

Welcome to Our Winter Edition

By Paula Arce-Trigatti | NNERPP



Welcome to the final edition of Volume 3 of NNERPP Extra, and thank you for joining us for this third year! We round out the year with a special issue focused on research-practice partnership (RPP) evaluation, an enduringly important topic to RPPs: As you'll see below, all articles in this edition examine different aspects of RPP evaluation, ranging from research

approaches one can use to develop measures of RPP effectiveness to exploring the role of the RPP evaluator and highlighting a new tool you can use to measure RPP health.

More specifically, in this edition, you'll find:

- **Research Insights:** We take a closer look at how a collaboration between the National Center for Research in Policy and Practice (NCRPP) and NNERPP is using Evidence-Centered Design to develop measures of RPP effectiveness.
- **RPP Deep Dive:** Our colleagues from the Computer Science for All space examine the emerging role of the external RPP evaluator.
- **Extra Credit:** We hear about a new tool for those working in partnerships to formatively assess and continuously improve their RPP.
- **Improving Improvement:** We hear how an improvement network evaluates the success of its improvement efforts.
- **Research Headlines:** We share a roundup listing all of our members' research from the past quarter.

Happy reading, and happy holidays! We look forward to seeing you again next year for Volume 4.

NNERPP | Extra Online

Be sure to check out the NNERPP | Extra website if you'd like to explore this issue's articles (and more!) online.

About NNERPP

NNERPP aims to develop, support, and connect research-practice partnerships in education to improve their productivity. Please visit our website at nnerpp.rice.edu and find us on Twitter: @RPP_Network.

IN THIS ISSUE

[p. 1] WELCOME

[p. 2] RESEARCH INSIGHTS

Developing Valid Measures of RPP Effectiveness: An Evidence-Centered Design Approach

[p. 6] RPP DEEP DIVE

RPP Evaluation: What Does An Evaluator Do?

[p. 11] EXTRA CREDIT: HOW TO

Measuring RPP Health: Introducing a New Tool Based on the Five Dimensions of RPP Effectiveness

[p. 14] IMPROVING IMPROVEMENT

How to Know an Improvement Effort is Succeeding?

[p. 17] RESEARCH HEADLINES

Developing Valid Measures of RPP Effectiveness: An Evidence-Centered Design Approach

Caitlin C. Farrell | National Center for Research in Policy and Practice

Introduction

In the past decade, research-practice partnerships (RPPs) have moved **from rare examples to widespread models in the field**. Despite this growth, we do not have a good sense of how well RPPs are achieving their aims, nor do we have validated tools to help RPPs monitor their progress towards their goals. Thinking more deeply about the effectiveness of RPPs is necessary to understand if they are a good long-term “bet” that can indeed contribute to equitable improvements to educational systems as intended.



The field-driven **RPP Effectiveness Framework** developed by Henrick and colleagues in 2017 offers an important beginning point in mapping how RPPs make progress toward common outcomes that partnerships pursue, as well as how they **attend to equity** in doing so. The next step is to develop a set of valid measures that are tied to these outcomes. Without such measures, it is hard for research, practice, and community partners to adjust their work towards more productive ends. In order to fill this gap, the **National Center for Research in Policy and Practice** and **NNERPP** are collaborating on a project funded by the William T. Grant Foundation to develop measures that can help an RPP assess its progress towards its goals.

In this edition of the Research Insights series, I describe our recent efforts and explain how we have used Evidence-Centered Design (ECD) to guide our work.

Our Design Principles

We aim to develop measures that are able to distinguish among RPPs that are just developing versus those that have mature practices in key areas, in order to provide any RPP with useful feedback to help strengthen their work together. Measures also need to be valid and connected to use, that is, how they will be interpreted to inform decision-making and action. Below are the set of design principles that have guided our work to date:

- We recognize that participatory methods support the development of tools that reflect the perspectives and needs of the end-users. Hence, our approach should involve RPP members and their perspectives at all stages of the work.
- We recognize that RPPs can benefit from low-stakes, formative evaluation to help them progress. Therefore, our tools should support this low-stakes formative evaluation of individual RPPs.
- We recognize that RPPs have different approaches to their work, including research alliances, design-based partnerships, community-engaged partnerships, networked improvement communities, and hybrids of these approaches. Hence, our tools should be relevant to and useful for any of these various types of RPPs.
- We recognize that RPPs may have made differing degrees of progress on their goals based on how long they have been working together. Therefore, our tools should be relevant to and useful for RPPs that have been working together for different time frames.

continued on the next page

Developing Valid Measures of RPP Effectiveness: An Evidence-Centered Design Approach, continued

- We recognize that RPPs include partner members coming from different organizational types and professional roles. As such, our tools should be able to capture multiple perspectives from those centrally involved in the RPP, including research, practice, and community members.
- We recognize that RPPs' progress on their goals is not binary but developmental. Therefore, our tools should allow for RPPs to move between different stages on the developmental trajectory as conditions change.
- We recognize that more equitable relationships are central to all aspects of RPP efforts. Therefore, we attend to equity within each dimension of the Henrick et al. (2017) framework.
- We recognize that RPPs may have multiple projects and lines of work. Hence, our tools should invite RPP members to select a single project to focus on for their responses. Specifically, we invite RPP members to select a project that (1) reflects the core aims of the partnership, and (2) for which the RPP would like more information about the project within the partnership.

Background: Evidence-Centered Design

In this project, we draw on a principled approach to assessment design called Evidence-Centered Design (ECD) to develop, refine, and test tools for measuring an RPP's development. The ECD approach centers the goals and purposes of the measurement activity, and works to ensure that evidence is gathered and interpreted in ways that are consistent with those goals (Mislevy & Haertel, 2006; Mislevy et al., 2003; Mislevy et al., 1999). ECD differs from older approaches to validity and reliability that treat validity and reliability solely as a property of the instruments themselves. In contrast, ECD foregrounds matters of validity for particular purposes, in our case, for low-stakes, formative evaluation of individual RPPs. It is construct-centered: At all phases of measures development, a clear sense of the respective construct guides decisions about what is to be measured, when and where measurement will take place, what items will be used to bring forward information about constructs, and how scores generated from responses to measures will support valid interpretation of data.

Broadly, ECD involves three sets of activities:

1. It begins by defining constructs to be measured, followed by identifying the range of situations and behaviors that might elicit evidence of these constructs. In our work, we drew on the 5 dimensions of the Henrick et al. (2017) RPP effectiveness framework, and we consulted research evidence on RPPs, past measures and tasks, and input from relevant stakeholder groups to describe our constructs.
2. Next, the team selects or creates tasks or measures to gather observable evidence of desired attributes for those constructs, such as valued activities, dispositions, or skills. In our case, we developed a set of survey and interview measures through an iterative design, testing, and revision approach. Our field test involved close to 300 RPP partners from 65 different RPPs to test our survey and interview measures.
3. Finally, the team analyzes how well the measures produce valid and consistent evidence for the constructs and determines what kind of guidance supports sense-making and interpretation of results for different purposes. With the documentation produced during the ECD process, the team then can adjust these different components, as ECD unfolds iteratively, with measures being refined as new information emerges from the performance of items or tasks. We are currently working on examining the validity of the measures, as well as producing guidance on when and how these measures can be used in practice.

continued on the next page

Developing Valid Measures of RPP Effectiveness: An Evidence-Centered Design Approach, continued

An Example: Measuring Trust and Cultivating Relationships

Below, we share a draft version of our hypothetical developmental progression for one of our key constructs, building trust and cultivating relationships (outcome 1 of the Henrick et al. framework; Table 1), and the progression for attending to equity when building trust and cultivating relationships (Table 2). In cultivating trust and relationships, partnerships can attend to equity by addressing historical imbalances of power among participating individuals, organizations, or communities. Imbalances of power can be related to historical relationships among research and practice organizations as well as to social identities related to race or ethnicity, language, gender, sexual orientation, ability, citizenship status, age, and more. Addressing these imbalances of power involves supporting authentic participation and voice among partnership members, and specifically elevating the perspectives of those with less power, as opposed to simply aiming for “everyone having an equal voice.”

Table 1. Description of the developmental progression for building trust and cultivating relationships in RPPs or RPP projects.

FORMING	MATURING	SUSTAINING
Partners have opportunities to follow through on their commitments to one another.	Partners generally follow through on their commitments to one another.	Partners almost always follow through on their commitments to one another.
Partners learn about each others’ contexts and establish routines of interaction and communication.	Partners have a general understanding of each others’ contexts, perspectives, and commitments through established routines of interaction and communication.	Partners have deep understandings of each others’ contexts, perspectives, and commitments through well-established routines of interaction and communication.
Partners may not yet feel comfortable raising difficult issues with one another.	Partners engage in honest discussion of difficulties by the partnership.	Partners have worked through both internal conflicts and challenges to achieving goals, with relationships strengthened.

Table 2. Description of the developmental progression for attending to equity when building trust and cultivating relationships in RPPs or RPP projects.

FORMING	MATURING	SUSTAINING
Partners identify differences in historically limited social power of participating individuals or organizations in the work together.	Partners elevate the voices of participating individuals or organizations with historically limited social power in their work together.	Partners regularly address the historically limited social power of participating individuals or organizations, removing barriers and creating structures that support authentic involvement, participation, and voice.

continued on the next page

Developing Valid Measures of RPP Effectiveness: An Evidence-Centered Design Approach, continued

From these trajectories, we identified situations where we might observe or gather evidence related to an RPP's placement on the trajectory. For instance, looking at the notes from an RPP's regular meeting could provide insight into the routines that support interaction that build trust, or the ways in which different voices are heard. As for key informants, we suspect leaders within each partner organization would be critical for evaluating trust, although the perspectives of a wide range of stakeholders would be appropriate to evaluate overall trust and shared commitment.

Two sample survey items that might help elicit information about an RPP's placement on this trajectory include:

1. What best describes your partnership?
 - a) My partners are reliable in completing tasks in an agreed upon time;
 - b) My partners sometimes complete tasks on time.
 - c) We have established some deadlines for our partner's tasks, but they have not come up yet.
 - d) My partners do not reliably complete tasks on time.

2. Our partnership privileges the ideas of practitioners, community members, or other stakeholder groups in our meetings. [Strongly Agree/Agree/Neither Agree nor Disagree/Disagree/Strongly Disagree]

In Conclusion

It is our goal to develop valid and reliable measures that RPP members can use to gather information that helps them evaluate their partnership and make progress on their goals. In terms of next steps, we are currently working through our collected data to make sense of the information, in addition to putting our constructs and claims through testing to ensure their reliability. We hope this article provided some insight into this work and the value of approaching it using Evidence-Centered Design. We welcome any **feedback** on these efforts and look forward to continuing to involve the NNERPP community in this work.

Acknowledgement

This work is funded through the generous support of the William T. Grant Foundation. Our team includes Caitlin Farrell, Bill Penuel, Robbin Riedy, Julia Daniels, Kristina Stamatias, Kristen Davidson, Sarah Wellberg, Paula Arce-Trigatti, and Jim Soland. We are so appreciative of all NNERPP members who have been involved in the project to date.

- *Caitlin Farrell is director of the National Center of Research in Policy and Practice (NCRPP).*

RPP Evaluation: What Does an Evaluator Do?

By Stacey Sexton | SageFox Consulting; Erin Henrick | Partner to Improve; Madeline Noya | Aldea Analytics; and Danny Schmidt | Partner to Improve

INTRODUCTION

Since its inception in 2017, more than 125 unique projects have been borne from the National Science Foundation's (NSF) **Computer Science for All (CSforAll: Research and RPPs)** (hereafter CSforAll: RPP) program, yielding many new RPPs in various stages of development. Given the **emergent nature** of our collective knowledge of RPPs and evaluation, we were motivated to better understand, and thus better support, external project evaluators associated with CSforAll: RPPs.



Specifically, we were interested in further exploring 1) characteristics of CSforAll: RPP evaluators and evaluation plans, 2) evaluation activities, 3) evaluation use and impact, and 4) RPP evaluator needs. To begin to accomplish this, we reviewed evaluation plans from 32 CSforAll: RPP proposals, and conducted interviews with 7 external evaluators representing 14 RPPs. The participating projects (who are all members of the **RPPforCS community**) come from the first three cohorts of the CSforAll: RPP program. The analysis includes 10 large grants (maximum \$2M), 16 medium grants (maximum \$1M), and 6 small grants (maximum \$300K). Here we share initial findings from this exploratory study and conclude with reflections to further support the RPP evaluation community.

BACKGROUND

Evaluation is the “**systematic investigation of the worth or merit of an object**” and involves gathering and analyzing data to inform learning, decision-making, and action in order to contribute to the improvement of a program or policy (Weiss, 1998). NSF expects every CSforAll: RPP to incorporate external review and feedback on the project's designs and activities, but project teams have the freedom to determine the role of the evaluative body, how (and what types of) feedback is provided to the team, the cadence of evaluation milestones, and how the team uses information from an evaluation. Currently, there are few models for RPP evaluation. Given this, understanding more about how the CSforAll: RPPs are using evaluation to support their work is a useful topic to investigate.

FINDINGS

(1) Characteristics of CSforAll: RPP evaluators and evaluation plans

External review design. Given the exploratory nature of this study, we first wanted to understand how CSforAll: RPPs designed their external review and feedback plan. 38% of the projects in our sample had an external evaluator, 31% of the projects had both an advisory board and external evaluator, and 16% of the projects only had an advisory board (15% of the evaluation plans did not specify). 50% of the evaluators in the sample were described as being affiliated with an independent consulting firm and 19% were affiliated with a university (the remaining projects either use an advisory board model or did not report evaluator affiliation).

continued on the next page

RPP Evaluation: What Does an Evaluator Do?, continued

Type and focus of evaluation. We then examined the evaluation plans to determine the type and focus of the evaluation. 34% of the plans included both formative and summative evaluations, 31% of the plans included formative evaluation only, and only one plan described only a summative evaluation (31% of the proposals did not specify the type of evaluation planned). In order to understand the foci of the evaluations, we categorized the evaluation questions into three categories: 1) 38% of the evaluation plans included questions to evaluate both the partnership and the overall CSforAll: RPP program, 2) 34% only included questions to evaluate the program, and 3) 16% only included questions to evaluate the partnership (13% did not specify evaluation questions).

Six out of seven evaluators we interviewed reported investigating the health of the partnership, with a focus on the **5 Dimensions of Effectiveness of RPPs** (Henrick, Cobb, Jackson, Penuel, Clark, 2017). While equity did not emerge as a central focus for the CSforAll: RPP evaluations (which was a surprising finding for us), two evaluators did specifically mention that they pay particular attention to whose voices are being heard, who is participating in RPP discussions, and power dynamics within the RPP, asking questions like “Are we practicing what we preach around equity?” and, “Talk about who is at the table. Whose voices are you listening to?”

Evaluator background. In order to understand evaluator backgrounds, we examined the brief biographies included in the proposals. While 25% of the plans did not specify the evaluator’s expertise, 63% reported experience in evaluation, 47% reported having experience in STEM/CS, and 34% of the evaluators reported experience in RPPs.

Interviewees were asked what characteristics they believed were most important to be an effective evaluator of RPPs. The most common themes related to building relationships with the partners. One interviewee said, “Engender trust first. It’s more important that people on both sides trust you. More important than being right. More important than showing how smart you are.” Strong communication skills and being able to report findings in an easy to understand and actionable format was another theme that emerged from interview responses. Another evaluator also emphasized honest communication saying, “Being a proactive communicator. It lends itself to that critical friend role, as well as just being honest about how things are going.”

Role of the evaluator. 6 out of 7 evaluators we interviewed described their role as a “critical friend.” One evaluator further clarified this term- “Unlike a traditional evaluation, you’re deeply embedded in the work routines. You have to be, to kind of understand them, that’s part of the process.” One evaluator described their role as a “thought partner” and two described their role as a “broker.” Some interviewees acknowledged that their role has changed over time. One evaluator stated: “Roles will always get renegotiated but that’s the primary work in the first year- figuring out the roles and how to collaborate. The focus of evaluation should be on process and relationships and workflow and maybe later it goes to outcome and impact. I wish there was more focus in RPPs on that evolution.” One interviewee indicated that they tend to play a more embedded role when evaluating an RPP than in traditional research grants. We conjecture this might be due to the complex nature of RPPs, challenges related to understanding how the partnership operates, and the need for RPP evaluators to be responsive.

(2) Evaluation activities

Evaluator activities. An analysis of the evaluation plans revealed the types of activities in which the evaluators planned to engage, as well as the planned data collection methods. 63% reported plans for the evaluator to collect data, 53% reported plans to analyze data, 44% described plans to report results, and 19% described plans to attend meetings (25% of the plans did

continued on the next page

RPP Evaluation: What Does an Evaluator Do?, continued

not specify evaluator activities). Interviewees reported engaging in three main types of activities: collecting data, attending meetings, and providing feedback to partners and NSF. All evaluators interviewed reported collecting data and developing evaluator reports for NSF; half of the interviewees reported also attending team meetings.

Types of data collection. The most common data collection methods described in the evaluation plans were surveys and interviews, each appearing in 47% of the evaluation plans. 31% reported plans to conduct focus groups, while conducting observations and document analysis each appeared in 25% of the plans. 13% of the evaluation plans did not specify data collection methods. Interview data indicated that evaluators typically utilize a variety of qualitative data methods to determine the project's success along the five dimensions identified in the Henrick, et al. (2017) framework.

(3) Evaluation use and impact

Interviewees identified three areas where evaluation feedback made an impact on the RPP with specific examples from their own projects.

Improving the functioning of the RPP. Evaluation activities and feedback cycles allowed some evaluators to improve the functioning of the RPP by helping to establish clearly defined roles and group norms amongst the partnership participants, while another emphasized their effects on the group's problem solving cycle saying, "From the first year we developed a problem solving cycle. It's become ingrained in [the organization]. It's how they orient all of their partnership work." Another evaluator also mentioned some smaller

logistical changes that benefited their RPP saying, "Nuts and bolts things they can change. Simple things like [using] action items."

Improving research efforts. Some evaluators described how evaluation feedback helped improve the project's research efforts. One evaluator described how they were able to use survey and interview responses to improve participant recruitment strategies. Another described using evaluation data to improve the parent consent process saying, "We had a suggestion early on on how to make it slightly less onerous."

Improving program offerings. Finally, evaluators reported their impact on the intervention. Some said they influenced the overall design and frequency of delivery of professional development workshops delivered based on feedback from program participants. One evaluator described how the program offered "a third round of PD that they weren't planning to do, and it was totally different from the first two rounds, because part of the evaluation data showed them that their pre-packaged curriculum wasn't meeting the needs of the teachers." Another evaluator stated, "They viewed our surveys and interviews as a way to check progress on how things are going. The feedback we provided would influence how the PD would go throughout the school year."

It's important to note that in addition to the examples above of ways evaluators felt they had meaningful impacts on their RPPs, some evaluators also expressed a lack of perceived impact on their projects. One evaluator stated, "We write formal reports that get appended to things. Not sure it's being used to improve their practice." While another said plainly, "I don't think it's



continued on the next page

RPP Evaluation: What Does an Evaluator Do?, continued

changed the function of the RPP.” Yet another expressed the variability in how feedback is received by different members of the team saying, “Feedback is not well received by the PI, but I get emails from other project team members explaining it was helpful, thanking me.”

(4) RPP evaluator needs

When asked what tools and supports they would like in order to be more successful, respondents reported being generally satisfied with the evaluation instruments currently available, but expressed a desire for a forum with which to have honest conversations with other evaluators. Five of the seven interviewees emphasized the importance of learning from their fellow RPP evaluators. Several mentioned wanting to learn through the results of this study.

LIMITATIONS

There are several limitations to this study. The proposals we were able to analyze were a small sample of the funded projects to date. The evaluation plans in the proposals varied greatly in the amount of detail provided about the specific methods and approaches, so some information we collected was incomplete. The biggest limitation to understanding the evaluators’ role and impact within the project in our view is that we were unable to interview other RPP members to understand their perspectives.

DISCUSSION

NSF has allowed significant flexibility in how the external evaluation role is fulfilled, and the findings from this study indicate great variation related to how the role is implemented across the CSforAll: RPP program. Findings suggest that it is important for RPP teams to discuss the role and responsibilities of external evaluation particularly regarding the concept of “critical friend.” From our perspective, one can only succeed as a critical friend if there is agreement across the RPP that 1) practitioners and researchers welcome critical feedback, 2) the evaluator will be included and have enough information to make a meaningful assessment, 3) critiques will be constructive and intended to improve processes and outcomes, and 4) feedback will be discussed openly.

Additionally, we believe that CSforAll: RPP evaluations would benefit from a more explicit equity focus, particularly given fundamental values related to equity and inclusion ingrained in the CSforAll: RPP program. If evaluators are unclear on what it means to focus on equity in RPP processes and outcomes, then we risk reinforcing, rather than addressing, the disparities that the CSforAll: RPP program is targeting. We believe that RPP evaluators can play a key role in [assessing equity in both RPP processes and outcomes](#). The Center for Evaluation Innovation (2017) outlines a framework and guiding principles for equitable evaluation that we believe are relevant to the CS for All community of RPPs, including ensuring that evaluative work “hold at its core a responsibility to advance progress towards equity” (p. 4).

This pilot study is a first step towards understanding more about RPP evaluation, including the characteristics of CSforAll: RPP evaluation plans and evaluators, types of activities RPP evaluations engage in, how RPP evaluation is supporting the goals of the project, and RPP evaluator needs. NSF has created a unique opportunity for RPPs by requiring a critical review process, but we need to learn more about how to support RPP evaluators and how RPP evaluations can be designed and implemented to best support the needs of RPP teams.

continued on the next page

RPP Evaluation: What Does an Evaluator Do?, continued

● Team

Our study team consists of four RPP evaluators with experience across a range of RPP types and evaluation experiences. Stacey Sexton has been an evaluator on three RPPs funded through the NSF and is also the facilitator of the RPPforCS community. Erin Henrick is the lead author of the Five Dimensions of RPP Effectiveness framework and an evaluator of several RPPs within and outside of CS. Madeline Noya has served in the evaluation role in two CS for All: RPPs, and Danny Schmidt is the research analyst for this study and several CS for All: RPP evaluations conducted by Partner to Improve.

References

Arce-Trigatti, P. (2021). What's New With RPP Effectiveness? *NNERPP Extra*, 3(2), 9-14.

Center for Evaluation Innovation, Institute for Foundation and Donor Learning, Dorothy A Johnson Center for Philanthropy, Luminare Group. "Equitable Evaluation Framing Paper." *Equitable Evaluation Initiative*, July 2017, www.equitableeval.org.

Henrick, E.C., Cobb, P., Jackson, K., Penuel, W.R., Clark, T. (2017). *Assessing Research-Practice Partnerships: Five Dimensions of Effectiveness*. New York, NY. William T. Grant Foundation.

Henrick, E., McGee, S., & Penuel, W. (2019). Attending to Issues of Equity in Evaluating Research-Practice Partnership Outcomes. *NNERPP Extra*, 1(3), 8-13.

National Science Foundation. 2020. Computer Science for All (CSforAll: Research and RPPs). https://nsf.gov/funding/pgm_summ.jsp?pims_id=505359

Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies*. Prentice Hall, Upper Saddle River, New Jersey.

Measuring RPP Health: Introducing a New Tool Based on the Five Dimensions of RPP Effectiveness

By Elizabeth Barkowski and Carrie Scholz | American Institutes for Research

Portions of this article are directly reprinted with permission from [this original blog post](#) by Regional Educational Laboratory (REL) Southwest. We are grateful to the authors, REL Southwest, and the Institute of Education Sciences (IES) for allowing us to help spread the word about this new tool with you!

There is limited literature on how to formatively assess the effectiveness of research-practice partnerships (RPPs). More tools and processes are needed so that RPP members, regardless of how long the partnership may be in place, are able to use their time more efficiently and to accelerate progress toward achieving their goals. We identified these needs in our own work, and developed a new tool to address these needs, which we introduce here.

REL Southwest partners with state education agencies, school districts, and other educational organizations to address pressing education issues through RPPs in Arkansas, Louisiana, New Mexico, Oklahoma, and Texas. Our website's [partnership pages](#) list partnership members and describe the focus of work in each state. After year one of the RPPs, we identified a need to improve their participation, engagement, and overall functionality. To inform our next steps, we needed a research-based tool to assess RPP health. For this reason, REL Southwest developed the [Tool for Assessing the Health of Research-Practice Partnerships](#).

How Did We Develop This Tool?

As the basis of the new tool, REL Southwest used an existing framework defined in [Assessing Research-Practice Partnerships: Five Dimensions of Effectiveness](#) (2017) by Henrick, et al., which readers of this magazine might be familiar with through previous articles discussing different aspects of this framework (see [here](#), [here](#), and [here](#), for example). The Henrick, et al. (2017) framework provides dimensions and indicators that researchers and practitioners must address to build and sustain an effective RPP. The authors of this framework intended the indicators to serve as a guide for future RPPs to develop more specific measures of RPP effectiveness, and the framework was intended to be customized for specific RPP needs.



In addition to meeting the research community's need to address the challenge of RPP evaluation and improvement, the tool provides a means for participation by practitioner partners. Teachers, administrators, and other practitioners can use the tool to identify challenges and issues with the RPP's process and goals, ensuring they have a voice in its assessment and continuous improvement. REL Southwest developed the new tool for individuals—researchers or practitioners—who want to formatively assess and continuously improve an RPP in collaboration with its members. Together the partners set short- and medium-term measurable goals, establish processes and structures to support these goals, and reflect on the degree to which their work accomplishes these goals.

How Can You Use This Tool?

The REL Southwest Tool for Assessing the Health of RPPs includes a guidance document and an Excel workbook that consists of three parts. The tool's components are also outlined in [this infographic](#).

The tool guidance includes information on how to use each of the workbook's three parts, including:

1. Guidance to prioritize dimensions and indicators.
2. Guidance to establish short- and medium-term goals and measures.
3. Interview protocol designed to help conduct interviews with the RPP members to assess the health of prioritized dimensions and indicators.

continued on the next page

Measuring RPP Health: Introducing a New Tool Based on the Five Dimensions of RPP Effectiveness, continued

Individual RPP members may use the tool effectively without a significant investment of time, even if they have limited experience evaluating or participating in RPPs.

By using a formative lens rather than a summative one, the tool encourages RPP members to purposefully and honestly reflect on their collaborative work and to make necessary adjustments and improvements over time to achieve the partnership's intended outcomes. Using the tool at the onset of a partnership could help ensure all members have a common understanding of the process and goals as the partnership gets underway. Periodic assessment could help keep the RPP focused on goals or identify a need to adjust them. By conducting the interviews with RPP members, evaluators can discover additional needs that the partnership may not be meeting for its members.

An Example of How the Tool Was Used and Adapted in an RPP

The REL Southwest team at American Institutes for Research (AIR) decided to pilot this tool with an RPP outside of the REL program. The goal was to see if the tool and related processes would be feasible in an RPP with a single district, one of the predominant configurations of RPPs across the country. Here is a brief overview of how the RPP used the tool and of some of the outcomes associated with its use.

PUMP-CS RPP pilots the tool. Preparing Urban Milwaukee Pathways in Computer Science (PUMP-CS) is an NSF-funded Computer Science for All RPP. We worked with them starting in 2019 to pilot our tool: We first introduced participating RPP members to the Henrick et al. (2017) framework so they could familiarize themselves with the five dimensions and the indicators for each dimension, then used our tool to facilitate a process for the RPP members to collaboratively determine quarterly priorities for the RPP based on the framework's dimensions and indicators.

The RPP enthusiastically decided to prioritize all five dimensions and prioritized at least one indicator per

dimension to focus on in the first quarter. AIR used the tool's protocol to interview each member of the RPP leadership team (2 district staff members and 1 researcher and 1 technical assistance partner). After discussing the interview results, the RPP leadership team identified the following action items aligned to the prioritized indicators listed below.

Dimension 1: Building Trust and Cultivating Partnership Relationships

Prioritized Indicator: Researchers and practitioners routinely work together

- Action items for PUMP-CS: Continue to develop the partnership's logic model during the RPP meetings and share it with other stakeholders (i.e., the partnership's steering committee) to further inform its development as well as establish a shared understanding of the partnership's efforts and activities.

Prioritized Indicator: The RPP establishes routines that promote collaborative decision making and guard against power imbalances

- Action items for PUMP-CS: Establish bi-weekly touchpoints for the RPP leadership team and ground conversations in the logic model so the prioritized activities and outcomes inform key decisions.

Dimension 2: Conducting Rigorous Research to Inform Action

Prioritized Indicator: The RPP establishes systematic processes for collecting, organizing, analyzing, and synthesizing data

- Action items for PUMP-CS:
 - Develop a timeline for district data requests
 - Identify the audiences for the data analyses and establish a portal for aggregating data so the RPP may access findings on an as-needed basis
 - Establish processes that invite the district to co-interpret results

continued on the next page

Measuring RPP Health: Introducing a New Tool Based on the Five Dimensions of RPP Effectiveness, continued

Prioritized Indicator: Findings are shared in ways that take into account the needs of the practice organization

- Action item for PUMP-CS: Identify existing research that may address the district’s immediate needs while waiting for findings from current data collection efforts

Dimension 3: Supporting the partner practice organization in achieving its goals

Prioritized Indicator: The RPP provides research and evidence to support improvements in the partner organization

- Action item for PUMP-CS: Collaboratively articulate measurable, timebound short- and medium-term goals for each prioritized indicator that align to the logic model

Dimension 4: Producing knowledge that can inform educational efforts more broadly

Prioritized Indicator: The RPP develops two dissemination plans, one that supports partnership goals and a second plan for broader dissemination

- Action items for PUMP-CS: Identify which district stakeholders need to increase their understanding of the partnership’s efforts and develop a strategy for communicating with them

Lessons learned. Meaningful conversations take time. Prioritizing dimensions, selecting indicators, and setting SMART (specific, measurable, achievable, relevant, time bound) goals required up to 1.5 hours of meeting time. Rather than rush these conversations and gather feedback via interviews every quarter, in the second year of the partnership, the leadership team decided to use the tool every 6 months and to prioritize 2-3 dimensions and a subset of their indicators during that 6-month period. This adjustment allowed the RPP to accomplish more significant, targeted work with fewer action items needing to be pushed to the following 6-month period.

In addition to interviewing the RPP members every six

months, AIR also added a brief survey modeled after the rubric developed by Henrick et al. called “Are We a Partnership Yet”? This helps to serve as a quick temperature check of how each partner is feeling about the different dimensions at play. The survey data along with the themes from the qualitative data are discussed every six months. According to the most recent survey findings, the RPP members reported that the RPP is on track toward maturing on the “building trust and cultivating partnership relationships” dimension. Feedback during the interviews suggested that: (1) the RPP members believe the partnership is changing the status quo of CS instruction in the district and (2) the diversity of expertise strengthens the partnership.

PUMP-CS RPP members’ reflection on the use of the tool

“I think that the structure that you guys are putting in place here is really essential. It’s going to change the way we do business, and I think that’s really important.”
– RPP district staff member

“It has been a breath of fresh air to have another partner joining us in this work, and to actually get us to look up...at the larger context of the project. That’s been something that has been fun and informative.”
– RPP researcher representative

In Conclusion

We hope that by providing guidance on how the dimensions and indicators of the Henrick et al. framework can be used to establish goals and monitor how an RPP is progressing towards these goals, our tool will be useful to individuals working in RPPs that wish to assess the health of their partnership. The REL Southwest team invites you to [contact us](#) with questions or to share your experiences using the tool.

- *Elizabeth Barkowski is a senior researcher and Carrie Scholz is a principal researcher at American Institutes for Research.*

Improving Improvement: How to Know an Improvement Effort is Succeeding?

By David Hersh | Proving Ground



This is the sixth installment of *Improving Improvement*, our quarterly series focused on leveraging the power of research-practice partnerships (RPPs) to build schools', districts', and states' capacity to improve. In our [previous article](#), we laid out the work ahead for our partnerships and the questions we hope to answer in the 2021-2022 school year. In this installment, we are taking a step back to reflect on how we evaluate the success of our improvement efforts.

Defining Success

As with any program evaluation, the starting point for Proving Ground's assessment of whether or not we are succeeding is a clear definition of success. What does success look like? In the last installment of *Improving Improvement*, we noted that our ultimate goal is for our partners to make evidence-based continuous improvement –which includes piloting and evaluating interventions– part of the ordinary course of business in their agencies. When we break down this goal, we see two components of success: 1) our partners practice evidence-based continuous improvement and 2), they do so routinely, in the ordinary course of business. We developed operational definitions for both constructs – “evidence-based continuous improvement” and “ordinary course of business”, outlined below.

For Proving Ground, evidence-based continuous improvement involves a set of core activities or competencies executed sequentially and repeated until the desired amount of improvement is achieved. The core activities are, in order:

1. Clearly define the problem and set an improvement goal for it
2. Identify root causes
3. Identify a set of potential interventions aligned to the root causes
4. Prioritize a potential intervention from that set to try
5. Design the intervention using user-centered design principles
6. Plan for implementation and progress monitoring
7. Pilot to generate evidence of impact
8. Use evidence from the pilot to decide whether to stop, scale or adapt
9. Reflect on the results in light of your improvement goal

Thus, our partners are practicing evidence-based continuous improvement if they are executing these activities. Success, however, implies doing so with some minimal threshold of fidelity. What if partners identify root causes that aren't really root causes? What if they pilot but in ways that generate only spurious evidence of impact? We therefore developed a rubric to define in some detail the characteristics of optimal execution. For example, the quality of piloting might range from not piloting at all to doing so in a way that supports a causal inference of impact.

In this way we can check how well each partner is doing while they work with us (all of our partners execute at least one improvement cycle during our engagement). However, doing it well while we are there to support says little about how well they would do it after our engagement ends, or whether they would do it at all. You can only make so much impact in one

continued on the next page

Improving Improvement: How to Know an Improvement Effort is Succeeding?, continued

cycle. This is why capacity building is an important part of our role. We need to assess not just the quality with which partners execute the competencies of improvement while they work with us, but whether they continue doing it and how well they do it when we are gone.

In a similar way, success also requires some minimal level of coverage. If partners are great at continuous improvement cycles but only do them on one team and/or for one narrow problem of practice a year, the likelihood of having meaningful long-term impacts is small. Therefore, our goal requires making it part of the ordinary course of business. It needs to be part of the fabric of the organization. Recognizing that it would be impractical to see thorough application in every area, we operationalized “ordinary course of business” by defining it to mean that partners apply evidence-based continuous improvement to all problems of practice aligned with strategic priorities.

Success for us, therefore, means our partners having institutionalized the practice of continuous improvement so thoroughly that they continue practicing the core competencies of improvement with fidelity for all strategically aligned problems of practice long after our engagement ends.

Measuring Progress

Operationalized definitions of success are necessary to track progress but they are not in and of themselves metrics that can be used to measure progress. The next step for us is to develop measurable targets (and measurement instruments if needed) based on the operational definitions [1]. For example, did we succeed with a partner that practices all but two of the core competencies of evidence-based continuous improvement, as defined earlier? What if they do them all and do them well but only for some of their strategic priorities? What if they do them all, for all strategic priorities, but each with limited fidelity? There is also a time component to each question. By when do partners need to be fully practicing evidence-based continuous improvement in the ordinary course of business? It is not realistic to expect that our partners will fully institutionalize the practices during our engagements, so we need both progress markers – leading indicators that suggest they will get there – and a way of gathering data about their practice after our engagement ends.

For our leading indicators, we are looking for necessary steps along the path to fully institutionalized continuous improvement, which we have conceptualized into three phases:

Phase 1: Do our partners execute a high-quality improvement cycle while working with us? How well did they execute each competency and to what degree was the decision outcome optimizing?

Phase 2: Based on their work with us, how confident are we that they are able to do this without us – can they generalize from the model we worked on together to other problems of practice? Additionally, how confident are we that they are willing to do this without us – have we created the internal demand?

Phase 3: Are they doing this after our engagement ends? How well and for how many of their strategic priorities?

The first two phases establish the leading indicators. Our first leading indicator is the quality of the work our partners do while they are engaged with us. Using the rubric we developed, we can assess the execution of each competency. We have (or will have) targets for the share of partners scoring at defined levels on the different competencies. The second leading indicator is more difficult to measure. While we can reasonably rely on interactions, artifacts, and a partner self-assessment developed

continued on the next page

Improving Improvement: How to Know an Improvement Effort is Succeeding?, continued

using the rubric referenced earlier to draw inferences about partners' capacity to practice evidence-based continuous improvement without us, we will have limited direct evidence of their likelihood of doing so. We will likely rely on a combination of self-reported intentions and, for some partners, continuity planning work we are doing with them.

Phase 3 is the goal itself but is challenging to measure with fidelity. First, while we have a self-diagnostic rubric that we could administer to partners after the engagement ends, it is reasonable to assume response rates will decline as we get further removed from the engagement. While the network is designed to be ongoing –all our partners continue to be members of the PG Network after the direct engagement with us– the intimacy of our connection is likely to decline. More problematically, the self-diagnostic is not designed for summative use, making it difficult to use for evaluating how well Proving Ground is doing. It is a formative tool we and partners can use to identify the competencies on which they need the most work. It is highly possible that as partners' sophistication in continuous improvement increases –a good outcome– their self-assessments will get more critical, resulting in lower scores – implying a bad outcome. Similarly, there are likely to be issues with inter-rater reliability, as the individual(s) using the rubric may change over time.

There are, therefore, challenges we face in tracking progress towards our goal. We will, however, practice what we preach and continuously reflect on and improve how we go about measuring how well we are doing.

The Implicit Ultimate Goal

What has so far gone unsaid is that the goal laid out here is instrumental to a larger goal. Practicing evidence-based continuous improvement is a means to an end – part of a theory that the best way to improve outcomes for students is to ensure education agencies are learning organizations that are highly skilled and disciplined in getting better. Framed this way, the real goal of our work is to improve student outcomes. If our theory is correct, partners will practice evidence-based continuous improvement in the ordinary course of business and, as a result, all the outcomes on which they practice it will improve. Improvement in our partners' priority outcome areas should be measurably greater than that for comparable agencies that have not worked with us or otherwise practiced evidence-based continuous improvement. We are working on a way to measure this that would also address the selection effects we are likely to encounter – we do not randomly choose partners. Stay tuned for more on where we land.

Looking Ahead

In the next installment of Improving Improvement, we'll share updates on the progress of our newest networks, the Georgia Improvement Network and the Rhode Island LEAP Support Network, including lessons learned from partnering with states to support districts on their improvement journeys.

We are also always open to additional suggestions for topics for future editions of Improving Improvement. Please reach out to us with any questions you have about our networks, continuous improvement process, or ideas you'd like to see us tackle.

[1] While we have developed these, including them in detail here is beyond the scope of this article.

Research Headlines From NNERPP Members: Last Quarter

CAREER AND TECHNICAL EDUCATION

GEORGIA POLICY LABS examines
 -- how state contexts affect participation in high school CTE programs
 -- CTE participation in Montana
 -- factors that determine equity and access in CTE participation

COMPUTATIONAL THINKING & CS

CHICAGO ALLIANCE FOR EQUITY IN COMPUTER SCIENCE examines CS teacher persistence and attrition

DIGITAL PROMISE provides resource on computational thinking for educators

COVID-19

DETROIT EDUCATION RESEARCH PARTNERSHIP examines how Detroit families' experiences during the pandemic shaped school attendance

EDUCATION POLICY INNOVATION COLLABORATIVE examines teacher turnover in Michigan's lowest performing schools during the pandemic

REL CENTRAL examines district strategies for remote learning during the Covid-19

REL MIDWEST examines changes in student learning in Illinois school districts during the cCovid-19 pandemic

TENNESSEE EDUCATION RESEARCH ALLIANCE examines changes in educators' experiences during the pandemic in the 2020-21 year

EARLY CHILDHOOD EDUCATION

MULTNOMAH COUNTY PARTNERSHIP FOR EDUCATION RESEARCH examines family perceptions of participating in a structured summer kindergarten transition program

NYC EARLY CHILDHOOD RESEARCH NETWORK examines role of New York City's pre-k site support personnel

EARLY CHILDHOOD EDUCATION, continued

REL NORTHWEST examines the implementation and effects of Alaska's pre-elementary grant program

URBAN EDUCATION INSTITUTE examines public pre-k supply & demand in Bexar county

EARLY WARNING SYSTEMS

REL MID-ATLANTIC compares the accuracy of a prior performance early warning system and a machine learning algorithm

REL NORTHWEST examines first-year effects of early indicator and intervention systems in Oregon

ENGLISH LEARNERS

UCHICAGO CONSORTIUM examines outcomes for English Learners in pre-k and the early grades

EVALUATION

REL CENTRAL provides program evaluation toolkit

FAMILY ENGAGEMENT

REL MID-ATLANTIC examines impacts of home visits on student outcomes and parent engagement

INSTRUCTION

EDUCATION POLICY INNOVATION COLLABORATIVE examines competency-based education in Michigan's 21j pilot districts

continued on the next page

Research Headlines From NNERPP Members: Last Quarter, continued

POST-SECONDARY

GEORGIA POLICY LABS
examines the effect of passing a CTE technical assessment on college enrollment

REL NORTHWEST
examines impacts of the Oregon Promise program on students' postsecondary attainment

UCHICAGO CONSORTIUM
examines student attainment from a community perspective

SCHOOL CLIMATE

MADISON EDUCATION PARTNERSHIP
examines classroom belonging in middle school

SCHOOL DISCIPLINE

HOUSTON EDUCATION RESEARCH CONSORTIUM examines
-- the relationship between student suspensions and interaction with the juvenile justice system
-- the timing of first school suspension and juvenile justice contact

STEM

RESEARCH ALLIANCE FOR NEW YORK CITY SCHOOLS
evaluates afterschool STEM program

STUDENT ENROLLMENT

REL MID-ATLANTIC
examines how well machine learning algorithms predict early fall student enrollment

STUDENT MOBILITY

HOUSTON EDUCATION RESEARCH CONSORTIUM examines
-- flows of student mobility
-- informal networks of elementary school student mobility
-- the link between student mobility during the school year and students' academic achievement and educational attainment
-- what predicts campus mobility at Houston area public schools
-- the relationship between school-year mobility and school performance in the Houston area

TEACHERS & PRINCIPALS

REL MID-ATLANTIC
examines teacher and principal residency grants

REL MIDWEST
examines supports that are associated with teacher retention

REL NORTHEAST & ISLANDS
examines participation in and impacts of a new teacher mentoring program

TURNAROUND

EDUCATION POLICY INNOVATION COLLABORATIVE
examines the implementation of Michigan's partnership model of school reform through its third year

End Notes

NNERPP | Extra is a quarterly magazine produced by the National Network of Education Research-Practice Partnerships (NNERPP), a professional learning community for education research-practice partnerships (RPPs) housed at the Kinder Institute for Urban Research at Rice University. NNERPP's mission is to develop, support and connect RPPs in order to improve the relationships between research, policy, and practice.



[@RPP_Network](https://twitter.com/RPP_Network)



nerpp.rice.edu

NNERPP is made possible through generous funding provided by the Bill and Melinda Gates Foundation, William T. Grant Foundation, Annie E. Casey Foundation, and The Wallace Foundation.