motivation and job satisfaction.

Moreover, the CP^{2}R tool addresses the issue of teacher attrition and retention by providing a more meaningful and supportive evaluation process. By focusing on aspects of teaching overlooked in traditional evaluations, such as the subtleties of classroom interactions and the evolving nature of instruction, this tool contributes to a more exact and nuanced understanding of teacher performance. The reflective nature of the tool and the data accumulated in the evaluation process provide space for teachers to improve their practice in contextual ways. The emphasis on supportive relationships between teachers and supervisors ensures participants view evaluation as a collaborative process aimed at professional growth rather than a punitive measure.

The CP^{2}R Classroom Observation and Evaluation Tool significantly advances teacher evaluation practices. Adopting a holistic, inquiry-agnostic framework provides a comprehensive, flexible, and meaningful assessment that supports teacher development and enhances student learning outcomes. As educational systems continue to evolve, tools like CP^{2}R can play a crucial role in fostering a more supportive and effective teaching environment.

References

Anderson, N. (2022). Synthesizing frameworks and tools to develop a plan for evaluating an online data utilization curriculum for teachers. *Evaluation and Program Planning, 94*. https://doi.org/10.1016/j.evalprogplan.2022.102148

Anderson, N. C., Brockel, M. R., & Kana, T. E. (2014). Disciplined inquiry: Using the A+ inquiry framework as a tool for eliminating data hoarding, mindless decision-making, and other barriers to effective ESA programming. *Perspectives: A Journal of Research and Opinion About Educational Service Agencies, 20*(3).

Anderson, N. C., Conn, D. R., Gamas, K. C., Borkhuis, B., & Lantto, J. (2018). Capacity, passion, relevance, and presence: A conceptual framework for the interpretation and study of success. *Journal of Research Initiatives, 4*(1), 11. https://digitalcommons.uncfsu.edu/jri/vol4/iss1/11

Anderson, N., Conn, D. R., & Tenam-Zemach, M. (2022). *CP2R overview, origin, and evolution*. Author.

Anderson, N. C., Conn, D. R., Tenam-Zemach, M., Clemente, I., Schaefer, L., & Zemach, J. (2020). Are schools "robbing" students? Resuscitating the purpose of school through CPR. *Curriculum and Teaching Dialogue, 22*(1–2), 183–196.

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo* (2nd ed.). Sage Publications. https://doi.org/10.1080/14780887.2014.992750

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7–74.

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models.* Educational Testing Service.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848.

Bryant, P. E., MacLean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26*(3), 429–438.

Conn, D. R. & Tenam-Zemach, M. (2019). Confronting the assessment industrial complex: A call for a shift from testing rhetoric. *Special Issue of Journal of Curriculum Theorizing: Curriculum Theorizing in the Post-Truth Era, 34*(3), 122–135.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches.* SAGE publications.

Darling-Hammond, L. (2013). *Getting teacher evaluation right: What really matters for effectiveness and improvement.* Teachers College Press.

Darling-Hammond, L. (2015). Can value-added add value to teacher evaluation? *Educational Researcher, 44*(2), 132–137.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Policy, 23*(1), 69–87.

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method.* John Wiley & Sons.

Fowler, F. J. (2014). *Survey research methods* (5th ed.). SAGE Publications.

Gilbert, J. K. (2008). Visualization: An emergent field of practice and enquiry in science education. In J. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 3–24). Springer.

Google. (2020). *Google forms.* https://www.google.com/forms/about/

Greenhaus, J. H., & Kossek, E. E. (2022). Meaningful work and family: How does the pursuit of meaningful work impact family life?

In M. L. Rockmann, J. A. Dutton, & J. S. Detert (Eds.), *Exploring meaningful work* (pp. 219–237). Oxford University Press. https://doi.org/10.1093/oso/9780197559094.003.0013

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching, 8*(3), 381–391.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Hill, H. C., & Grossman, P. (2013). Learning from teacher observations: Challenges and opportunities posed by new teacher evaluation systems. *Harvard Educational Review, 83*(2), 371–384.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38*(3), 499–534.

Jennings, P., & Greenberg, M. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, *79*, 491–525. https://doi.org/10.3102/0034654308325693

Kabat-Zinn, J. (2003). Mindfulness-based interventions in context: Past, present, and future. *Clinical Psychology: Science and Practice, 10*(2), 144–156.

Keller, G., & Warrack, B. (2003). *Statistics for management and economics (with Info Trac)*. Thomson Higher Education.

Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction*. Corwin Press.

Knowles, M. S. (1984). *Andragogy in action: Applying modern principles of learning.* Jossey Bass.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Educational Research, 88*(4), 547-588

Kraft, M. A., & Gilmour, A. F. (2016). Can principals promote teacher development as evaluators? A case study of principals' views and experiences. *Educational Administration Quarterly, 52*(5), 711–753.

Kraft, M. A., & Gilmour, A. F. (2022). Revisiting the widget effect: Teacher evaluation reforms and the distribution of teacher effectiveness. *Educational Evaluation and Policy Analysis, 44*(1), 71–95.

Lea, R. B., Rapp, D. N., Elfenbein, A., Mitchel, A. D., & Romine, R. S. (2008). Sweet silent thought: Alliteration and resonance in poetry comprehension. *Psychological Science, 19*(7), 709–716.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). SAGE Publications.

Milanowski, A. (2017). Lower performance evaluation practice ratings for teachers of disadvantaged students: Bias or reflection of reality? *AERA Open, 3*(1). https://doi.org/10.1177/2332858416685550

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). SAGE Publications.

Nightingale, E. (2013). *The strangest secret*. Rough Draft Printing.

Papay, J. (2012). Refocusing the debate: Assessing the purposes and tools of teacher evaluation. *Harvard Educational Review, 82*(1), 123–141.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). SAGE Publications.

Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational Researcher, 38*(2), 109–119.

Pontenila, S., Stephens, E., & Anderson, N. C. (2025). Using A+ Inquiry as a framework for exploring faculty needs related to program assessment workload. *Intersection: A Journal at the Intersection of Assessment and Learning, Early View*. https://doi.org/10.61669/001c.126752

Richards, K. A. R., & Hemphill, M. A. (2018). A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical Education, 37*(2), 225–231. https://doi.org/10.1123/jtpe.2017-0084

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67.

Sparr, J. L., Miron-Spektor, E., & Lewis, M. W. (2022). From a label to a metatheory of paradox: If we change the way we look at things, the things we look at change. *Academy of Management*. https://doi.org/10.5465/AMC.2021.0006

Stoll, E. (1940). Poetic alliteration. *Modern Language Notes, 55*(5), 388-390.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). *A coming crisis in teaching? Teacher supply, demand, and shortages in the U.S.* Learning Policy Institute.

Taut, S., & Sun, Y. (2014). The development and implementation of a new teacher evaluation system. *Educational Measurement: Issues and Practice, 33*(3), 12–20.

Tenam-Zemach, M., Conn, D. R., & Parkison, P. T. (2021). *Unraveling the assessment industrial complex: Understanding How Testing Perpetuates inequality and injustice in America.* Routledge.

Tufte, E. R. (1990). *Envisioning information*. Graphics Press.

Ware, C. (2000). *Information visualization: Perception for design*. Morgan Kaufmann Publishers.

Author Bio

Michelle Tenam-Zemach is a former Professor of Curriculum and Instruction and the Executive Director of Strategic Initiatives at Drop The Knowledge consulting agency.

Dr. Daniel R. Conn is Chair of Education at Minot State University and President-Elect of the American Association of Teaching and Curriculum.

Dr. Nathan C. Anderson, Director of Institutional Assessment at Minot State University, focuses on data-driven decision-making, educational frameworks, and community service, while savoring family and outdoor activities.

Erica D. Alstad, M.Ed., former school principal, now education consultant and realtor, is active in her community and loves a healthy lifestyle, dogs, and family.