CDE Educator Evaluation Research Activities
(8/16/2013)

2010-11: Rubric Development	**Development of:** • Rubrics for Teachers and Principals • Initiation of Resource Bank • Pilot test activities (Pilot, integration, partner, rogue) **Provide:** • Training for pilot districts		
Research Activities	**Details of Activity**	**Deadline**	**Comments**
Determine face validity, preliminary content validity, credibility /relevance, and usability of early drafts of rubrics.			

Conduct a variety of formal and informal feedback sessions designed to inform the field of the work of the EE unit and invite them to continue to communicate about their ideas, concerns, and things they thought would work well.	• Interviews
	 • Focus Groups
	 • Feedback through CDE website
	 • Discussions with professional organizations
	    o ACEE
	    o CEA
	    o CASB
	    o CASE
	    o Summer Institutes
	    o HR Directors
	    o EE Staff members
	 • Revision of Rubrics following data collection
		8/31/11	Draft rubric used in 2011-12 was based on the results of this process.

2011-12: Development and Beta Testing	Development of: • System • Rubrics • Technical Guidelines on Professional Practices • State data collection and monitoring system • Tools for district implementation system (Excel spreadsheets and Bloomboard)	Populate and Launch Resource Bank Beta-Testing: • Rubrics • Tools Provide: • Differentiated support to districts

Research Activities	Details of Activity	Deadline	Comments
Determine face validity, preliminary content validity, credibility/relevance, and usability of early drafts of rubrics.			

Preliminary content validity, face validity, credibility, fairness and usability of revised instruments	• Interviews
	• Focus Groups
	• Feedback through CDE website
	• Discussions with professional organizations
	  o ACEE
	  o CEA
	  o CASB
	  o CASE
	  o HR Directors
		6/30/12	Assistance from CLF staff members

Inter-Rater Agreement Training	-To determine whether MET videos may be used to determine inter-rater agreement -Development and pilot testing of IRA process and training for evaluators -Identification of problems and potential problems with rubric and system with respect to IRA	8/31/12	Financial and other supports from CLF
2011-12 Pilot study on principal rubric - Develop preliminary data analysis strategies and test them on principal evaluation data collected during 2011-12.	-Pilot test of data collection and analysis strategies; preliminary look at evaluation results -Collect and analyze data from districts: • Self-assessment ratings • Mid-year review progress reports, when available • End-of-year ratings • Survey data • Matched with other state data (accountability measures, student demographics, educator demographics, etc.)	8/1/12	

<table>
<tr>
<td rowspan="2" style="background-color:brown; color:white">2012-13: Pilot and Rollout</td>
<td>
Development of:

SSP Rubrics
Technical Guidelines on Student Growth
Student growth tools for district implementation

Populate and Launch Assessment Resource Bank

Beta-Testing:

Student growth tools

Provide:

Training of all non-pilot districts
Resources

</td>
<td>

Tools
Informal supports
Convenings to share lessons learned
Recommendations on Specialized Service Professionals standards and elements
Rubrics for teachers and principals Raters recruited or available to conduct observations
Rater trainings
Certification process (if appropriate)
Scoring design (including student growth)[1]
Number, frequency, and length of observations and number of raters for each
Recommended annual timeline for completing components

</td>
</tr>
</table>

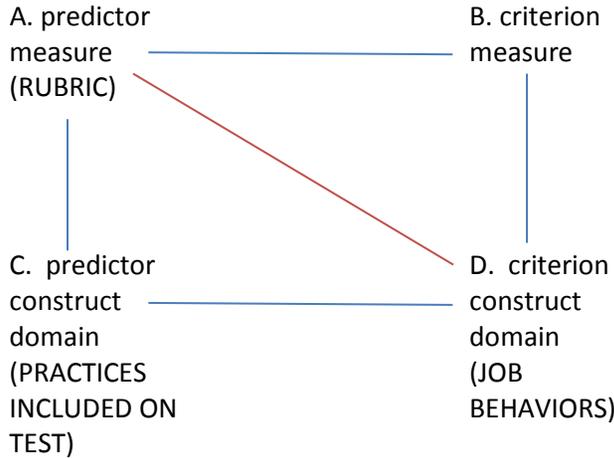| Research Activities	Details of Activity	Deadline	Comments
Finalize System Components			
Usability study of teacher and principal rubrics	Interviews and Focus Groups with Pilot and Integration Sites: • Integration Summit • CEA • CASE • NMI • Webinars • Other Survey Data	April - June 2013	To determine • Utility, fairness, user-friendliness of instruments • Changes necessary to improve quality of all materials • Viability of proposed changes to materials and processes

---

[1] Hill, H. C., Charalambous, C. Y, and Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher*, Vol. 41 No. 2, pp. 56-64.

Research Activities	Details of Activity	Deadline	Comments
2012-13 Pilot study on teacher and principal rubrics (preliminary validation study). Analysis of pilot district data in order to: - Provide preliminary evidence for fairness (e.g., rubrics are equally applicable for different teacher types) - Provide preliminary analysis for reliability (e.g., ratings are highly correlated with each other) - Provide preliminary analysis for validity (e.g., ratings are highly correlated with other measures of classroom success) - Guide revisions of teacher and principal rubrics and users' guides	Collect and analyze pilot data: • Self-assessment ratings (from districts) • Mid-year review progress reports, when available (from districts) • End-of-year ratings (from districts) • Student perception survey data (from CLF) • Teacher perception survey data (from CLF) • Teacher, principal, and superintendent feedback survey data (from districts) • TELL data (from NTC) • Student demographics (from CDE October files) • Educator demographics (from CDE HR files) • School characteristics (from CDE school files) • District characteristics (from CDE district files) • Assessment data (from CDE assessment files and integration districts [when student-teacher links are required]) • SPF data (from CDE files) Specific analyses: • Analyses at all levels: check box, element, standard, and overall • Distributions, including sub-groups • Descriptives, including sub-groups • Correlations, within and between elements and standards, and with other measures • Factor analysis on check boxes (within elements), elements (within standards), and standards (within overall)	7/1/13 for high-priority teacher rubric analyses Remaining analyses are ongoing	
Determination of critical issues	Critical review of findings from 2012-13 qualitative and quantitative data to determine necessary changes to the process and/or materials.	8/1/13	
Review and revise teacher and principal rubrics and users' guide	Apply findings from qualitative and quantitative data to the revision of all materials associated with the CO State Model Evaluation System for teachers and principals.	8/16/13	

Research Activities	Details of Activity	Deadline	Comments
Study to determine whether a 4-point or 5-point rubric is most appropriate	This study is to determine whether the CO rubric contains the appropriate number of rating levels.	2/15/13	
Develop a new decision matrix/framework	Come up with a more precise alternative to the SCEE decision matrix, make recommendations to CDE leadership	5/1/13	
Student growth analyses	Conduct preliminary analyses related to: • Assessments that could be used for educator effectiveness • Growth models on those assessments, when student-teacher links are available • Other student growth components such as SPF	Ongoing	

<table>
<tr>
<td rowspan="2">2013-14: Pilot and Rollout</td>
<td>Provide:
• Statewide TA on rollout of teacher/principal systems
• Support to ALL districts through resources, trainings, tools, etc.
• Convening of pilot districts to share lessons learned

Develop:
• Evaluation system for Other Licensed Personnel/Specialized Service Professionals
• Criteria for evaluator training courses for approval</td>
<td>Finalize System Components:
• Rubrics for Specialized Service Professionals
• Raters recruited or available to conduct observations
• Rater trainings
• Certification process (if appropriate)
• Scoring design (including changes to process, student outcomes and determination of effectiveness ratings)
• Number, frequency, and length of observations and number of raters for each</td>
</tr>
</table>

Research Activities	Details of Activity	Deadline	Comments
2013-14 Validation study - Analysis of pilot district data in order to: - Provide preliminary evidence for fairness (e.g., rubrics are equally applicable for different teacher types) - Provide preliminary analysis for reliability (e.g., ratings are highly correlated with each other) - Provide preliminary analysis for validity (e.g., ratings are highly correlated with other measures of classroom success) - Guide revisions of teacher and principal rubrics and users' guides	Collect and analyze pilot data: • Self-assessment ratings (from districts) • Mid-year review progress reports, when available (from districts) • End-of-year ratings (from districts) • Student perception survey data (from CLF) • Teacher perception survey data (from CLF) • Teacher, principal, and superintendent feedback survey data (from districts) • TELL data (from NTC) • Student demographics (from CDE October files) • Educator demographics (from CDE HR files) • School characteristics (from CDE school files) • District characteristics (from CDE district files) • Assessment data (from CDE assessment files and integration districts [when student-teacher links are required]) • SPF data (from CDE files)	7/1/14 for high-priority teacher rubric analyses Remaining analyses are ongoing	

Research Activities	Details of Activity	Deadline	Comments
Validate teacher and principal scores by testing the relationships specified in the diagram to the right (Source: *Standards for Educational and Psychological Testing (1999).* American Psychological Association, pp. 153-155.) Determine the strength of the relationships in the diagram above by answering the following questions: 1. A to B: Are predictor scores related to scores on the criterion measure? 2. A to C: Does the measure represent the educator's standing on the domain of interest? 3. C TO D: Are the predictor and criterion domains related? 4. B to D: Does the measure represent the educator's standing on the domain of interest? A to D: Establishing this relationship is the "hallmark" of content validity. This linkage is established by designing predictor measures (rubric) as samples of the criterion construct domain.	Specific analyses: • Analyses at all levels: check box, element, standard, and overall • Distributions, including sub-groups • Descriptives, including sub-groups • Correlations, within and between elements and standards, and with other measures • Factor analysis on check boxes (within elements), elements (within standards), and standards (within overall) -- A. predictor measure (RUBRIC) B. criterion measure C. predictor construct domain (PRACTICES INCLUDED ON TEST) D. criterion construct domain (JOB BEHAVIORS) 1. Measured through empirical examination of the relationship. 2. Logical analysis, expert judgment, and convergence with or divergence from conceptually similar or different measures. 3. Theoretical and logical analysis. 4. Logical analysis, expert judgment, and convergence with or divergence from conceptually similar or different measures. 5. May be established by pairing linkages 1 and 4 or linking 2 and 3. May also be established directly if the predictor measure is designed as a sample of the criterion construct domain.		

Research Activities	Details of Activity	Deadline	Comments
Analyze statewide data and make adjustments to the system as needed	Analyze data from CDE HR collection (districts submit 2013-14 data in December 2014 and these data are cleaned and available in April 2015): • End-of-year ratings at Standard level only • Matched with other state data (accountability measures, student demographics, educator demographics, etc.)	6/1/15	
Inter-rater agreement on teacher and principal rubric ratings	Determine specifications for certifying that inter-rater agreement is sufficient for evaluators.	3/1/14	
Student growth analyses	Conduct preliminary analyses related to: • Assessments that could be used for educator effectiveness • Growth models on those assessments, when student-teacher links are available • Other student growth components such as SPF	Ongoing	
Generalizability study	This study will attempt to partition the variance in teacher and principal scores into different components, (e.g., teacher, lessons, evaluators), their interactions, and measurement error. This will provide information about how to identify optimal numbers and types of observations and scoring designs.	Following completion of validation study	This type of study will eliminate the uncertainty surrounding sampling of lessons, differences among raters, and other factors that affect observation ratings.
Simulation studies	• Determination of whether the number of items can be shortened without affecting performance and consequently the characteristics of resulting teacher scores. • Are there key clusters of items that consistently provide the same ratings for teachers as using the entire rubric? What information will be lost if certain practices are eliminated?	Following completion of validation study	This strategy will enable CDE to use validity study data for multiple purposes and to answer some key research questions.

Research Activities	Details of Activity	Deadline	Comments

Pilot test Specialized Service Professionals rubrics and processes, including examining content validity	• Identify districts to pilot instruments
	   o Invitations in EE newsletter
	   o Individual communications with pilot, integration, partner, and non-pilot districts
	   o Individual communications with districts who provided work group members
		8/15/13	

	• Notify pilot sites of requirements
	   o Self-Assessment and final ratings
	   o Participation in webinars, surveys, interviews, and/or focus groups to share feedback on the instruments and process.
		9/1/13	

	• Schedule feedback sessions and notify pilot sites of schedule
	   o Professional organization meetings
	   o Regularly scheduled state-level meetings
	   o Unit meetings within CDE
	   o Specially scheduled meetings with districts or collections of districts and BOCES
	   o Webinars
	   o Surveys
	   o Others
		Ongoing	

	• Collect feedback data through CDE/EE website: Interviews Focus Groups Feedback through CDE website Surveys Discussions with professional organizations	April – June 2014	To determine : • Utility, fairness, user-friendliness of instruments • Changes necessary to improve quality of all materials • Viability of proposed changes to materials and processes

2014-15: Pilot and Rollout	**Provide:** • Support to districts via resources and training **Finalize:** • Statewide implementation of teacher/principal systems • Recommendations to SBE for this year and all following years for CI

Research Activities	Details of Activity	When By Whom	Comments
Validate Specialized Service Professional rubrics			
Analyze data and make adjustments as needed			
Implementation Fidelity Studies			