Educational Renaissance Vol. 6 (2017)

Research Article

Sharing the Responsibility for Teacher Preparation: Collaborative Design of a Teaching Observation Rubric

Cathy Yun, PhD, California State University, Fresno
email: cyun@csufresno.edu

Lisa Bennett, PhD, California State University, Fresno
email: lbennett@csufresno.edu

Abstract

This case story describes the collaborative development and use of a teaching observation rubric to support and scaffold evidence-based changes in the focal educator preparation program (EPP). The case highlights the power of EPP-district collaborations for improving the teaching and coaching practices for both stakeholders while strengthening trust. Additionally, this case demonstrates how the focal EPP worked with districts to ease preservice-to-inservice transitions for novice teachers, through development of a feedback instrument that spans the transition, thereby decreasing teacher attrition within the first three years on the job.

Based on current research and literature, scholars, experts, and school districts are calling for transformative change in how teachers are prepared toward programs built around a strong clinical practice model in partnership with school districts (Grossman, 2010; National Research Council, 2010). The National Center for Teacher Residencies (NCTR, 2015) identified three necessary shifts for moving toward a clinically orientated teacher preparation program: 1) Restructure clinical experiences, emphasizing competency-based assessments aligned to district and/or state measures; 2) Revise courses to include more theory-to-practice connections and opportunities for simulations and rehearsals; and 3) Build authentic and substantive collaborations with schools and school districts. The ongoing rubric work described in this proposal addresses all of these three shifts.

The first shift is toward focus on clinical experiences in the field, with meaningful feedback, as a critical part of learning for novice teachers. Many traditional models of teacher preparation privilege coursework and position fieldwork as a place where candidates have the option to try out coursework learning. Recent critiques of such traditional models call for the clinical experience to be the central component, with prolonged time in the field, multiple opportunities to practice in authentic contexts, and intentional guidance or scaffolding (Grossman, 2010). The rubric was created to help coaches focus on specific action-oriented competencies aligned to both district and state standards. The rubric provides a "common language," vetted by multiple stakeholders, supporting consistency and helping to prevent mixed messages or misinterpretations of jargon. Using the rubric, various constituents can provide specific feedback and next steps for strengthening candidate practice.

The second shift is toward stronger theory-to-practice connections, with rehearsal opportunities. A new course ("Inquiry and Puzzles of Practice") provides a structured safe space for candidates to engage in their own inquiries. Within this course, candidates will become familiar with the rubric, use it to guide curriculum design, and practice with it as they observe each other. This course will continue throughout the program so as candidates grow in their practice and gain experience with the rubric, they will be able to internalize what "good teaching" looks like and self-monitor development across different competencies.

The third shift is toward building authentic and substantive partnerships with schools and school districts. The collaborative rubric work has fostered deeper relationships between the two IHEs and three partner districts based on shared vision and trust. The rubric was developed by synthesizing existing district observation tools, and aligning this synthesis to the new state standards. Because the school districts have committed to using the rubric through induction for new teachers, program completers will have a more seamless entry into the local workforce (NCTR, 2015). The rubric will support not only better prepared first year teachers but also new teacher retention by easing the preservice-to-inservice transition.

In order for feedback to bridge the worlds of preservice and inservice, it must be relevant to the intersecting standards and goals of EPPs and partnering districts. Maintaining relevance to a dynamic policy landscape can be challenging for EPPs as state and national standards for curriculum, preservice teacher performance, and inservice teacher performance continue to evolve. Partnering with school districts can help EPPs remain relevant to practitioners and smooth the transition into the field for novice teachers. For school districts, partnering with EPPs can provide access to resources and cutting-edge evidence-based practices. Frequently, school and district administrators struggle to operationalize "good teaching" and to provide meaningful feedback to classroom teachers that will help improve their practice in non-threatening ways. The collaborative process described in this case story resulted in development of an observation feedback instrument that explicitly bridges the preservice-to-inservice transition, aligns to both preservice and inservice teacher state standards, operationalizes evidence-based pedagogical practices, and provides consistent language for action-oriented constructive feedback throughout the novice to expert continuum.

Addressing Puzzles of Practice

Rubric development emerged as a solution that could address multiple "puzzles of practice" for the case EPP which leveraged these puzzles to initiate conversations with regional partners. The EPP was facing newly revised preservice state standards, a need to update candidates' clinical experiences, and a campus initiative to reduce fees for students -- all while remaining responsive to regional needs. In response to these puzzles of practice, the case EPP chose to locally develop a standards-aligned, research-based teaching observation instrument for providing meaningful feedback to candidates in the field.

To develop as reflective practitioners, preservice candidates require consistent high-quality, standards-based feedback and opportunities for facilitated next-steps planning that builds toward effective instruction. Many traditional models of teacher preparation privilege coursework and position fieldwork as a place where candidates have the option to try out coursework learning. Recent critiques of such traditional models call for the clinical experience to be the central component, with prolonged time in the field, multiple opportunities to practice in authentic contexts, and intentional guidance or scaffolding (Grossman, 2010). Various feedback instruments are available; however, most are aligned to national rather than state standards (e.g., INTASC), and they frequently require usage fees that are passed along to candidates (e.g., Danielson, 2013). Such fees are problematic for regions with high

poverty rates and EPPs with food-insecure candidates (Jargowsky, 2015). To address these puzzles related to state-specific standards alignment and equitable access for teacher candidates, and based on recommendations in the extant literature, the EPP developed the Continuum of Reflective, Engaging, and Accessible Teaching Rubric (CREATe Rubric) in partnership with another EPP and three school districts.

This case story describes the collaborative process of these two EPPs and three school districts in developing a common rubric to provide action-oriented formative feedback for new teachers while they are teaching in the field. The rubric focuses on specific action-oriented competencies aligned to both district and state standards, and provides a "common language," vetted by multiple stakeholders, supporting consistency and helping to prevent mixed messages or misinterpretations of jargon. Using the rubric, various constituents can provide specific feedback and next steps to strengthen practice for any teacher along the novice to expert continuum. Furthermore, because each competency is represented in the rubric as a separate continuum, the rubric can help identify each teacher's individual zone of proximal development (Vygotsky, 1978) and help teacher educators provide tailored and targeted scaffolding. Thus, the rubric provides a way to create practice profiles across multiple competencies that can help document teacher development over time.

Collaborative Partners and Context

The CREATe Rubric was developed over a period of 18 months in collaboration with a second EPP (EPP-2) and three of the major area school districts. The focal EPP leveraged existing relationships to engage program coordinators from both EPPs and high-level district administrators from the three districts in this collaborative process. The focal EPP is a regional comprehensive university in the western United States and is one of the largest EPPs in the state. The region that the EPP serves is lingustically linguistically and culturally diverse but is challenged by high poverty rates (25.5%) and low educational attainment (19.7% have a BA or higher) (FC Economic Development Corporation, 2017). CREATe Rubric development was launched in concert with other initiatives and innovations at the focal EPP, which were all part of a comprehensive program revision and rebranding effort. Incorporating rubric development into a larger strategic plan and vision for the EPP allowed the focal EPP to use resources and leverage existing relationships in creative ways to accomplish the work.

Rubric Development

Rubric development began with close examination of the new state standards. The focal EPP first met with each school district team and the EPP-2 team separately to carefully review the new preservice standards. Each team analyzed the six standards and 45 substandards to identify those that are critical for novice teachers and could be directly observed in a classroom setting. The selected observable standards were compared and synthesized across teams, with nearly 100% consensus across all constituents regarding which observable standards should be represented on the rubric. From this analysis, 14 rubric items were developed, aligned to 17 preservice substandards.

Next, extant district inservice rubrics were synthesized and incorporated into CREATe so that it could be aligned and used across districts as a continuum of development from novice to expert teacher. Different districts' inservice rubrics varied in regard to complexity and explicitness of how inservice standards were operationalized. These three district inservice rubrics were coded and synthesized to develop evidence-based language for common "look-fors" as descriptive anchors in the more advanced performance categories of CREATe. Based on the analysis and synthesis of the three inservice district rubrics, seven performance categories were developed for CREATe: five (unobserved, attempting, exploring, emerging, developing) spanning the expected developmental trajectory of preservice teachers and two (skillful,

masterful) extending teacher development into inservice practice. Integration of the district inservice rubrics into CREATe explicitly bridges the instrument from preservice to inservice.

Once the structure of CREATe was built, the most recent literature and evidence in teacher preparation was consulted in operationalizing the practices and strategies associated with culturally sustaining pedagogy (Paris, 2012), developmentally appropriate practices (Bredekamp, 1997), and universal design for learning (Meyer, Rose, & Gordon, 2014; Hall, Meyer, & Rose, 2012; Meyer, Rose, & Gordon, 2014). These three frameworks were operationalized based on evidence-based practices, theoretical foundations, and language woven throughout the preservice and inservice state standards. Research on rubric development and testing was also consulted during initial development. For example, in a comprehensive review of published empirical research on rubrics Jonsson and Svingby (2007) outlined five major considerations of reliable rubrics:

1. Benchmarks increase agreement, but should be chosen with care (Denner, Salzman, & Harris, 2002; Popp, Ryan, Thompson, & Behrens, 2003);

2. Analytical scoring is preferable (Johnson, Penny, & Gordon, 2000; Johnson, Penny, & Gordon, 2001; Penny, Johnson, & Gordon, 2000a; Penny, Johnson, & Gordon, 2000b);

3. Training is necessary but not sufficient for inter-rater agreement (Stuhlmann, Daniel, Dellinger, Denny, & Powers, 1999; Weigle, 1999);

4. Topic-specific rubrics are more generalizable and dependable than generic rubrics (DeRemer, 1998; Marzano, 2002);

5. Careful consideration should go into the number of levels on the rating scale.

These research-based recommendations were used as guidelines to ensure development of a reliable and valid instrument.

Rubric Structure

The CREATe Rubric includes 14 items, aligned to preservice and inservice state standards. The 14 items are organized into four sections: Positive Environment, Instructional Design and Implementation, Rigorous and Appropriate Content, and Reflection-In-Action. Each of the 14 items is rated along a seven-point developmental continuum with the following rating categories: Unobserved, Attempting, Exploring, Emerging, Developing, Skillful, and Masterful. Each rating category has an anchor descriptor that operationalizes each of the 14 items with action-oriented, observable "look-fors." Table 1 summarizes the 14 items.

District partners and EPPs participated in initial CREATe testing and refining through joint classroom visits and debriefing sessions. District administrators, EPP teacher educators, and classroom teachers completed surveys to provide expert feedback on each item. After this initial round of field testing in Spring 2017, each rubric item was revised based on feedback data. During Fall 2017, the CREATe Rubric was piloted on a small scale, with data and feedback loops regarding training, calibration process, and usability (including time, costs, challenges). The CREATe Rubric is currently undergoing reliability and validity testing.

Rubric Testing

CREATe is currently undergoing a rigorous testing and vetting process. Multiple sources of data are being used to evaluate the validity, reliability, and practice-based value of CREATe. In the initial design, construct validity was addressed through alignment of the rubric anchor descriptors to theory, evidence-based practices, local district standards, and state standards. Internal consistency will be explored using Cronbach's Alpha and exploratory factor analysis. Content validity can be addressed through careful selection of items (Anastasi & Urbina, 1997) as well as expert feedback. For every item on the rubric, experts (e.g., teacher educators, school administrators, classroom teachers, fieldwork coaches) complete a short researcher-developed feedback survey (Foxcroft, Paterson, le Roux, & Herbst, 2004) to

rate: 1) extent to which the item is observable, 2) degree of alignment between rubric item/indicator and standard, 3) degree of face validity, and 4) extent to which the item and indicator reflect positive teaching practices. Criterion validity will be examined through comparing candidate scores on rubric items with qualitative observation notes, course grades, Teacher Performance Assessment scores, and post-completion employer ratings. Inter-rater reliability is being established through rigorous initial training and paired observation sessions, comparing observer rubric ratings with those of the rubric developers. Observers will be required to recertify annually. Throughout the validation process, rubric data usability will be examined for tracking individual candidate growth, providing targeted just-in-time scaffolds for candidates, fostering reflective practice in candidates, and tracking overall program efficacy. In addition to previously mentioned comparisons, usability analyses will utilize data from semi-annual focus groups, candidate exit surveys, and district administrator feedback.

Implications

Decontextualized, summative teacher observation protocols do not foster engagement in cycles of reflective practice and continuous improvement. Development and adoption of the CREATe Rubric across EPPs and partner districts has the potential to support evidence-based practices at multiple levels for multiple constituents (i.e., candidate, coaches, program, regional) and will provide both qualitative and quantitative evidence to inform practice on a system-, and potentially, statewide level. CREATe Rubric adoption will provide data for monitoring progress of candidates and for guiding EPP continuous improvement efforts.

In addition to the innovation of the rubric itself, the development process has deepened the level of collaboration and understanding between EPP and district stakeholders. In developing, testing, revising, and validating the CREATe Rubric, it has been necessary to have difficult conversations around pedagogy, equity, curriculum, and standards. These conversations have fostered growth and new understanding for all involved parties. District stakeholders have had to acknowledge that every teacher (even star veteran teachers) has potential to grow; EPP stakeholders have had to recognize the need to more substantively model and require rigor, pedagogical content knowledge, and inclusive culturally sustaining pedagogical practices. Future collaborative work also includes alignment of CREATe to other standards and frameworks (e.g., NGSS; INTASC) to increase usability.

Development of the rubric and buy-in of partner districts and EPPs to utilize it are examples of the kinds of work that can be accomplished through the research-based practice of working in partnership with the larger community. This case story demonstrates an example of how collaboration between EPPs and districts can result in powerful initiatives that can inform the field. By sharing our work in progress, we hope to encourage other IHEs to consider the potential impact of similar collaborations within their own contexts.

Table 1
CREATe Rubric Items and Item Description

Section & Item	Item Description
A. Positive Environment	
1: Caring Community	Implementation of systems and routines for behavior intervention, restorative justice, and conflict resolution
2: Inclusive Learning Environment	Classroom climate including considerations regarding physical safety, psychological safety, tone, peer interactions, and inclusive practices
3: High Expectations	Classroom includes scaffolded high expectations for all learners, opportunities for students to practice self-regulation, differentiated instruction based on the learners in the context.
4: Positive Behavior Expectations	Routines, procedures, and norms established and maintained in the classroom
5: Reflection, Assessment, & Self-Assessment	A culture of reflective practice, which includes self-assessment, teacher assessment with high quality feedback, opportunities for revisions based on feedback, a culture of intellectual risk-taking, and encouragement of metacognitive thinking about engagement, learning, and behavior.
6: Funds of Knowledge	Familiarity with and leveraging of students' funds of knowledge
B. Instructional Design & Implementation	
7: Student Motivation, Engagement, & Active Learning	Engagement and motivation of students through design and implementation of a range of instructional activities that incorporate the needs and interests of the learners in the classroom.
8: Varied Strategies	Use of multiple developmentally- appropriate strategies and resources, technology, principles of UDL, tiered support and scaffolds, and differentiation
9: Research-based Instruction for Emergent Bilinguals and Students with Special Needs	Use of research-based instructional approaches to provide a supportive learning environment for first and/or second language acquisition and student with special needs.
C. Rigorous & Appropriate Content	
10: Critical & Creative Thinking	Opportunities for students to engage in critical and creative thinking
11: Subject Matter Knowledge	Teacher application of subject matter knowledge, standards, and frameworks
12: Content Accessibility	Supports and scaffolds to ensure curriculum and content are accessible to all students
13: Interdisciplinary Integration	Subject specific pedagogy, integration across content areas, and incorporation of visual and performing arts
D. Reflection-in-Action	
14: Monitoring Student Learning & Adjusting Instruction	Teacher monitoring of student learning, instructional modifications, and differentiation

Item 5: Student Reflection, Assessment, & Self-Assessment

5.3 Involve all students in **self-assessment and reflection** on their learning goals and progress and provide students with **opportunities to revise or reframe their work based on assessment feedback.**

Unobserved: Not yet evident	**Attempting: Aware, may not be effective**	**Exploring: Attempting, minimally effective**	**Emerging: Consistently attempting, limited effectiveness**	**Developing: Consistently attempting, somewhat effective**	Skillful:	Masterful:
Students have no opportunity to self-assess and reflect on their own progress Students are not given opportunities to revise work No assessment feedback provided, or feedback provided but is not the basis for student revision **Mistakes and/or intellectual risk-taking penalized** Students not encouraged to think metacognitively about their engagement, learning, and behaviors	One observed opportunity for at least one student to self-assess and/or reflect on progress Students might have one opportunity to work, with teacher direction student self-revision Mistakes and/or intellectual risk-taking not penalized Students might be encouraged once to think metacognitively about their own engagement, learning, and behaviors	Some students have opportunity and are guided to self-assess and reflect on progress Specific students given some opportunities to revise work, with teacher direction Mistakes and/or intellectual risk-taking positively acknowledged and recognized Some students minimally encouraged to think metacognitively about their engagement, learning, and behaviors	All students have opportunity and/or guidance to self-assess and reflect on progress All students given opportunity to revise work, with teacher direction Mistakes and/or intellectual risk-taking encouraged All students encouraged to think metacognitively about their engagement, learning, and behaviors	Students formally guided to self-assess and reflect on progress All students given opportunities to revise work, using their own ideas revision Mistakes and/or intellectual risk-taking encouraged and valued All students encouraged more than once to think metacognitively about their engagement, learning, and behaviors	Students formally guided to set goals, self-assess, and reflect on progress All students given opportunities to revise work, using their own ideas and/or learning goals revision Mistakes and/or intellectual risk-taking encouraged and valued, and modeled by the teacher All students encouraged frequently to think metacognitively about their engagement, learning, and behaviors	Systematic, frequent, opportunities for student self-assessment, goal setting, and monitoring progress All students given multiple opportunities to revise work, using their own ideas and goals connected to student's own goals, and used for revisions Mistakes and/or intellectual risk-taking expected for all students as well as the teacher Students initiate thinking metacognitively about their engagement, learning, and behaviors

Figure 1 Sample Item from CREATe Rubric (Yun & Bennett, 2017)

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing (7th ed.)*. Upper Saddle River, NJ: Prentice-Hall.

Bredekamp, S. (1997). NAEYC Issues Revised Position Statement on Developmentally Appropriate Practice in Early Childhood Programs. *Young Children, 52*(2), 34-40. Retrieved from http://www.jstor.org/stable/42727311

Danielson, C. (2013). *The framework for teaching: Evaluation instrument*. Princeton, NJ: Danielson Group.

Denner, P. R., Salzman, S. A., & Harris, L. B. (2002, Feb.). *Teacher work sample assessment: An accountability method that moves beyond teacher testing to the impact of teacher performance on student learning.* Paper presented at the 54th Annual Meeting of the American Association of Colleges for Teacher Education, New York, NY.

DeRemer, M. L. (1998). Writing assessment: Raters' elaboration of the rating task. *Assessing Writing, 5*(1), 7-29.

Foxcroft, C., Paterson, H., Le Roux, N., & Herbst, D. (2004, July). *Psychological assessment in South Africa: Aa needs analysis: the test use patterns and needs of psychological assessment practitioners: final report.* Retrieved from http://repository.hsrc.ac.za/bitstream/handle/20.500.11910/7498/1716_Foxcroft_Psychological assessmentin%20SA.pdf?sequence=1

Grossman, P. (2010). *Learning to practice: The design of clinical experience in teacher preparation*. Washington DC: AACTE. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=825DA42BB0721BACFC03ADC3AEF4F7F9?doi=10.1.1.178.4088&rep=rep1&type=pdf

Hall, T., Meyer, A., & Rose, D. (2012). U*niversal design for learning in the classroom: Practical applications.* New York, NY: Guilford Press.

Jargowsky, P. A. (2015). *Architecture of segregation: Civil unrest, the concentration of poverty, and public policy.* Washington DC: The Century Foundation. Retrieved from https://tcf.org/assets/downloads/Jargowsky_ArchitectureofSegregation.pdf

Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. A*pplied Measurement in Education, 13*, 121-138.

Johnson, R. L., Penny, J., & Gordon, B. (2001). Score resolution and the interrater reliability of holistic scores in rating essays. *Written Communication, 18*, 229-249.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2,* 130-144.

Marzano, R. J. (2002). A comparison of selected methods of scoring classroom assessments. *Applied Measurement in Education, 15*, 249-268.

Meyer, A., Rose, D. H., & Gordon, D. (2014). *Universal design for learning: Theory and practice.* Wakefield, MA: CAST.

National Center for Teacher Residencies (NCTR). (2015). *Clinically oriented teacher preparation.* Washington DC: NCTR. Retrieved from https://nctresidencies.org/wp-content/uploads/2015/07/NCTR-COTP-Final-Single-Pgs.pdf

National Research Council (NRC). (2010). *Preparing teachers: Building evidence for sound policy.* Report by the Committee on the study of teacher preparation programs in the United States. Washington, DC: National Academies Press.

Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher, 41*, 93-97.

Penny,Penny, J., Johnson, R. L., & Gordon, B. (2000). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing, 7*, 143-164.

Penny, J., Johnson, R. L., & Gordon, B. (2000). Using rating augmentation to expand the scale of an analytic rubric. *The Journal of Experimental Education, 68*, 269-287.

Popp, S. E. O., Ryan, J. M., Thompson, M. S., & Behrens, J. T. (2003, April). *Operationalizing the Rubric: The Effect of Benchmark Selection on the Assessed Quality of Writing*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Stuhlmann, J., Daniel, C., Dellinger, A., Kenton, R., & Powers, T. (1999). A generalizability study of the effects of training on teachers' abilities to rate children's writing using a rubric. *Reading Psychology, 20*, 107-127.

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children, 23,* 34-41.

Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing, 6,* 145-178.

Yun, C., & Bennett, L. H. (2017). *Continuum of Reflective, Engaging, and Accessible Teaching (CREATe) Rubric* [Rubric and manual]. Unpublished instrument. Fresno, CA: Fresno State.