

DO YOU SEE WHAT I SEE?

**DISTRICT DESIGNS LEARNING PLAN TO DEVELOP
A CLEAR VISION OF EFFECTIVE INSTRUCTION**

As Larry Gerardot, a principal in Fort Wayne (Indiana) Community Schools, sat in front of a computer, he had no idea how the new project in which he had been asked to participate would affect his work and the work of other principals. Yet he knew that Fort Wayne Community Schools had decided that the district would approach inter-rater reliability as professional learning, starting with the principals. Though he was uncertain of the outcome of this work, he was intrigued with the power of principals working and learning together on the RISE Indiana Teacher Effectiveness Rubric — the district's instrument for evaluating instructional practice — and improving his practice in supporting teacher learning.



By Kay Psencik, C. Todd Cummings, and Larry Gerardot

The district leadership team of Fort Wayne Community Schools, Indiana’s largest school district, has focused on becoming a learning organization for many years, due primarily to the leadership of Superintendent Wendy Robinson (Hirsh, Psencik, & Brown, 2014). Valuing professional learning, she partnered with organizations such as The Wallace Foundation and Learning Forward and consulted Michael Fullan’s work to build leadership capacity. She wanted to ensure the district focused on developing a skilled and committed district and principal leadership core to achieve the district’s moral purpose.

In 2010, as one of its major initiatives, the district implemented the RISE Indiana Teacher Effectiveness Rubric — a principal and teacher evaluation system to clarify for teachers and administrators what highly effective, rigorous instruction really looks like. The district uses the rubric, which was developed by Indiana Department of Education and guided by Charlotte Danielson’s Framework for Teaching, to evaluate classroom teachers’ instructional practice.

The rubric’s 24 measures cover four major domains: purposeful planning, effective instruction, teacher leadership, and core professionalism.

At the same time, the district received a Teacher Incentive Fund grant to provide stipends for teachers based on student performance data and their evaluation. The evaluation carried 60% of the weight in determining stipends. As a result, the district paid nearly \$8 million in teacher effectiveness stipends in 2012-14.

District leaders began to analyze the teacher evaluations to determine whether principals were rating teachers across the district with the same lenses. They wanted to be sure that principals were observing instruction in the same way and in agreement on ratings teachers received. Data from five years of implementation of the RISE Indiana Teacher Effectiveness Rubric showed principals were all over the map in scoring instruction.

Through data analysis, they found that not all principals had a clear or common understanding of the rubric’s elements. They also realized that the district had little professional learning in place for principals that focused on teacher evaluation.

STEPS IN THE PROCESS

Establish a theory of change and logic model.



Establish a leadership team.



Develop clear definition of terms from the RISE Indiana Teacher Effectiveness Rubric.



Establish exemplars.



Hire external partners to provide video and manage teacher ratings.



Hire external partner to make 12 videos for the district of exemplar teachers.



Engage the entire principal corps and district leaders in defining terms.



Establish inter-rater agreements among the leadership team members.



Establish norms for videos.



Begin conversations.



Establish protocols to guide conversations and bring principals to rater agreement.



Test to identify areas of agreement and discrepancies.



Support.

DEVELOPING CLARITY

District leaders determined that principals needed professional learning with an emphasis on inter-rater reliability — especially in purposeful planning and effective instruction, the first two of the rubric’s domains.

Modeling their work after the Measures of Effective Teaching project (Bill & Melinda Gates Foundation, 2012), district leaders sought a process that would meet the unique needs of the district and work toward ensuring principal rater agreement. They realized that the most effective way to do this work is to increase principals’ conversations about high-quality instruction.

This focus deepens the district’s efforts at becoming a learning system. The district leadership team has established a clear vision and definition of standards-driven professional learning to ensure that all in the organization are learning in powerful ways. Team members know that if they are to achieve their moral purpose — educating all students to high standards — they need to engage teachers and principals in a cycle of continuous improvement.

The district superintendent and district leadership team believe professional learning is the central process for continuous improvement. Leaders focused their work on inter-rater reliability to establish effective approaches to engaging principals in deep conversations around instruction and key elements of the RISE Indiana Teacher Effectiveness Rubric.

District leaders knew they must start with a clear vision of inter-rater reliability and build an effective change process that made sense to everyone in the organization in order to develop this process with fidelity, so they did their homework and lined up strategic partners to buttress the work.

THE RESEARCH

As a starting point, district leaders relied heavily on the work of the Measures of Effective Teaching project and one of its principal authors, Tom Kane. Kimball & Milanowski (2009) and Graham, Milanowski, & Miller (2012) found that quality observation verified by a well-trained observer added validity to the evaluation process and that adding even a second observer creates even stronger ratings. Consequently, inter-rater reliability is an essential learning design to support principal and teacher learning that results in highly effective instruction every day for every child.

In addition to the research, district leaders drew on support from the Harvard University Strategic Data Project in the Center for Educational Policy Review. Having access to Measures of Effective Teaching project’s principal authors Tom Kane and Andrew Ho helped guide early theoretical underpinnings.

PARTNERSHIPS

Realizing that the process would need key partners to ensure success, the district selected Empirical Education’s Calibration

GUIDING ASSUMPTIONS

- Establishing clear guidelines and thorough, intense practice through experience with peers strengthens inter-rater reliability and observer agreement. If observers and raters have clear and concise instructions about how to rate behavior and can come to agreement about this rating, this agreement increases consistent ratings across the district.
- When using qualitative data using two or more observers, developing inter-rater reliability and observer agreement ensures that results generated will be useful in understanding the effectiveness of all teachers based on common vocabulary and can be used to design professional learning.
- If even one of the observers is erratic on his or her scoring system, the entire system may be jeopardized as perceptions of others may interfere with its effectiveness.
- Developing inter-rater reliability and observer agreement is more about having clear distinguishing descriptors, exemplars, and conversations than about simple agreement.

and Certification Engine as the vehicle to host videos and the calibration tests. The district also chose Edivate — from School Improvement Network — to create videos of a diverse group of teachers to highlight teachers teaching at a highly effective level.

Learning Forward senior consultant Kay Psencik provides district leadership with the tools to build an inter-rater reliability process grounded in the principles of professional learning and guided by a framework that includes KASAB (knowledge, attitudes, skills, aspirations, and behavior), theory of change, and logic model.

ESTABLISHING EXPECTATIONS

The district leadership team considered several statistical approaches during the design phase. What the district really wanted was absolute agreement among all principals when they observed lessons. The team established the expectations and percentage of agreement as a standard all principals must meet.

Those expectations include:

- In order to become a trainer, principal leaders had to develop 90% agreement on all elements in planning and instruction (domains 1 and 2 of the rubric).
- To receive certification, principals and raters must agree at a minimum of 85% of all elements in planning and instruction. For example, there are five components in the purposeful planning category. The group of observers of a common teacher plan and observed video lesson must agree on an exact rating with the norm established by the leader-

ship team on four of those five elements.

- The team will discuss any discrepancy for any element. Team members will share their thinking and, using their observation notes, the definitions, and the rubric, they will work toward coming to agreement about its rating.
- Ultimately, every principal will meet the standard through taking a test and matching the norms established by the principal leadership team.

ASSESSMENTS

Principals pretested for agreement before engaging in the learning process to determine significant areas of agreement and disagreement. Facilitators monitor progress by giving assessments regularly throughout the learning sessions so that they focus on the needs of the learners and differentiate instruction.

Principals are required to be certified to rate teachers. If a principal is unable to reach the standards of agreement required, the district provides intense coaching and support and a certified second evaluator for the school.

Principals who don't meet the standard have multiple opportunities to learn and meet the certification requirement.

THE LEARNING PROCESS

At the first meeting of the leadership team, Gerardot became excited about the work. He knew it would be a challenge to do the work well, but he believed that if they could do a great job, it would have a significant impact on teaching and learning in the district.

As he became clear about the work to be done, he and his teammates jumped right in. He realized that the first task was to analyze the terms in the rubric and consider those that might be troublesome. The team found many words that might be easily interpreted in different ways and some that had multiple definitions. Furthermore, he knows that the team's work was to develop a definition of terms that would mirror the district's purpose and definition of rigorous instruction.

First, the district established a clear purpose for the learning process and worked to ensure that everyone understood it. The purpose of this program is to establish a professional learning and certification system for all principals and assistant principals to ensure reliable use of the rubric.

The process includes six steps:

1. Ensure everyone knows the purpose and process of the work.
2. Develop precise and clear definitions of terms unclear in the district's rubric.
3. Develop a training manual for a group of trainers to ensure consistency.
4. Establish ongoing districtwide collaboration and support.
5. Clarify certification procedures.
6. Establish post-certification support and monitoring.

The superintendent, the cabinet, district leaders for this project, and the facilitator developed the district's proposed

initiatives to improve the inter-rater reliability of classroom observation rating. These included a train-the-trainer approach and developing a user's manual for the rubric. In a vision statement, district leaders laid out the program design and described the implementation of the program.

The district's goal was to ensure that every principal engaged in conversation about the rubric, the definitions, and their observations of instruction by July 2015 and was certified by June 2016. Because of the short timeline, several components of the project needed to be managed at the same time. The district needed to hire a consultant to guide the process, identify an effective certification calibration engine and use it effectively, and create videos showing highly effective teachers, as well as launch a stringent professional learning program for all district administrators.

The district established a district leadership team of principals from all school levels. The team also included district leaders responsible for teacher evaluation as well as those responsible for curriculum, assessment, and instruction. The district set criteria for this team, approved that criteria through the superintendent's cabinet, and requested the principals to join the group.

This leadership team defined the terms in the rubric and sought exemplars to be sure all were seeing with the same eyes. As the leadership team became proficient in identifying terms and recognizing those indicators in video of lessons, the group turned to ways to engage the entire administrative team.

The team sought feedback from the larger community of administrators, then used the feedback to make revisions. The goal was to develop as clearly articulated definitions as possible so that people could see the definition in the same way.

The principals became engaged in the process and could see the value of the work they were doing together. As Gerardot reported, "I shared this process with my teachers, and they are so excited about this work. They are eager to deepen their understanding of the definition because we all want to improve our practice."

At the same time that principals were working on definitions, district leaders contracted with School Improvement Network to create 12 teaching videos mirroring the terms being defined by the principal leadership team. The leadership team set criteria for the selection of these teachers, balancing the list by race, gender, and sexual orientation in order to capture the widest view of the district. Most importantly, the teachers had to be rated highly effective on the rubric.

Once the principal leadership team was satisfied with its definitions, the work of calibration began. This team began observing videos and, using the rubric and their definitions, they rated teachers in the videos on each element in the first two domains of the rubric.

The process of viewing the videos, scoring the elements, discussing the rationale for the ratings, and working toward consensus proved to be time-consuming. The leadership team

spent one day a month viewing the videos and hosting the essential conversations around their observations. The leadership team had to meet that standard of agreeing on 90% of the elements in each domain.

As the principals in the district leadership team began to use the definitions while viewing video of classroom teachers instead of just the rubric, they had an aha moment. One principal reported, “When we just used the rubric, we were all over the place in our ratings — there was no agreement. We have used just the rubric for five years, and we were in a habit of just using the rubric. Our facilitators had to remind us to pay attention to the rubric term definitions that we had been working on for almost eight months. When we used the definitions, we realized we were in closer agreement on our ratings.”

The district principal leadership team viewed video after video, stopped and discussed each element, working toward agreement, and continued the process until the team met the standards of agreement.

After celebrating their success, the team began viewing video and establishing the norms all other principals would have to meet. They realized they were still learning.

Once the videos were normed, the principal leadership team began to host afternoon sessions with all principals to give them the opportunity to work through the same processes and to have the same conversations the team had been having. They began with observing video, scoring that video in domains 1 and 2 using the definitions as well as the rubric, and hosting conversations.

They held multiple afternoon meetings at elementary, middle, and high school levels. Two principals worked together at each level. They created a protocol to engage all principals in the conversation and work toward agreement. They were all working toward their first assessment date 12 months later. Everyone began to deepen their understanding of the elements and what they looked like. They were beginning to wear the same glasses.

On July 14, 2015, almost 12 months after the start of the work, all principals and assistant principals met to take their preassessment and engage in meaningful conversations around the instruction they were observing. All principals in the district were at 65% absolute agreement on all elements. Sessions for the next school year will focus on areas where they were not in agreement. After 10 months of deep conversations, they will take their full test.

PERCEIVED INITIAL IMPACT

Participants say that having collegial conversations around definitions, constructing common meaning regarding instruction and the rubric, and engaging with vertically aligned teams have already impacted their system of support for teachers. They report:

- Increased precision and quality of feedback comments;
- More consistent ratings across all forms of feedback;

- Greater clarity and understanding in the relationship between domains 1 and 2; and
- Better understanding on the part of teachers and coaches of the terms and vocabulary in the rubric.

CHALLENGES

Participants encountered several challenges:

- **Staying focused on this process as professional learning and not certification.** The district did not choose to just certify the principals, but to ensure there was ample time for learning from each other. However, when principals know they will be tested and certified through the process, they naturally concern themselves with that process rather than what they are learning. The leadership team is key to ensuring that principals stay focused on their own learning and the learning of their peers.
- **Ensuring that all principals can distinguish between evaluation and rater agreement.** Evaluation of teacher effectiveness has many components. Principals do multiple drop-in observations and view artifacts in the classroom, such as unit designs and student work, to make a final rating. The classroom formal observation is only one component. Inter-rater agreement is a focus on the lenses the principal uses to see the rubric in the classroom.
- **Developing precision in observations and descriptive language** to distinguish differences in observations and move toward agreement.

LESSONS LEARNED

By engaging in this process, participants came to understand a few key concepts.

Definition of terms matters. The rubric gives principals and teachers clarity around quality instruction, but terms in the documents often have multiple meanings and lack clarity of vision. When principals come to agreement about what the terms mean, and then have multiple opportunities to discuss what those terms look like while viewing lessons, they begin to see together.

Collaboration time matters. Principals spend many hours evaluating teachers. They really value time together to discuss their observations and work together toward common agreement about their observations. The leadership team is adamant about ensuring that all principals have extensive time to work with each other, discuss video lessons together, and learn from each other. They requested and were granted longer time than planned to engage their peers in the same level of conversation and dialogue that the leadership team had experienced.

Principals were concerned that the process would be rushed and they would not have the same rich experience as the leadership team. They appealed to the cabinet for longer working time and multiple windows to certify. This request led to many

Continued on p. 23

est and curiosity to validate and encourage. Then ask teachers to describe ideas in more detail so you can picture how it would transpire in the classroom. When time permits, have teachers use you as a mock audience to teach the content or skill. This short trial run can help uncover vague language or plans that lack specifics. In many cases, it will also reveal critical sequences in the teaching process that were missing altogether.

RICHER OPPORTUNITIES FOR LEARNING

Diligently and consistently modeled and implemented, practical unpacking strategies can help an educator community develop shared understanding of underlying ideas, uncover gaps in grasp of instructional practices, and prepare lessons with improved clarity and richer opportunities for student learning.

A central goal of communication is to cohere — “to coalesce fragments of information back together into a single understanding” (Atkinson, 2003). This definition describes well one of the most difficult tasks of teaching. And it’s actually the origin of the word communication: to “make common” or “bring together.”

Vague words produce underdeveloped conceptions, limit-

ing teacher growth and understanding of practice and leaving students with ambiguous ideas. Well-defined and specified language paves the way for purposeful classroom interaction, minimizes unproductive struggle, and creates opportunities to learn.

REFERENCES

Atkinson, C. (2003, April 22). *Bullet points kill (effective communication)* [Web log post]. Available at www.marketingprofs.com/3/atkinson6.asp.

Ermeling, B.A. & Graff-Ermeling, G. (in press). *Teaching better: Igniting and sustaining instructional improvement*. Thousand Oaks, CA: Corwin Press.

Genevieve Graff-Ermeling (genevieve.ermeling@lhsoc.org) is chief academic officer at Orange Lutheran High School in Orange, California. Bradley A. Ermeling (brad.ermeling@gmail.com) is principal research scientist at Pearson Research and Innovation Network. Ronald Gallimore (ronaldg@ucla.edu) is distinguished professor emeritus at University of California, Los Angeles. ■

Do you see what I see?

Continued from p. 18

afternoons of rich conversations with their peers.

Leadership matters. The final, most important lesson from the project was how principals took over the leadership and facilitation. From writing the protocol and implementation to planning for the districtwide assessment, leadership team members were vocal advocates for the power of a thoughtful, reflective, conversational process.

NEXT STEPS

As the district moves closer toward rater agreement among all principals, it plans to take other approaches.

First, the district will work with teachers to understand the definitions and use them with precision in their collaboration to design curriculum maps, units of study, assessments, and lessons to match the descriptors in the first two domains.

The district will also work to develop inter-rater agreement among those who evaluate principals and program directors.

Finally, the district will work to ensure that the conversations principals are having around quality instruction continue through ongoing professional learning and district leadership meetings.

One principal sums up the impact of the professional learning on his work: “I learned today that I need to pay more attention to the rubric and the definitions when I do my observations,” said Chad Hasong, principal of North Side High School. “I had begun to make assumptions about what this

rubric says, and this work is going to reshape the way I observe teachers and give them feedback.”

REFERENCES

Bill & Melinda Gates Foundation. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Author.

Graham, M., Milanowski, A., & Miller, J. (2012, February). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*. Center for Educator Compensation Reform. Available at <http://files.eric.ed.gov/fulltext/ED532068.pdf>.

Hirsh, S., Psencik, K., & Brown, F. (2014). *Becoming a learning system*. Oxford, OH: Learning Forward.

Kimball, S. & Milanowski, A. (2009). Examining teacher evaluation validity and leadership decision making within a standards-based evaluation system. *Educational Administration Quarterly*, 45(1), 34-70.

Kay Psencik (kay.psencik@learningforward.org) is a Learning Forward senior consultant. C. Todd Cummings (c.todd.cummings@fwcs.k12.in.us) is manager of the System of Support and Compensation and Larry Gerardot (larry.gerardot@fwcs.k12.in.us) is principal of the Career Center at Anthis in Fort Wayne Community Schools in Indiana. ■