Refining Evaluation Tools for Evidence-Based Professional Development

Meet the team

Kassie Haan



Communication Liaison kassie.haan@du.edu



Cecilia Ball

Research Lead

Kate Gladson
Harbaugh



Worksheet & Rubric Analyst



Mariana Romero Chong

Revisions & Recommendations Lead

Key Takeaways

The current SPDG rubric includes redundancy, ambiguous language, and limited alignment with research. Our revisions streamline content, clarify expectations, and strengthen the rubric's foundation in evidence -based practices.

Methods Overview

		DONE	NOT
Reviewed	existing SPDG rubrics and worksheets	√	
Analyzed	real grantee responses for redundancy/confusion	√	
Conducted literature review		√	
Collaborated on iterative edits and revisions		√	
Integrated	feedback from OSERS and each other		√

Research Foundations: Professional Learning and Adult Learning

Learning Forward Standards (2022)

- Equity-centered, job-embedded, and aligned with system goals.
- Organized into 3 categories:
 - Rigorous Content (curriculum-aligned)
 - Transformational Processes (collaborative, active learning)
 - o Conditions for Success (resources, leadership, systems)
- Backed by meta-analysis linking standards to improved outcomes.

Dunst & Trivette (2012)

- Most effective PD uses 4-5 active strategies over 20+hours.
- Key methods: modeling, coaching, reflection, self-assessment
- Short lectures = minimal impact; real-world practice = sustain change.

Research Foundations: Implementation Science & Systems Alignment

Fixsen et al. (2005)

- Effective implementation requires:
 - Competency Drivers: coaching, training
 - Organization Drivers: admin support, data systems
 - Leadership Drivers: adaptive leadership
- Fidelity, feedback loops, and stage-based support are essential
- Systems must build capacity, not just adopt practices.

ESSA-aligned PD (Learning Forward, n.d.)

- High Quality PD is:
 - Sustained (not one-off)
 - Collaborative &data-driven
 - o Focused on student outcomes & classroom relevance.

Key Findings

Redundancy

Redundancy across domains created confusion and over -reporting

Language

Ambiguous or jargon -heavy language led to scoring inconsistencies and grantee confusion

Research Alignment

Limited research alignment with best practices in PD and implementation science

Timeline

April 22 –26 Finalized scope with SPDG and gathered project materials

April 29 – May 3 Conducted literature review and began rubric analysis

May 6–10 Analyzed submissions for redundancy and missing elements

May 13–17 Drafted individual rubric edits and rationale

May 20 - 24 Finalized revisions, created comparisons, and built slide deck

May 27–31 Sent materials to SPDG for input and made final revisions for presentation

June 4 Present final recommendations to SPDG

June 25 Final report to SPDG

Logic Model

Inputs	Rubric versions, worksheet, SPDG guidance, grantee submissions, research
Activities:	Document analysis, rubric editing, literature integration, logic modeling, presentation prep
Outputs: Revised rubric & worksheet, change comparison, recommendations report, presentation	
Outcomes: Short-term – Improved clarity & reduced confusion for grantees Medium-term – More consistent data & reduced redundancy Long-term – Stronger EB-PD implementation & improved support for	

Rubric Revisions We Made

Replaced vague descriptions with clear, observable expectations

Added verification tools like resumes, MOUs, and fidelity rubrics

Used plain-language bullets instead of dense paragraph text

Included score 4 exemplars and balanced trainer/coach roles

Integrated citations and models like Dunst & Trivette and Guskey throughout

Example Before & After: B(5) — Trainer Coaching, Observation, and Evaluation

Description of training provided to trainers	Description of the training provided to trainers, including duration and content focus	
Description of coaching provided to trainers	Description of the structure and frequency of coaching provided to trainers (e.g., individual, group sessions)	
Description of procedures for observing trainers	Description of the procedures used to observe trainers, including the fidelity instrument used	
(Not present)	Description of procedures to obtain participant feedback	
Description of how observation and training fidelity data were used	Description of how observation and training fidelity data were used (e.g., to determine if changes should be made to the content or structure of trainings, such as schedule, processes; to ensure that trainers are qualified).	

Recommendations

Create clear "score 4" Help grantees better self -evaluate and **Provide Annotated Examples** exemplars with side -by-side improve consistency comparisons Offer guidance through Clarify common misinterpretations and **Develop Reviewer & Grantee Training** webinars or how -to rating challenges documents Pilot the Revised Rubric Before Full Test usability and scoring Use feedback to refine before statewide reliability with a small cohort Rollout or national use

Address Self-Scoring Bias



Acknowledge the risk of both over- and under-rating due to misunderstanding rubric criteria or organizational culture



Recommend validation strategies, such as reflection checklists, calibration sessions, or prompts embedded in the rubric, to improve scoring accuracy

First Recommendation

Action/	Means/	Outcome/
Recommendation	Implementation	Intended Impact
Provide Annotated Examples	Create clear "score 4" exemplars with side -by- side comparisons	Help grantees better self - evaluate and improve consistency

Example from B(5):

Applicants who scored a 4 typically provide:

- A clear defined training plan for trainers
- Structured and frequent coaching
- Description of how trainers are observed using fidelity tool
- A process for collecting and using participant feedback
- Clear expectations of how data were used to revise delivery and provide targeted trainer support

Second and Third Recommendation

Action/Recommendation	Means/Implementation	Outcome/Intended Impact
Develop Reviewer & Grantee Training	Offer guidance through webinars or how -to documents	Clarify common misinterpretations and rating challenges
Pilot the Revised Rubric Before Full Rollout	Test usability and scoring reliability with a small cohort	Use feedback to refine before statewide or national use

Fourth Recommendation

Action/Recommendation	Means/Implementation	Outcome/Intended Impact
Address Self-Scoring Bias	Acknowledge the risk of both over- and under-rating due to misunderstanding rubric criteria or organizational culture	Recommend validation strategies, such as reflection checklists, calibration sessions, or prompts embedded in the rubric, to improve scoring accuracy

Final Steps & SPDG Input

What We're Asking From SPDG

Final feedback on rubric clarity and structure Suggestions for any last revisions

The Team's Next Steps

June 4 (today): Present final recommendations to SPDG

June 4-11:Edits and finalizing report

June 25: Submit final written report to SPDG

(See Slide 5 for full project timeline)

References:

- Dunst, C. J., & Trivette, C. M. (2012). Moderators of the effectiveness of adult learning method practices. *Journal of Social Sciences*, 8(2), 143–148.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature* (FMHI Publication #231). University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network.
- Learning Forward. (2022). Standards for professional learning.
- Learning Forward. (n.d.). Definition of professional development.
 Powered by Title II. Retrieved May 28, 2025.