

# Practical Measurement for Continuous Improvement in the Classroom: A Toolkit for Educators

REL 2023-139  
U.S. DEPARTMENT OF EDUCATION

*A Publication of the National Center for Education Evaluation and Regional Assistance at IES*



# Practical Measurement for Continuous Improvement in the Classroom: A Toolkit for Educators

*Jill Walston and Marshal Conley*

October 2022

This toolkit is designed to guide educators in developing and improving practical measurement instruments for use in networked improvement communities (NICs) and other education contexts in which principles of continuous improvement are applied. Continuous improvement includes distinct repeating processes: understanding the problem, identifying specific targets for improvement, determining the change to introduce, implementing the change, and evaluating if and how the change led to improvements. This toolkit is intended for a team of educators who have already identified specific student learning needs and strategies to improve instruction to address those needs and are ready to test these strategies using continuous improvement processes. The toolkit aims to help the team with the final step in the cycle, which includes collecting data to measure implementation of changes and intended outcomes and using those data to inform future action. Measures for continuous improvement should be closely aligned to student learning goals and implementation of instructional strategies driving the continuous improvement effort, and they should be practical to use in a classroom setting. A team of educators can use this toolkit to proceed through a series of steps to identify what to measure, consider existing instruments, draft instruments, evaluate and refine instruments, plan data collection routines, and plan for data discussions to interpret the data and inform action. Regional Educational Laboratory (REL) Southwest developed the resources in the toolkit in partnership with the Oklahoma State Department of Education team working with the Oklahoma Excel NICs.

# CONTENTS

---

<b>Introduction</b> .....	<b>1</b>
<b>Overview of the toolkit</b> .....	<b>2</b>
<b>Using the tools</b> .....	<b>4</b>
<b>Who should use this toolkit?</b> .....	<b>4</b>
<b>Where to start.</b> .....	<b>5</b>
<b>How to implement the toolkit.</b> .....	<b>6</b>
<b>Practical measurement in the context of continuous improvement.</b> .....	<b>7</b>
<b>What is improvement science?</b> .....	<b>7</b>
<b>What are networked improvement communities?</b> .....	<b>10</b>
<b>What is practical measurement?</b> .....	<b>10</b>
<b>What are the principles of high-quality measurement?</b> .....	<b>12</b>
<b>The six steps of practical measurement.</b> .....	<b>15</b>
<b>Step 1. Identify what to measure</b> .....	<b>16</b>
<b>Step 2. Consider existing instruments</b> .....	<b>20</b>
<b>Step 3. Draft instruments</b> .....	<b>23</b>
<b>Step 4. Evaluate and refine instruments</b> .....	<b>25</b>
<b>Step 5. Plan data collection routines</b> .....	<b>28</b>
<b>Step 6. Plan for data discussions</b> .....	<b>29</b>
<b>References.</b> .....	<b>32</b>
<b>Appendix A. Tools</b> .....	<b>A-1</b>
<b>Appendix B. Examples from Oklahoma Excel networked improvement communities.</b> .....	<b>B-1</b>
<b>Appendix C. Overview of the Southwest Networked Improvement Communities Research Partnership</b> .....	<b>C-1</b>
<b>Boxes</b>	
1 Key terms .....	13
2 Sources to explore related research and existing instruments. ....	21
<b>Exhibits</b>	
1 Driver diagram .....	7
2 The Plan-Do-Study-Act cycle .....	9
3 Steps to develop practical measures and related activities and tools .....	15
6.1 Types of graphs .....	31

## Contents

B1	Early Childhood Education Networked Improvement Community example from Tool 1A	B-4
B2	Early Childhood Education Networked Improvement Community example from Tool 1B	B-4
B3	Early Childhood Education Networked Improvement Community example from Tool 1C	B-5
B4	Early Childhood Education Networked Improvement Community example from Tool 4B	B-7
B5	Early Childhood Education networked improvement community example from Tool 5	B-10
B6	Early Childhood Education Networked Improvement Community example from Tool 6B	B-11
B7	Early Childhood Education Networked Improvement Community example from Tool 6B	B-12
B8	Early Childhood Education Networked Improvement Community example from Tool 6B	B-14

# INTRODUCTION

---

This toolkit is designed to guide educators in developing and improving practical measurement instruments for use in networked improvement communities (NICs) and other education contexts in which principles of continuous improvement are applied. Measures for continuous improvement should be closely aligned to student learning goals and implementation of instructional strategies driving the continuous improvement effort, and they should be practical to use in a classroom setting. Continuous improvement includes distinct repeating processes: understanding the problem, identifying specific targets for improvement, determining the change to introduce, implementing the change, and evaluating if and how the change led to improvements (Langley et al., 1996).

This toolkit is intended to help a team of educators with the final step in the cycle, which includes collecting data to measure implementation of changes and intended outcomes and using those data to inform future action. A team of educators can use this toolkit to proceed through a series of steps to identify what to measure, consider existing instruments, draft instruments, evaluate and refine instruments, plan data collection routines, and plan for data discussions to interpret the data and inform action. [Appendix A](#) includes supporting tools associated with each of the steps.

Regional Educational Laboratory (REL) Southwest developed resources in the toolkit in partnership with the Oklahoma State Department of Education (OSDE) team working with the [Oklahoma Excel \(OK Excel\)](#) NICs. Through the Southwest Networked Improvement Communities Research Partnership, REL Southwest has supported OSDE's OK Excel project with a series of coaching and training projects. These projects were designed to build state capacity for implementing content-area NICs at the district level to test and scale up high-impact, evidence-based instructional strategies selected by the NICs. The projects were also intended to increase state and district understanding of improvement science and the use of data for continuous improvement. Examples of how the toolkit resources are applied in the OK Excel NICs' work are in [appendix B](#).

# OVERVIEW OF THE TOOLKIT

---

This toolkit provides resources and guidance for educators to develop classroom-based practical measurement instruments for use in networked improvement communities (NICs) or other continuous improvement contexts such as professional learning communities. The toolkit is intended for a team of educators that has already identified specific student learning needs and strategies to improve instruction to address those needs and are ready to test these strategies using continuous improvement processes. The toolkit will empower educators to develop instruments such as short surveys, checklists, rubrics, exit tickets, and quizzes to measure improved instruction and student learning and that are practical to implement in the classroom. The goal of these “practical measures” is to provide data that a team of educators can examine together to inform their efforts to continuously improve instruction and student learning. By applying the processes described in this toolkit, educators can design practical measures that will provide accurate and actionable data to guide their improvement efforts.

*Unlike traditional measures (for example, accountability tests), practical measurement is data collection that is used both as a summative tool after a change idea is implemented and as a formative tool to check progress and change course during the implementation of a new change idea (that is, instructional strategies or classroom practices aimed at improving student outcomes).*

NICs and other continuous improvement initiatives are concerned primarily with building a system that is constantly seeking ways to improve how it tackles its consistent problems of practice. This kind of system relies on easy to collect, accurate, and usable feedback mechanisms that can come from practical measures. The primary goal of the toolkit is to support educators in understanding the characteristics of quality measurement and how to apply that understanding to develop and use practical measures. The processes described in the toolkit derive from the field of improvement science as well as from practices used to develop instruments and measures for other research purposes.

This toolkit includes a brief introduction to NICs, continuous improvement, and principles of high-quality measurement. However, it assumes a baseline understanding of the main concepts of continuous improvement processes such as Plan-Do-Study-Act (PDSA) cycles.<sup>1</sup> The main sections of the toolkit describe a process to develop measures that are practical for this context and that will provide useful data as the NIC team explores its implementation and related outcomes.

---

1. You can learn more about PDSA cycles in the infographic, Using Improvement Cycles to Promote Equitable Instructional Practices: Oklahoma Excel (<https://ies.ed.gov/ncee/rel/Products/Resource/90120>).

## Overview of the toolkit

This toolkit includes the following sections:

- **Using the tools** provides suggestions about users and contexts for which the toolkit would be appropriate and how teams in different situations might implement sections of the toolkit.
- **Practical measurement in the context of continuous improvement** provides background information about improvement science, NICs, practical measurement, and principals of high-quality measurement. Key terms are defined in [box 1 on page 13](#).
- **The six steps of practical measurement** describes processes to identify what to measure and how to consider existing instruments, draft instruments, evaluate and refine instruments, plan data collection routines, and plan for data discussions. Each section contains information about the step, suggestions for activities that teams can work on together to make progress on the step, tools to support that progress, and links to additional resources.
- [Appendix A](#) includes supporting tools associated with each of the steps.
- [Appendix B](#) includes descriptions of how these tools have been used and examples of practical measures created by NICs that have worked with the Regional Educational Laboratory (REL) Southwest NICs partners in the Oklahoma State Department of Education (OSDE).
- [Appendix C](#) includes background information on the Southwest Networked Improvement Communities Research Partnership between REL Southwest and OSDE.

# USING THE TOOLS

---

## Who should use this toolkit?

This toolkit is intended for educators working in continuous improvement contexts within their schools and districts. The resources are designed to be used by a team of educators that has already identified specific student learning needs and strategies to improve instruction to address those needs and are ready to systematically test these strategies. The purpose of the toolkit is to support developing and improving the instruments used to collect data to measure improvements in instruction and student learning and then to use the data to measure the success of the improvement strategies. The toolkit describes processes to develop, improve, and implement practical measures in a team setting and includes supporting resources for each step.

Improvement teams or networked improvement communities (NICs) can comprise educators from schools, districts, and state education agencies, and depending on the focus of improvement, NICs can also include parents and community members. This toolkit is intended to support teams or NICs across multiple contexts. New NIC teams may use these tools to find, adapt, or create measures. NIC teams that are up and running may use these tools to improve and refine existing practical measures, build team capacity to use practical measures effectively, and demonstrate the quality of existing practical measures and the data they yield to stakeholders external to the NIC. Below are some examples of the many potential contexts for use of this toolkit.

### School contexts

- A group of grade-level subject area teachers who are implementing and testing a new instructional practice or curriculum material through Plan-Do-Study-Act (PDSA) cycles and who need support in developing multiple practical measures to assess improvement beyond the interim and summative assessments available to them, such as periodic progress assessments and annual statewide tests.
- A principal who wants to participate in a new districtwide school climate initiative and wants teachers in her building to implement it. The principal envisions a model in which she trains a core set of lead teachers on NIC practices and continuous improvement processes to drive implementation of the initiative in the school. The principal is familiar with NICs and improvement science, but she needs support on practical measurement, including how to create instruments to collect data specifically aligned to implementation of the initiative and intended outcomes.

### District contexts

- A district-level administrator who oversees the district's five elementary schools and is in year 3 of a NIC implementation aimed at improving reading scores. To enhance the way that teachers in the NIC prepare and use data displays, the administrator wants to



## Using the tools

set up professional development and planning sessions with the teachers focused on data displays.

- Math department chairs at three district high schools who seek to improve algebra scores through PDSA cycles and who want to ensure that the teachers in their departments show interrater agreement when they use the practical measure rubrics they have developed. The department chairs need protocols and process steps to use with teachers as they test and measure interrater agreement on the rubrics.

## Regional or state contexts

- A state education agency's curriculum and instruction office that needs to develop new practical measures for use as part of a NIC initiative to increase use of evidence-based instructional strategies in participating schools. Office staff leading the initiative want both to use existing measures and to develop new surveys and rubrics as part of the initiative in order to ensure that measures are well aligned to the new instructional strategies.
- Staff at a regional office of education who are piloting a technology intervention to increase student attendance in a cohort of low-performing schools in the region and who need to develop and test short surveys for use in PDSA cycles to understand how the intervention is being implemented and perceived. The staff have some experience creating surveys but want to test the surveys before using them so that they can be confident in the quality of the data they collect to measure implementation.

## Where to start

NIC teams may engage with this toolkit at different stages depending on their individual contexts. NIC teams should consider where they are in the continuous improvement process as it pertains to measurement. This toolkit is organized around six steps:

1. Identify what to measure.
2. Consider existing instruments.
3. Draft instruments.
4. Evaluate and refine instruments.
5. Plan data collection routines.
6. Plan for data discussions.

Each section contains information, activities, and a tool or tools that team members can use to complete that step and move to the next. A team that has completed a [driver diagram](#) (see [box 1 on page 13](#) for definitions of this and other key terms) and is ready to implement and test a new strategy but that does not have a plan for collecting needed data will start at step 1. A team that has already selected or developed measures and wants to evaluate them could start at step 4. A team that is satisfied with its data collection instruments but would like guidance for planning data conversations can find support in the last section, step 6.

## How to implement the toolkit

Educators using these tools can implement them in several ways depending on where they need the most support. NIC organizers, leaders, and participants can:

- Develop a measurement plan for a new NIC using the toolkit resources as guidance.
- Extract content for training NIC teams before they develop measures.
- Provide ready-to-use templates for educators in the NIC as job aids to reference and complete as they develop and test measures.

[Appendix B](#) provides examples from REL Southwest’s partners in the OSDE to illustrate how the Oklahoma Excel (OK Excel) teams applied the processes and resources for their NICs. The examples provided here illustrate how the OK Excel NIC leaders employed the processes described in this toolkit to create and use data collection instruments given their goals and context. **The OK Excel instruments included in the appendix are not intended to be adopted as is.** Other teams of educators should carefully consider the instruments’ alignment to their specified aim, theory of action, and local context.

# PRACTICAL MEASUREMENT IN THE CONTEXT OF CONTINUOUS IMPROVEMENT

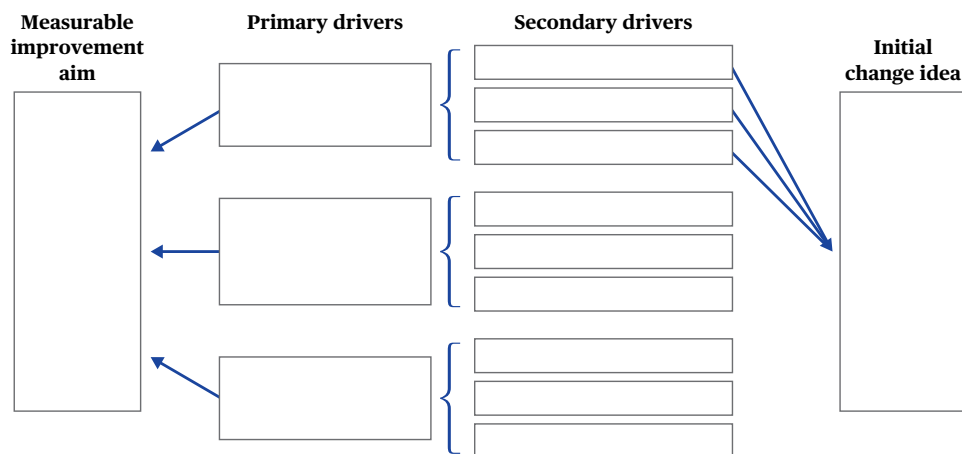
Users of this toolkit should have background knowledge of improvement science and networked improvement communities (NICs). This section provides a brief overview of these topics and the fundamental ideas of practical measurement.

## What is improvement science?

Improvement science is a systematic approach to implementing and testing strategies to solve problems and drive change in complex systems (Improvement Science Research Network, 2016). Educators using an improvement science approach begin by defining the problem or student learning need to be addressed, consider the teacher and student factors that are related to the need, determine a strategy to drive improvements, and collect evidence about whether the strategy is working. A driver diagram, like the one in exhibit 1 and in [appendix A](#), is a common tool in improvement science for displaying the “drivers” hypothesized to influence progress toward solving the problem and achieving an “improvement aim.” Each driver represents a hypothesis about a change essential to improvement and may include details about what the driver means. Together, the primary drivers offer an overview of the landscape for change.

A series of secondary drivers sit under each primary driver and comprise the strategies hypothesized to activate the primary drivers. Each secondary driver can have extensive details about how to implement strategies, and these strategies make up a coherent “change idea” to be tested and revised through a process of continuous improvement. For example, a NIC may identify “teachers’ capacity to implement instructional practices” and “students’ perceptions of their ability to be successful in math class” as primary drivers because they hypothesize that creating change in these areas could create positive change

**Exhibit 1. Driver diagram**



## Practical measurement in the context of continuous improvement

in student outcomes. Under those primary drivers, the NIC might identify “weekly professional learning community meetings” and “math mentoring with older students,” respectively, as secondary drivers that are specific strategic levers for developing a change idea.

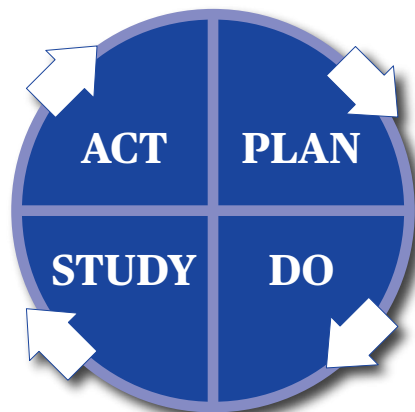
Plan-Do-Study-Act (PDSA) cycles are often used as a model for the process of continuous improvement. The phases of a PDSA cycle are described in exhibit 2.

### Learn more about improvement science and Plan-Do-Study-Act cycles

*[Learning to Improve: How America’s Schools Can Get Better at Getting Better](#)*: This book by leaders from the Carnegie Foundation for the Advancement of Teaching is considered a foundational text on the application of improvement science to solve persistent education problems or practice (Bryk et al., 2015).

*[Continuous Improvement in Education: A Toolkit for Schools and Districts](#)*: Produced by REL Northeast & Islands, this toolkit provides an overview of the implementation of PDSA cycles in schools, along with resources that educators can use to implement continuous improvement initiatives (Shakman et al., 2020).

Exhibit 2. The Plan-Do-Study-Act cycle



**Plan:** During this phase, the team of educators will develop the foundation for the work, including developing a driver diagram. The team targets a problem (for example, low student achievement or engagement), identifies its root causes using strategies such as driver diagrams, and identifies an intervention or change in practice that is expected to drive improvement based on research evidence. This change idea does not need to be radical; it can be a simple tweak in practice. The change should directly address root causes of the problem and be motivated by a theory about why the change will help move targeted outcomes toward their desired levels. Team members develop a plan for implementing the change in their practice. In addition, they develop a plan to measure implementation of the change and the expected outcomes. The instruments used to collect data should be well aligned to the developed theory of improvement and should be practical—that is, the instruments should produce data that clearly answer questions identified in the NIC team’s hypotheses about its problem and change idea and that can be incorporated as part of regular classroom routines without too much burden (Bryk et al., 2015). Developing instruments for these purposes is a focus of this toolkit.

**Do:** Next, educators implement the planned changes over a set period of time, and data are collected to measure implementation and expected outcomes.

**Study:** After implementing the intervention for a specified period—typically several weeks—team members reconvene to analyze the data collected and interpret the findings. Team leaders should provide easy-to-understand visualizations of the data and findings, and participants review these graphic displays summarizing the data and discuss their observations, questions, interpretations, and potential next steps for implementing the change idea.

**Act:** During the final phase of the PDSA cycle, team members act on what they learned during the Study phase. They decide how to proceed with the change idea. The team decides whether to *adopt* or *expand* the idea if the data suggest that the change was implemented as intended and improvements were realized, *adapt* the change idea if the data suggest necessary changes that could be made to improve the change idea, or *abandon* the change idea altogether if the data suggest that improvements were not realized and try something else.

The cycle is then repeated by beginning another planning phase. Each PDSA cycle builds on a prior cycle, providing participants with the opportunity to continuously refine their change idea or “stack” or combine the initial change idea with additional change ideas until they achieve the initial aim.

---

## What are networked improvement communities?

NICs are continuous improvement initiatives that bring together and engage educators and researchers in an applied research process to solve a clearly specified, high-priority problem that they are empowered to address through their shared work. NICs apply the processes of improvement science and continuous improvement described above. Educators are well positioned to determine what changes they can implement to drive improvement toward a specific goal, called an *aim*. NIC members leverage expertise and working relationships between researchers and educators to complete PDSA cycles so that shared testing of interventions and communication of findings within NICs informs learning at scale (Bryk et al., 2011). The NIC model also posits that when these continuous improvement questions are considered in a networked setting, there is the potential for accelerated learning.



### Learn more about networked improvement communities

[\*Getting Ideas into Action: Building Networked Improvement Communities in Education\*](#): This essay from the Carnegie Foundation for the Advancement of Teaching makes the case for improvement science as another approach to conducting research to solve education problems.

## What is practical measurement?

Measurement for continuous improvement efforts differs from measurement for theory development, school accountability, or student assessment. Practical measurement involves a team of educators (for example, a NIC) implementing a specific change in practice, actively collecting data using instruments embedded in that practice, then using that data to inform iterative refinement of that practice (Yeager et al., 2013).

In general, practical measures:

- Are linked to high-impact, attainable improvement goals.
- Orient educators to aspects of classroom learning environments associated with student learning (such as daily routines, instructional formats, and dynamics of teacher-student and student-student interactions), thereby serving as levers for—as well as measures of—improvement.
- Feature data collection and analysis routines that are relatively undemanding and easy to interpret.
- Accurately measure observed elements of instruction, thereby producing data that reflect what happened in a classroom.

## Practical measurement in the context of continuous improvement

- Provide data that can be examined by the team to monitor and evaluate progress toward goals for implementing the change idea with fidelity and improving teacher and student outcomes (Yeager et al., 2013).

Practical measures can be instruments such as short surveys, observation checklists, rubrics, exit tickets, or quizzes that are practical to implement in the classroom. See the discussion of [different kinds of instruments](#) for more information, and see examples of [instruments used by the OK Excel Early Childhood Education NIC](#) in appendix B.

As NIC teams identify and adopt practical measures, they should consider the following criteria:

- The instruments should be aligned to the theory of improvement so that the resulting data will help educators understand whether the theory is right, whether there is improvement, and whether changes are needed for the next cycle.
- For each element to be measured (for example, student engagement or teacher self-efficacy), measures should be short and easy to complete (Proger et al., 2017).
- The instruments should be easy for educators to embed in normal practice and provide data that are timely and interpretable.
- Practical measures enhance the “networked” aspects of NICs by enabling collaborators to learn together by sharing well-developed measures they can use to compare experiences and develop a collective wisdom across the NIC. As smaller groups of educators within a NIC develop measures that are working well for them, they may expand the use of the measures to include additional educators and contexts. By doing so, they may realize greater benefit from a larger set of implementation contexts and resulting data than by continually adapting measures in the original, smaller group.



### Learn more about practical measures

- [\*Practical Measurement\*](#): This resource published by the Carnegie Foundation for the Advancement of Teaching describes how practical measurement differs from typical measurement in education research.
- [\*Practical Measures, Routines and Representations\*](#): Known as PRM2, this is a partnership of researchers and district educators who are developing and testing a set of practical measures for middle school math interventions.
- [\*On the Development of Content-Specific Practical Measures Assessing Aspects of Instruction Associated with Student Learning\*](#): This paper from the National Center on Scaling Up Effective Schools describes criteria for developing practical measures in education settings and presents examples from the field.

## What are the principles of high-quality measurement?

Measurement is a process for classifying or quantifying attributes, such as physical properties, skills, behaviors, and attitudes. Instruments such as assessments, surveys, and rubrics are commonly used in education to measure many constructs relevant to teaching and learning (for example, student learning, teacher practices, and student or teacher attitudes). Validity and reliability are the two main concepts when considering the quality of an instrument.

**Validity** is the extent to which inferences made based on the data are accurate and meaningful for their intended use. In this case, validity refers to how well a measure aligns to key elements of the construct. Validity can be evaluated by experts who review the content of the measure and determine whether it covers all relevant aspects of the construct. (For example, does a survey about a specific teaching practice include all the important aspects of that practice?) Validity also can be evaluated by looking at how well data collected from a selected measure align to data from a well-established measure of the same construct. (For example, do the data from a short classroom measure of kindergarten students' phonemic awareness correlate highly with results for those students on a larger assessment of the same skill?)

**Reliability** is the extent to which a measure produces consistent data under different conditions and for different people. Reliable measures produce dependable results when they are unambiguous, so that all respondents (or raters) interpret items similarly. For example, interrater reliability refers to the extent that raters scoring student essays interpret the scoring protocol similarly so that an essay would receive the same score regardless of who rated it. Another type of reliability is called internal consistency reliability and refers to how well a set of items (for example, survey items or assessment items) measure the same thing and therefore can be combined to create a single scale measure. Internal consistency reliability is evaluated by examining how well the values on the set of items correlate with each other. (For example, do people with high values on one of the items tend to have high values on the other items in the set?)

## Developing high-quality practical measures

**A single instrument may provide high-quality data for one purpose but not for another.** For example, a grade 3 math assessment might provide valid and reliable school-level data for school accountability purposes but not useful formative data about individual student performance that could inform instruction. When thinking about the quality of measures used in classroom continuous improvement efforts, it is important to consider the purposes of measurement in improvement science as follows:

- Examine implementation fidelity—a measure must provide a picture of how teachers in the NIC implemented the change so they can compare implementation data and understand how differences in implementation may have affected the change idea and student outcomes.



## Practical measurement in the context of continuous improvement

- Assess change— a measure must be sensitive to short-term changes to allow NIC teams to understand whether their predicted change occurred.
- Predict outcomes—a measure must accurately capture information on changes in key processes that are expected to link to improvement aims based on the theory of improvement to ensure that measures answer the research questions on which those predicted outcomes are based.
- Inform educator actions—a measure must be easily administered and analyzed with clear implications for improvement so that teachers in the NIC can decide to adopt, adapt, or abandon tested changes based on the data it yields.

The process for developing measures and testing those measures to collect evidence of validity and reliability can be extensive. High-stakes instruments (such as assessments used for school accountability, statewide surveys used to inform resource allocations or other policies, or rubrics used to rate dimensions of teacher–student interactions to measure the causal impact of a professional development program) undergo extensive development, review, pre-testing, and analysis to verify the validity and reliability for their purposes. The development and testing processes can be time- and resource-intensive, and the resulting instruments can be lengthy and burdensome to administer and analyze. It is possible, however, to create high-quality measures that also are *practical* for use in continuous improvement settings. A key tenet of this toolkit is that teams can create instruments that are high-quality by following a systematic process that borrows techniques from traditional instrument development efforts—such as cognitive interviewing and interrater reliability testing—in a way that a team of educators with limited time and resources could feasibly implement.

### Box 1. Key terms

Below are key terms used throughout this toolkit to describe roles and processes. Some key terms have been selected from the glossary of *Learning to Improve: How America's Schools Can Get Better at Getting Better*, adapted for education contexts, and are used throughout the toolkit (Bryk et al., 2015).

**Change idea.** An alteration to a school or classroom system or a process or instructional intervention that is to be tested through a Plan-Do-Study-Act (PDSA) cycle to examine its efficacy in improving some driver(s) in the working theory of improvement. Examples are a new protocol for meeting with students who request help, an evidence-based teaching method aligned to algebra skill gaps, and implementation of graphic organizers in a writing class.

**Driver diagram.** A tool that visually represents a group's working theory of practice improvement. The [driver diagram](#) creates a common language and coordinates the efforts among the many different individuals across a district or school joined together in solving a shared problem.

**Improvement aim.** An improvement effort that answers the following question: What are we trying to accomplish? Improvement aims should clearly specify how much improvement and for whom and should clearly articulate how a networked improvement community (NIC) will determine progress toward or attainment of the aim. Improvement aims are often related to changes in leader or teacher practice or student performance.

**Improvement team.** The core group of educators, leaders, and, in some instances, content experts and researchers who carry out the work of the NIC. They identify the problem and the improvement aim, select the related change idea, develop the tools and artifacts necessary, and then implement and measure the change idea. For the most part, this report uses the terms *team*, *improvement team*, *NIC*, and *NIC teams* interchangeably.

**Instrument.** The measurement instrument for collecting the data that will be used to answer the research questions (for example, a survey, a short assessment, a log, checklist, or rating protocol). An instrument might include multiple items or might consist of a single item.

**Item.** A single question, rating scale, or other single element on an instrument.

**Network hubs.** A core group formed either centrally in an education agency or school or distributed across network leaders and teachers that carries out critical functions necessary for the support and effective operations of a NIC. These functions and capacities include improvement science expertise, analytics, knowledge management, professional development, implementation support, convenings, communications, and technological support. The hub makes the network work for the educators engaged in the PDSA cycle.

**Practical measurement.** As discussed in the [“What is practical measurement?”](#) section, practical measurement consists of frequent, agile tools embedded in school and classroom routines and practices to generate data to inform improvement. These tools may be referred to as *measures*, *instruments*, or *assessments* based on their type, purpose, and the data they yield. Because the intent is to inform continuous improvement, practical measures are collected frequently to assess whether positive changes are occurring. Because practical measures are focused on and used by educators, the measures are framed in language that is natural and comprehensible to teachers and students.

# THE SIX STEPS OF PRACTICAL MEASUREMENT

The following sections describe the six steps to develop and use practical measurement in continuous improvement: identify what to measure, consider existing instruments, draft instruments, evaluate and refine instruments, plan data collection routines, and plan for data discussions (exhibit 3). Each section contains information about the step, suggestions for activities that a team can work on together to make progress on the step, tools to support that progress, and links to additional resources.

**Exhibit 3. Steps to develop practical measures and related activities and tools**

Step	Activity	Tool to support activity (see appendix A)
1	Identify what to measure	1A. Start with completed driver diagram and change idea description
		1B. Write research questions
		1C. List data elements and instruments needed to answer your questions
2	Consider existing instruments	2A. Review the literature and identify relevant instruments
		2B. Create a bank of potential items mapped to research questions
3	Draft instruments	3A. Learn about and practice writing instrument items
		3B. Review drafted instruments
4	Evaluate and refine instruments	4A. Collect reviewer feedback
		4B. Conduct cognitive interviews
		4C. Evaluate interrater reliability
5	Plan data collection routines	5. Plan for timing and logistics of data collection
6	Plan for data discussions	6A. Create data displays
		6B. Plan data inquiry activities

## Step 1. Identify what to measure

### 1A. Start with completed driver diagram and change idea description

Before developing practical measures to use in your Plan-Do-Study-Act (PDSA) cycles, you will want to make sure that you have consensus among members of your team about your driver diagram, including the aim statement, primary and secondary drivers, and the elements of the change idea (for example, specific classroom activities and strategies) that will be implemented to achieve the expected improvements. To start the work around practical measures, your team will have already had to develop a driver diagram.



#### Learn more about driver diagrams

- Driver diagram: <https://carnegienetworks.zendesk.com/hc/en-us/articles/115001189967-Driver-Diagram>
- Vermont Department of Education driver diagram template: <https://education.vermont.gov/sites/aoe/files/documents/Driver%20Diagram.pdf>
- Institute for Healthcare Improvement driver diagram: <http://www.ihl.org/resources/Pages/Tools/Driver-Diagram.aspx>



**Supporting tool:** Your team can use a diagram that has already been developed, or you can adapt [Tool 1A. Template: Driver diagram and checklist](#) to record your aim statement, primary and secondary drivers, and notes about your change idea. The tool includes a checklist to assess the completeness of your driver diagram.

### 1B. Write research questions

To clarify what should be measured, networked improvement communities (NICs) will want to identify what questions to answer during the Study phase of the PDSA cycle. The final driver diagram will provide the information you need to develop specific research questions. You will want to spend time thinking about which

aspects of the change idea, drivers, and aim you want to test first and what you might want to test in another PDSA cycle.<sup>2</sup> Consider what questions are most important to help your team understand what is working and make decisions about what should be changed for the next cycle of implementation and testing. Research questions should describe what

*NICs should prioritize questions about how the change idea was implemented, the status or change in the drivers, and the student or educator outcomes the NIC has as a measurable improvement aim.*

2. If needed, refer to [Continuous Improvement in Education: A Toolkit for Schools and Districts](#) (Shakman et al., 2020) to plan for and prioritize the changes to test in your NIC.

## Step 1. Identify what to measure

you want to know and about whom. This step can be completed as a group activity during a team meeting, with a designated leader facilitating the discussion.

Your team will want to consider the following types of questions:

- **Implementation questions** are about how the change idea was implemented. These questions may be about how well or how often the teacher integrated the elements and activities of the change idea into classroom activities or about whether there were roadblocks to implementing the change idea, such as inadequate resources. These kinds of questions are sometimes called process or formative questions. Implementation questions might refer to implementation during a specific period or a change in implementation over time, or they might explore differences in implementation between groups (for example, did the implementation of a new instructional strategy occur with the same frequency among grade 1 and grade 2 teachers?).
- **Outcome questions** address the extent to which the change idea was associated with improvements in the primary drivers, student or teacher behaviors, attitudes or knowledge, or skills. Continuous improvement efforts often focus on short- and medium-term outcomes in a single or small series of PDSA cycles. Longer-term outcomes may be measured annually or after a longer series of PDSA cycles over an appropriate amount of time to fully implement the change and realize results. Like implementation questions, outcome questions might refer to a single time period or change over time, or they might address differences between groups.



### Learn more about writing research questions

[Program Evaluation Toolkit](#), Module 2, Chapter 2: How to write quality evaluation questions, is from a Regional Educational Laboratory Central toolkit about program evaluation. This chapter describes how to write questions for program evaluations.



**Supporting tool:** Your team can use [Tool 1B. Discussion guide: Writing research questions](#) to help identify research questions.

## 1C. List data elements and instruments needed to answer your questions

Once your research questions are finalized, you will want to identify the data needed to answer the questions, determine the best source for those data, and consider the timing of data collection to answer the research questions.

*Keep in mind that the most successful measurement strategy in a rapid-cycle improvement effort involves short instruments that are easy to administer and record data, so be sure that your team is prioritizing the most critical data elements.*

## Step 1. Identify what to measure

**Your team will want to consider the following about each research question:**

- What specific data elements are needed to answer the question?
- Are any related data currently being collected that could be accessed (for example, attendance records)?
- For data not already collected, what is the best type of instrument to collect these data (for example, a student survey, a rating scale to evaluate instructional materials, a log to count classroom behaviors, or a set of assessment items)?
- Could comparisons with other groups of students, teachers, or classrooms help you gauge whether the change is having its intended effect?
- Can you combine data from different perspectives that help you assess the effect of a change?
- When and how often should each data element be collected?

**Consider different kinds of instruments that could be used to collect the needed data:**

- **Surveys** are best suited for collecting data directly from respondents about their experiences, behaviors, beliefs, or attitudes.
- **Teacher surveys** can collect information directly from teachers about their implementation of the change idea and other related behaviors or attitudes or about their perceptions of student behaviors and learning outcomes. These surveys can be short, including only one or two items, or can consist of just a checklist with yes/no responses.
- **Student surveys** can collect information directly from students about their own behaviors and attitudes or their perceptions of their teacher's implementation of aspects of the change idea.<sup>3</sup> Some topics are best understood through multiple perspectives, so you might want to include similar items on teacher and student surveys (for example, items about how often the teacher implements a practice could be reported on by the teachers and the students). However, it is important to write surveys designed for students at an appropriate reading level for all students. Surveys can have multiple items or may include only one item. Surveys can be administered on paper, online, or by some other method. For example, students can answer a simple question on their way out the door (such as how much of the lesson did they understand or how helpful was teacher feedback that day) by dropping a card in one of three pockets labeled with answers.



**Yes**



**Some**



**No**

---

3. When collecting data, especially from students, apply appropriate guards to protect student identity. Ensuring anonymity can encourage candid answers, and following through with strict data security procedures will ensure that student privacy is protected.

## Step 1. Identify what to measure

- **Rating protocols** are best suited for collecting data by a rater who uses the protocol to systematically score the actions, environments, or products on a predefined scale, such as quality or completeness. Rating protocols to score multiple characteristics or dimensions are also called rubrics. Scale points should capture meaningful variation within each dimension (typically using a 3-5 point scale). Scales using fewer scale points may be easier for raters to complete quickly (which would be more practical), but scales with more scale points may more precisely capture differences and changes over time.
- **Observation rating protocols** can collect data about teacher or student behaviors or interactions between teachers and students. Rating protocols might be useful to collect data about teachers' implementation of an instructional practice critical to the change idea or about a student's performance related to a student learning outcome.
- **Document rating protocols** can collect data about an artifact, such as a teacher's lesson plan, aligned to implementation of the change idea, or data about student written work related to a student learning outcome.
- **Logs** to track occurrences of behaviors may be a practical way to frequently collect data relevant to the implementation of a change idea or indicators of outcomes. Logs may be an activity count of how frequently a teacher or student does something. For example, a teacher might tally how often he initiates classroom greetings or how often a student participates in cooperative work.
- **Assessments** are best suited to collecting data directly from students related to aspects of students' learning outcome. Assessments can be created to collect data aligned to the desired student learning outcome. A practical student assessment might take the form of a short quiz or exit ticket (for example, a single question for all students to answer and hand in at the end of class).



**Supporting tool:** The improvement team can use [Tool 1C. Table: Data map](#) to identify data needed to answer each research question, what kinds of instrument to use to collect the data, and timing considerations for when and how often the data should be collected.

## Step 2. Consider existing instruments

### 2A. Review the literature and identify relevant instruments

Once you have completed step 1 to identify what you want to measure and the kinds of instrument needed to collect the data, consider what instruments might already exist that you can use or adapt. Your team could begin with the same research literature and resources for evidence-based practices that you examined when determining your change idea and improvement efforts. Measures used in that related research may be useful. For example, is there academic research that describes implementation of your change idea or that has tried to find a causal relationship between the problem you are addressing and your change idea? How was the implementation of the change idea measured? How have changes in the drivers been measured, and how have related student attitudes, behaviors, or learning goals been measured?

Assessments, surveys, observation protocols, and other instruments used in formal research contexts are typically longer than would be appropriate in a rapid-cycle improvement effort, but you might want to adapt some elements of longer instruments to your purposes. You will want to look for evidence that the measures are high quality and practical for your purposes. Consider whether the items have been used with a population similar to yours (in grade level and demographics, for example). You will want to consider any published evidence of the validity and reliability of the instrument but understand that evidence for an entire scale (set of items) may not apply if you are using only selected items or modifying the items substantially. It may be helpful to review the information about validity and reliability in the section “[What are the principles of high-quality measurement?](#)”

Sources to find research related to your change idea and other sources to locate well-developed existing instruments are in [box 2](#).

This step in the development process can be performed by your team members over the course of two meetings. At one meeting, you can discuss possible sources your team is aware of (including those found in [box 2](#)) and select ones to explore before the next meeting. Your team members can bring what they find to the next meeting for group discussion.



**Supporting tool:** Your team members can use [Tool 2A. Checklist: Considering existing instruments](#) to guide considerations about how potential instruments align to the content and context of the change idea and elements of their research questions.

### 2B. Create a bank of potential items mapped to research questions

After examining and selecting possible instruments that could be used or adapted, you can organize all the identified items by research question for easy consideration. An item could be a survey item, rating scale, checklist, log, assessment item, or other unit of data collection. This can include items identified in step 2A that already exist but are not currently



## Box 2. Sources to explore related research and existing instruments

---

### Sources to explore related research

[What Works Clearinghouse](#). The What Works Clearinghouse (WWC) reviews research on programs, products, practices, and policies in education. [This video](#) covers how to use WWC.

[Education Resources Information Center](#). This resource, known as ERIC, is an online library of education research and information sponsored by the Institute of Education Sciences of the U.S. Department of Education. [This video](#) discusses how to search ERIC.

### Sources to explore existing instruments

[National Center for Education Statistics](#). This resource, known as NCES, is the primary federal entity for collecting and analyzing data related to education. NCES has many instruments to collect data from students, teachers, school staff, administrators, and parents about a multitude of education topics across their survey and program areas.

[EdInstruments](#). This resource is a curated and organized collection of measurement tools spanning domains, including academic knowledge and skills, student well-being, teaching, and school climate. The instruments in this database measure a variety of outcomes for children from birth through post-secondary education, parents, educators, administrators, and schools. The database includes links to free and publicly available information about the validity and reliability of each instrument, as well as peer-reviewed studies using the instruments.

[Compendium of Student, Teacher, and Classroom Measures Used in NCEE Evaluations of Educational Interventions](#). This report from the National Center for Education Evaluation and Regional Assistance can help evaluators and researchers select outcome measures for their future studies and assist policy-makers in understanding the measures used in existing Institute of Education Sciences studies.

[ED School Climate Surveys \(EDSCLS\)](#). The National Center on Safe Supportive Learning Environments has a suite of school climate surveys for middle and high school students, instructional and noninstructional staff, and parents and guardians.

[Measuring Student Engagement in Upper Elementary through High School: A Description of 21 Instruments](#). This report reviews the characteristics of 21 instruments measuring student engagement in upper elementary school through high school. It summarizes what each instrument measures, describes its purposes and uses, and provides information on its psychometric properties.

[A Review of Instruments for Measuring Social and Emotional Learning Skills among Secondary School Students](#). This resource was designed to support state and local education agencies in identifying reliable and valid instruments for measuring collaboration, perseverance, and self-regulated learning among secondary school students.

[University of Washington PMR2 site](#). This resource has practical measures for middle school math.

## Step 2. Consider existing instruments

used for students in your school (for example, survey items from a study that examined the change idea your networked improvement community is interested in implementing) or items that exist and are already collected for students in your school, such as daily attendance records. Once the items are compiled in this way, your team can:

- Prioritize items for each research question.
- Identify which items will need to be modified to better match the research questions.
- Identify gaps where new instruments or items will need to be developed to fully answer the research questions.

During this step, team members exploring existing measures for potential items can use an item bank to record the items they find. Then, your team can create a combined item bank and use it to inform the discussion of which items could be adapted or adopted.



**Supporting tool:** The improvement team can use [Tool 2B. Template: Item bank](#) to organize existing items for each instrument to be developed. This template can be used to compile potential items from existing surveys, checklists, and assessments that the team will want to consider.

## Step 3. Draft instruments

The third step is to draft the instruments (using the prioritized items in your item bank) and to add new items where needed. For items related to implementation research questions, you will want to consider how you could describe different levels of implementation (percentage of teachers, frequency, intensity, duration, quality defined by a standard, number of students involved, or something else). For items related to an expected outcome, you will want to consider what level of attainment is meaningful (percentage of students achieving a given level, frequency, intensity, duration, quality of a classroom practice or student behavior defined by a standard, or something else).

As described above in step 1C, “[List data elements and instruments needed to answer your questions](#),” you may be drafting one or more of the following types of instruments:

- Student surveys.
- Teacher surveys.
- Checklists.
- Logs.
- Document rating protocols.
- Observation rating protocols.
- Assessments.
- Exit tickets.

You can complete this step with the full team working together to draft all needed instruments or with smaller groups working separately on each instrument.

### 3A. Learn about and practice writing instrument items

It is critical to write survey items and assessment items following best practices to ensure that they yield reliable data. The team leader should review learning resources related to different kinds of instruments and plan to share them with the team before it drafts the instruments.



#### Learn more about writing survey items

[Survey Methods for Educators: Collaborative Survey Development \(part 1 of 3\)](#): This Regional Educational Laboratory (REL) Northeast & Islands guide describes a five-step

### Step 3. Draft instruments

collaborative process that educators can use with other educators, researchers, and content experts to write or adapt questions and develop surveys for education contexts.

#### [Workshop on Survey Methods in Education Research: Facilitator’s Guide and Resources:](#)

This REL Midwest toolkit consists of a facilitator guide and workshop handouts. The toolkit is intended for use by state or district education leaders and others who want to conduct training on developing and administering surveys.



### Learn more about assessments

- [NWEA assessment resources](#) provide an overview of classroom assessment issues and strategies.
- [University of Washington test construction website](#) provides guidance on writing high-quality assessment items.
- [Walker Center for Teaching and Learning](#) at the University of Tennessee at Chattanooga’s website provides examples of vetted classroom assessments and provides support for writing new items.
- Rhode Island Department of Education guidance:
  - [Guidance for Developing and Selecting Quality Assessments in the Primary Classroom](#)
  - [Guidance for Developing and Selecting Quality Assessments in the Elementary Classroom](#)
  - [Guidance for Developing and Selecting Quality Assessments in the Secondary Classroom](#)

After reviewing learning resources about writing effective survey items and before developing draft items, the team leader may plan a group activity using tool 3A to practice what the team learned.



**Supporting tool:** After reviewing resources about writing survey items, the team can use [Tool 3A. Worksheet: Activity to practice writing instrument items](#) to practice putting the best practices into action. This activity provides practice to detect problems in survey items and suggest revisions. Team members can work on these individually or in small groups for 10-15 minutes before sharing answers with the full team. Include time to review and discuss the answer key after each item.

### 3B. Review drafted instruments

After drafting instruments, the team should set aside time to review the instruments together against the quality criteria in the [Tool 3B](#) checklist before moving to the next step. Before using the Tool 3B checklist, it may be helpful to review the information about validity and reliability in the section “[What are the principles of high-quality measurement?](#)”



**Supporting tool:** The team can use [Tool 3B. Checklist: Criteria for high-quality practical measurement](#) to evaluate the instruments against quality criteria.

## Step 4. Evaluate and refine instruments

The next step is to collect evidence that your measure will elicit reliable and valid data that answer your research questions and inform your conversations about whether there is improvement in teaching and student learning. This section includes tools to support three types of pretesting activities: collecting reviewer feedback, conducting cognitive interviews, and evaluating interrater reliability.



### Learn more about evaluating instruments

- [Workshop on Survey Methods in Education Research: Facilitator’s Guide and Resources](#) is a Regional Educational Laboratory (REL) Midwest toolkit. Module 4 includes information about pretesting survey items.
- [Survey Methods for Educators: Collaborative Survey Development \(part 1 of 3\)](#) is a REL Northeast & Islands toolkit on writing and testing survey items.

### 4A. Collect reviewer feedback

Reviewers may provide valuable insights about how best to capture important constructs in your measures, what important items might be missing, or how to improve the wording of the instrument. For example, the validity of an instrument to measure student learning can be verified through expert review of the instrument in light of the learning outcomes or standards it is intended to measure. Depending on the context, local curriculum experts might include district- or school-based instructional coaches or content-area leaders, consultants working with a district or school, and regional or state education agency experts. For a teacher survey, you also might want to collect feedback from teachers about the wording of the items. The likelihood of collecting useful feedback will be improved if you provide your reviewers with information about the intended purpose of the instrument and a set of questions.



**Supporting tool:** Use [Tool 4A. Template: Reviewer feedback instructions](#) to help your team write targeted reviewer feedback instructions and questions.

### 4B. Conduct cognitive interviews

A common method for evaluating drafted survey items and other instruments is to conduct cognitive interviews, which are one-on-one interviews that are designed to find out how respondents understand, interpret, and respond to the items. The respondent completes the survey items with the interviewer present. The interviewer asks the respondent in real time to explain how they came up with their responses to each item and asks additional probing questions to uncover any misconceptions or areas that need more clarity (Willis, 2005).

#### Step 4. Evaluate and refine instruments

Cognitive interviews can include:

- A concurrent “think-aloud” process during which the participant is asked to verbalize what they are thinking as they consider and select answers to questions.
- Unscripted probes that the interviewer might ask in response to something the participant says or does. For example, the interviewer might ask a teacher participant who seems to be having difficulty deciding between two rating points in scoring a student’s work to explain how they interpret the two ratings and what made it difficult to choose.
- Scripted probes are prepared in advance to target predetermined potential problems. For example, if the team has drafted a survey item that it thinks might be hard to understand, it may prepare a question such as this: “In your own words, what is this question asking about?”

Cognitive interviews are often used to test survey items but can also be used to test any kind of instrument that requires someone to respond to questions or follow instructions to provide information (such as rating scales, checklists, or assessment items).

You will want to recruit different kinds of respondents for cognitive interviews so that you can gather a variety of perspectives. For example, if you are testing teacher survey items, you will want to include new and veteran teachers and teachers from relevant grade levels. In general, three or four interviews can provide adequate feedback for an instrument used in a continuous improvement context when time and resources are limited; additional interviews may be beneficial if time and resources allow.

Notes taken during the cognitive interviews can be used to inform revisions to improve the clarity and wording of the items. To ensure a complete record of what was said during the interviews, the team should assign someone to take notes or audio record the interviews (with respondents’ permission).

The results of the cognitive interviews will provide insight into how well the items are interpreted as intended. Instruments that are clearly and accurately understood by each type of potential respondent will provide more reliable data. Common misconceptions about an item across multiple interviews would indicate a strong need to improve clarity, but revisions might be considered even if a single interviewee encounters a problem. The team should spend adequate time reviewing the results to reflect on the implications of the feedback and make decisions about potential revisions.

To help the team understand how to conduct and use cognitive interviews, the team leader should plan a learning activity for the team using the mock interview script provided in [tool 4B](#). After the team acts out or reads the script, the team leader can lead a discussion around the questions, answers, and any potential changes to the instrument based on the interview results.



**Supporting tool:** The team can have a group activity using [Tool 4B. Activity: Learn about cognitive interviews](#) to act out or read a mock cognitive interview between a participant and interviewer. The team can then discuss the interviewer’s technique and how the results of the interview might inform item revisions.

## Step 4. Evaluate and refine instruments

### 4C. Evaluate interrater reliability

Measures that include rating protocols (for example, to evaluate student work, teacher behavior, or instructional materials) or that rely on rater (or respondent) judgment should be checked for interrater reliability. Interrater reliability refers to the degree that different raters assign the same rating to identical observations. Disagreement might occur if the raters are not well trained in using the protocol or if the protocol does not include well-defined descriptions of the rating scale. Rating protocols should closely align to the construct being measured and offer clear descriptions of the rating scheme and each rating point, so that all raters will apply the scoring consistently. Rating scale points should describe objective, observable aspects of the behavior or material being rated rather than subjective descriptions—such as “the instructor implemented the lesson well”—which would be open to varying interpretations. If a checklist or rubric is not interpreted and used consistently by all raters, the resulting score data will not be reliable and will offer little value to the team.

Before implementing a rating protocol, you should conduct an activity with the raters to check for evidence of interrater agreement and detect and discuss potential revisions needed to improve the protocol if the activity shows inconsistencies between raters.



**Supporting tool:** The improvement team can use [Tool 4C. Worksheet: Evaluate interrater reliability](#) to have multiple raters score the same observation (for example, a student’s work or a teacher’s lesson plan), calculate interrater agreement, and examine reasons for discrepancies. The data and discussion from this activity can inform revisions to the protocol.

## Step 5. Plan data collection routines

### 5A. Plan for timing and logistics of data collection

The timing and logistics for administering each practical measure should be planned well in advance to minimize burden and ensure consistency across classrooms. You should include repeated measures across multiple timepoints to address questions about improvement over time. Networked improvement communities (NICs) may structure their measurement routines in different ways. NICs implementing relatively simple change ideas may use the same measure over multiple Plan-Do-Study-Act (PDSA) cycles. Some NICs may use a larger-scale measure aligned to the aim statement at the start and end of a series of PDSA cycles and use a more focused set of measures within the PDSA cycles aligned to the discreet changes in a change package that builds toward that ultimate outcome.

A measurement plan is key to creating routines that ensure that measurement activities occur on the same schedule and in the same way for all participants. The measurement plan is organized by instrument.

For each instrument, the plan will describe:

- Who will collect the data and from whom. (These might be the same person.)
- What action steps are needed to ensure that the instrument is administered and that data are recorded correctly.
- When and how frequently measurement will occur (for example, daily, at the end of each week, every two weeks, at the beginning and end of the cycle) and a plan to ensure that there is sufficient time in the schedule to collect the data.
- What other resources, training, and instructions related to data collection will be needed.
- Plans to obtain parental consent for student surveys or other new data collection from students, if needed.

Teams may introduce new measures incrementally as they build a change idea over multiple PDSA cycles. As the measurement plan evolves, NIC teams should also ask the following questions about their measures and data collection processes:

- Will you have data that can be linked across students, teachers, and timepoints in a way that will allow you to answer your research questions?
- Are you promising confidentiality to those providing survey data? Should you? What impact might that have on data quality?



**Supporting tool:** TheRegardless of the model in use, a NIC team can use [Tool 5. Template: Measurement plan](#) to plan data collection routines for the PDSA cycles. This template represents a simple project planning format for measures that can be adapted based on the model in use.



## Step 6. Plan for data discussions

### 6A. Create data displays

During the Study phase of the Plan-Do-Study-Act (PDSA) improvement cycle, you will want to ensure that the team is guided by clear and organized representations of the data that are designed to inform actionable discussion about your research questions. People tend to understand information better when it is presented visually. Data visualizations are most effective when they accurately reflect the data, using clear labels and uncluttered design and the graph type that is best suited to the data and the questions that you want the data to help answer (Evergreen, 2017).

Typically, a subset of the networked improvement community (NIC) who are experienced working with data and graphics will plan for and create the data displays to share with the full NIC team. To begin planning for data displays, organize your data by research question. Identify which data (instruments and item numbers) will help answer each research question. Some questions may be answered by just one item at one timepoint, and others may be answered by multiple items within an instrument, across instruments, or across timepoints. For example, a question about how often the teacher implemented an element of the change idea might be examined by looking at data from a teacher's self-report log, a student survey, an observer's checklist, or a combination of all three. Data about student academic outcomes may come from student assessments and one or more items on a teacher survey.

Group all data sources that relate to each research question. Data from separate instruments should be shown in separate data displays, except for instances where an item has exactly the same wording and response options (or scale) across instruments. For example, if the same item is asked of teachers and students, you could display a summary of both sets of responses in the same graphic clearly labeling the teacher and student data.

For each research question and set of related data, make notes about how best to display the data to facilitate NIC members' exploration. Consider what kinds of display are appropriate. For example, if you are looking at change over time, graph the values with consistent scales so that comparisons are easy to see. Include the exact wording of a survey item with the data and use labels for response categories.

You may want to calculate raw totals, percentages, or an average value across one or more items. Averages are appropriate only for items with response options that are ordered and can be assumed to have equal distance

*Include the **response rates** for each item in your data displays. Response rates are the number of respondents with item data expressed as a percentage of the number of total possible respondents (the intended population). The response rate is one way to gauge whether survey results represent the intended population:*

- A high response rate maximizes the chance that the results are representative of the population.*
- A low response rate increases the chance of biased results, which cannot be generalized to the population.*

## Step 6. Plan for data discussions

between each point. For example, an average response can be calculated for a survey item with points labeled strongly disagree, disagree, neutral, agree, and strongly agree (numbered 1-5) but would not be appropriate for a survey item with points labeled never, a few times a year, about once a month, about once a week. Be sure to include the range of possible values and scale definition so that average scores are interpretable. Include the number of responses and total possible number of respondents so that it will be easy to take into account nonresponse rates when interpreting the data.

Consider which types of graphs are best suited to each kind of data and research question (exhibit 6.1).



### Learn more about [creating graphs in Excel](#)



**Supporting tool:** The team can use [Tool 6A. Worksheet: Plan for data displays](#) to plan data displays best suited to the types of data and research questions being explored.

## 6B. Plan data inquiry activities

The PDSA cycle includes the Study phase in which NIC members examine and interpret the data to inform decisions to act on for further implementation of the change idea. Data inquiries support the data interpretation part of the Study phase. This step will include planned activities to examine and interpret the data for evidence of expected changes, considerations for how subgroup comparisons can help team members understand factors that may be related to changes, and examination of data from different perspectives for a fuller picture of the changes.

Discussions about how to interpret the data should be semistructured. It is helpful to have some structure and planned activities to make the most of data discussions for informing next steps and for discussing implications for continued implementation of the change ideas. Using a structured inquiry process can be key to building capacities for school- and classroom-level improvements (Copland, 2003; Timperley, 2008). Keep in mind, though, that continuous improvement efforts can yield unexpected discoveries about how the quantitative data on drivers interact and that unexpected research questions can arise from team members' experiences in implementing their change ideas.

Typically, the improvement team or any educators who join the NIC should participate in data inquiry discussions. The improvement team and participating educators may be supported by leaders who sponsor the NIC's activities; researchers who assist in data collection, presentation, and reporting; and content experts who consulted on the design and development of the NIC and its materials. These supporters might also participate in some aspects of the data inquiry discussions, but the primary determining factor for participation in decisionmaking is whether the person is directly affected by the condition the NIC is attempting to change and has the power to implement or enable any decisions resulting from the data inquiry. In some instances, it may be appropriate to include students in

## Step 6. Plan for data discussions

this process to add context based on their experiences participating in the change. When involving students, the improvement team must be cautious about protecting confidentiality, and when adults characterize student actions and outcomes and hypothesize reasons behind those outcomes, they should be careful in their use of language.

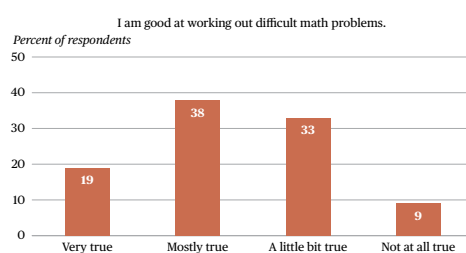
### Learn more about data inquiry

The [Practitioner Data Use in Schools: Workshop Toolkit](#) from Regional Educational Laboratory Northeast & Islands is designed to help practitioners systematically and accurately use

#### Exhibit 6.1 Types of graphs

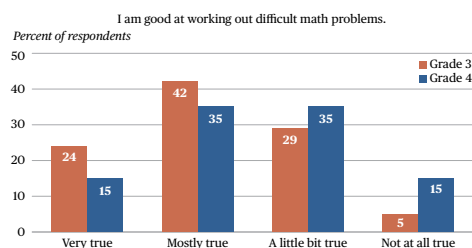
To examine a distribution.

Use a **bar graph** to show the distribution of student scores, observation ratings, or answers to survey items. For example, this bar graph summarizes hypothetical responses to a student survey item. The data are expressed as percentages but could also be displayed as frequencies.



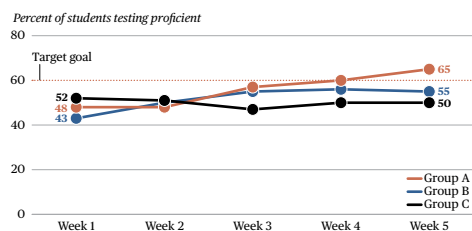
To compare distributions between groups.

Use a **clustered bar graph** to display group differences in distributions of responses. For example, this clustered bar graph shows hypothetical data from the survey item for two grade levels. Use percentages to make it easier to compare data between subgroups (in this case, grades 3 and 4).




To show trends over time.

Use a **line chart** to show how values change over time. For example, this line chart shows student performance over five weeks for three groups of students. To avoid cluttering the graph, only the first and last timepoints include data value labels. If there is a target goal related to these data, you can add a marker aligned to that goal to help with data interpretation. For example, if a target goal is for 60 percent of students to reach proficiency by the end of a five-week cycle, you could highlight and label the 60 percent line.



data to inform their teaching practice.

 **Supporting tool:** In this step, you can use [Tool 6B. Worksheet: Plan for data inquiry](#) to plan a process for sharing the data displays and develop questions in advance to guide the data inquiry process to include three phases: orient, explore, and interpret.

# REFERENCES

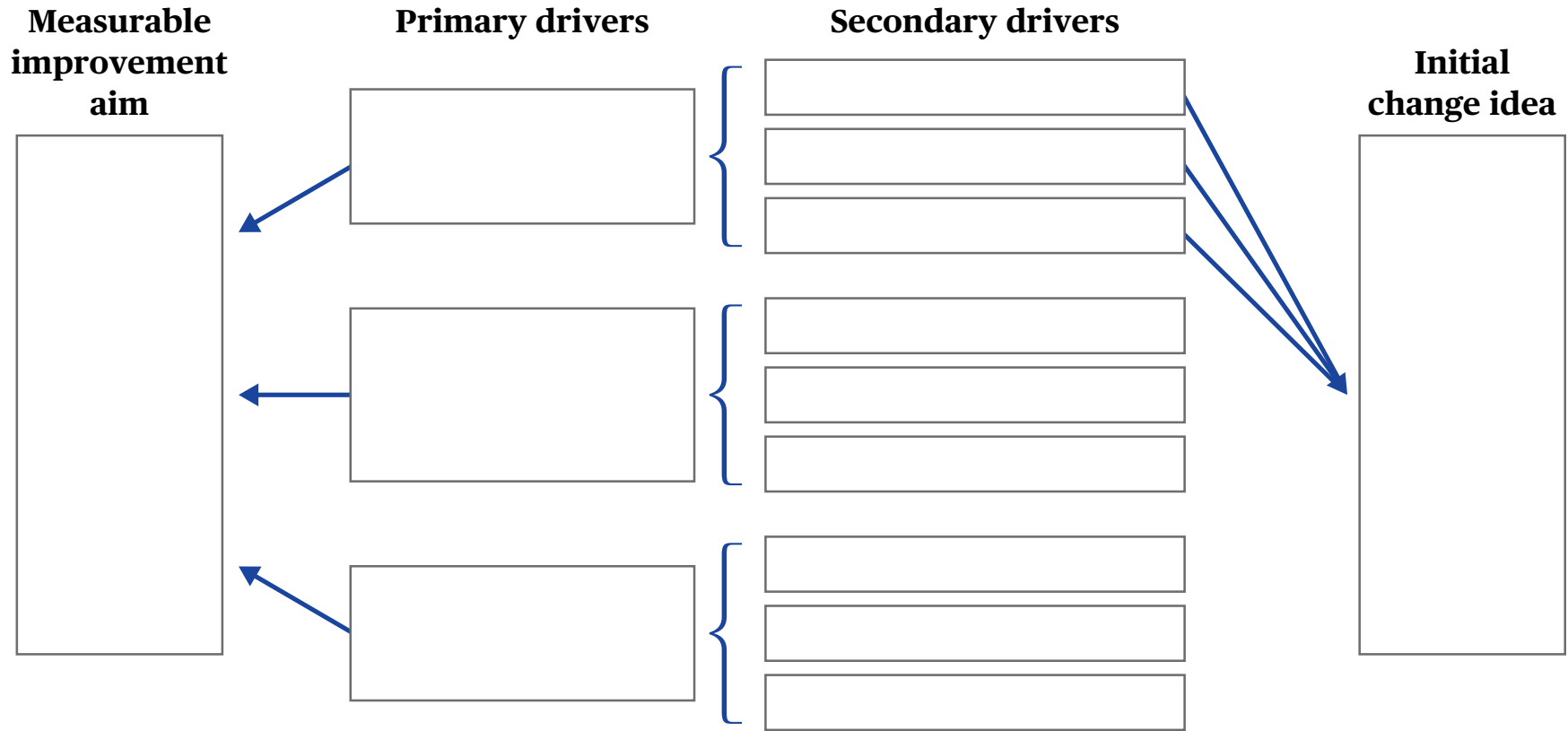
---

- Bryk, A. S., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. T. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 127-162). Springer.
- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.
- Copland, M. A. (2003). Leadership of inquiry: Building and sustaining capacity for school improvement. *Educational Evaluation and Policy Analysis*, 25(4), 375-395.
- Evergreen, S. D. H. (2017). *Effective data visualization: The right chart for the right data*. Sage.
- Improvement Science Research Network. (2016). *What is improvement science?* Retrieved October 16, 2021, from [https://isrn.net/about/improvement\\_science.asp](https://isrn.net/about/improvement_science.asp).
- Langley, G. L., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (1996). *The improvement guide: A practical approach to enhancing organizational performance*. Jossey-Bass.
- Proger, A. R., Bhatt, M. P., Cirks, V., & Gurke, D. (2017). *Establishing and sustaining networked improvement communities: Lessons from Michigan and Minnesota* (REL 2017-264). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. <https://ies.ed.gov/ncee/rel/Products/Publication/3853>.
- Shakman, K., Wogan, D., Rodriguez, S., Boyce, J., & Shaver, D. (2020). *Continuous improvement in education: A toolkit for schools and districts* (REL 2021-014). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. <https://ies.ed.gov/ncee/rel/Products/Publication/4005>.
- Timperley, H. (2008). Evidence-informed conversations: Making a difference to student achievement. In L. M. Earl & H. Timperley (Eds.), *Professional learning conversations: Challenges in using evidence for improvement* (pp. 69-79). Springer.
- Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Sage.
- Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement*. Carnegie Foundation for the Advancement of Teaching. Retrieved August 24, 2022, from <https://www.carnegiefoundation.org/resources/publications/practical-measurement/>.

# APPENDIX A. TOOLS

---

### Tool 1A. Template: Driver diagram and checklist



## Appendix A. Tools

After completing the driver diagram, you can use this checklist to ensure that the diagram is sufficiently detailed to support development of practical measures.

Driver diagram component	Requirements	Check
Measurable improvement aim	✓ A preset intended population.	<input type="checkbox"/>
	✓ A metric of interest.	<input type="checkbox"/>
	✓ A general timeline when the change should occur.	<input type="checkbox"/>
Primary drivers	✓ A factor or category of factors that influence the aim.	<input type="checkbox"/>
	✓ Contextual factors experienced by populations and stakeholder groups who influence the aim (students, teachers, families, and so forth).	<input type="checkbox"/>
	✓ Conditions that instigate the problem of practice and can influence the aim.	<input type="checkbox"/>
Secondary drivers	✓ Are linked to a specific primary driver.	<input type="checkbox"/>
	✓ Identify conditions or practices that can be changed or influenced by educators in a networked improvement community (NIC).	<input type="checkbox"/>
	✓ Can be translated into an actionable change idea.	<input type="checkbox"/>
Initial change idea	✓ Identifies a concrete, visible change from current processes, behaviors, or practices.	<input type="checkbox"/>
	✓ Is within the locus of control of the educators in the NIC and can be implemented consistently across educator contexts using readily available resources.	<input type="checkbox"/>
	✓ Is measurable in a practical way—rapidly, continuously, and during educators' regular activities and practices.	<input type="checkbox"/>

## Tool 1B. Discussion guide: Writing research questions

### Discussion points to clarify priorities for the research questions

#### Discussion points for implementation research questions

- What are the most important elements of the change idea?
- What would full and ideal implementation of the change idea look like? What specifically could be observed?
- Do we want to look at how well or how often a practice was implemented?
- Do we want to look at changes related to implementation over time?
- Do we want to be able to compare subgroups?

Select 1-4 critical questions about implementation and add below.

#### Discussion points for outcome research questions

- What are the most important aspects of the desired outcome(s) of the change idea?
- What are observable indicators that change has occurred?

Select 1-4 critical questions about outcomes and add below.

#### Considerations for the set of research questions

- Keep the total number of research questions to no more than four for each Plan-Do-Study-Act cycle. There may be exceptions, but generally you want to prioritize only the most important questions so that the measurement plan will not be too burdensome.

Draft research question	Implementation or outcome	Priority
1.		
2.		
3.		
4.		



## Tool 1C. Table: Data map

Fill in information about the data that you will need in order to answer each research question. Adapt the table so that you have enough rows for each data element that will contribute to answering each question.

**Data element:** Describe the specific data needed to answer the questions (for example, how frequently the teacher did something or evidence of student learning).

**Available data source:** What is the name of the database or system that contains the data (for example, attendance records). Add “na” if the information is not available in an existing data source.

**New data collection:** Indicate the kind of instrument needed to collect the data (for example, survey, checklist, rubric, or logs). See information about [different instrument types](#) for ideas. For implementation questions, consider who can best report on the implementation of the change idea (for example, teacher, student, or outside observer)? Do you need to consider implementation from multiple perspectives (for example, might you need a teacher survey and student survey to get an accurate picture)?

**Timing considerations:** Include notes about when existing data are collected and available or when new data should be collected (and how often, for questions about change over time).

Draft research question	Data element	Available data source	New data collection	Timing considerations
1.				
2.				
3.				
4.				

## Tool 2A. Checklist: Considering existing instruments

### Considering existing instruments

After reviewing the research literature or external sources for data collection instruments related to your research questions (see [box 2](#) in the toolkit), consider each existing instrument for its suitability for your measurement plan:

- How well does the content align to the needed data elements?
- What grade level was the instrument intended for?
- Has the instrument been used for students with similar demographic characteristics as yours?
- Is there published evidence about the validity and reliability of the instrument? Note if the reliability or validity evidence is for an entire scale (set of items) and whether you are using the full scale or only selected items, in which case the evidence may not apply.

Existing instrument	Degree of content alignment (low/partial/high)	Who has used the instrument? (For example, grade level or demographic characteristics)	Notes on validity and reliability evidence	Notes on overall suitability
1.				
2.				
3.				
4.				

## Tool 2B. Template: Item bank

### Item bank

After examining and selecting possible instruments in Tool 2A, you might find it useful to organize all the identified items by research question for easy consideration. For each research question record:

- Existing item (verbatim).
- Original source of the item.
- Networked improvement community (NIC) instrument: Name the instrument for which you may use the item (example, teacher survey, student survey).
- Notes: Record ideas for whether the item can be adopted as is for your instrument or how it could be adapted.
- Gaps: Identify any gaps in how well the existing items (whether used as is or adapted) cover the topic of the research question. Add notes about new items the team will need to write.

Add rows to the table as necessary.

Research question	Existing item	Original source	NIC instrument	Notes on whether item could be used as is or how it could be adapted	Gaps: What aspects of the research questions will not be covered by these items? Add notes on what new items will need to be drafted
1.					
2.					
3.					
4.					

## Tool 3A. Worksheet: Activity to practice writing instrument items

### Practice writing instrument items

Consider this short survey to be completed by grade 6 math students at the end of the week to gauge their perceptions of how well they understood the material and the teacher's use of feedback.

Each item is flawed. For each item, describe the problem and suggest a revision.

This week, math class was interesting, and I understand the new material. (*Select one.*)

- a. strongly agree
- b. agree
- c. disagree
- d. strongly disagree

*Problem:*

---

*Revision:*

---

How many times did the teacher ask me to explain how I got my answer? (*Select one.*)

- a. 1-2 times
- b. 3-5 times
- c. 5 or more times

*Problem:*

---

*Revision:*

---

My teacher always uses a positive tone when discussing my work with me. (*Select one.*)

- a. always
- b. sometimes
- c. never

*Problem:*

---

*Revision:*

---

When talking to me about my work, my teacher helps me comprehend and articulate the specific steps and processes I need to take to strengthen the quality of my work. (*Select one.*)

- a. always
- b. most of the time
- c. sometimes
- d. never

*Problem:*

---

*Revision:*

---

## Appendix A. Tools

### Answer key: Practice writing instrument items

This week, math class was interesting, and I understand the new material. (*Select one.*)

- a. strongly agree
- b. agree
- c. disagree
- d. strongly disagree

*Problem:* This item includes two ideas. This kind of double-barreled item is difficult to answer for a student who has different perceptions about each idea (for example, if the student disagrees that the class was interesting but strongly agrees that they understood the new material). The data from an item like this could be misleading.

*Revision:* This item should be revised to focus on just one of the ideas. If both ideas are critical, it should be rewritten as two separate items, such as:

This week, math class was interesting. (*Select one.*)

- a. strongly agree
- b. agree
- c. disagree
- d. strongly disagree

I understand the new material in math class this week. (*Select one.*)

- a. strongly agree
- b. agree
- c. disagree
- d. strongly disagree

An additional consideration would be to use a different rating scale than the agree/disagree scale. For example:

How interesting was math class this week? (*Select one.*)

- a. very interesting
- b. somewhat interesting
- c. a little interesting
- d. not at all interesting

How well do you understand the new material this week? (*Select one.*)

- a. very well
- b. somewhat
- c. a little
- d. not at all

How many times did the teacher ask me to explain how I got my answer? (*Select one.*)

- a. 1-2 times
- b. 3-5 times
- c. 5 or more times

*Problem:* The response options are not mutually exclusive; two categories include 5. There should be an option for zero. No timeframe is mentioned, so a student will not be sure how far back to count.

*Revision:* The response options could be rewritten

How many times did the teacher ask me to explain my work in the past week? (*Select one.*)

- a. 0 times
- b. 1-2 times
- c. 3-4 times
- d. 5 or more times

## Appendix A. Tools

### Answer key: Practice writing instrument items

My teacher always uses a positive tone when discussing my work with me. (*Select one.*)

- a. always
- b. sometimes
- c. never

*Problem:* The response options always, sometimes, and never are not a good match for an item that states the teacher always does something. For example, it does not make sense for a teacher to “sometimes” always use a positive tone.

*Revision:* The item could be reworded to ask about how often the teacher uses a positive tone, or the response options could include an agree/disagree scale.

My teacher uses a positive tone when discussing my work with me. (*Select one.*)

- a. always
- b. most of the time
- c. sometimes
- d. never

My teacher always uses a positive tone when discussing my work with me. (*Select one.*)

- a. strongly agree
- b. agree
- c. disagree
- d. strongly disagree

When talking to me about my work, my teacher helps me comprehend and articulate the specific steps and processes I need to take to strengthen the quality of my work. (*Select one.*)

- a. always
- b. most of the time
- c. sometimes
- d. never

*Problem:* The item is unnecessarily wordy. “Comprehend” and “articulate” may not be appropriate for the reading level of every student. You do not want reading ability to be a barrier to accurate understanding. Strive for simplicity without losing critical meaning.

*Revision:*

When talking to me about my work, my teacher helps me think about and explain how I can improve it. (*Select one.*)

- a. strongly agree
- b. agree
- c. disagree
- d. strongly disagree

## Tool 3B. Checklist: Criteria for high-quality practical measurement

Before you collect information to evaluate the instruments, review the set of instruments. Consider whether the set of instruments is valid for answering the research questions, reliable, and practical by considering each element in the checklist.

Quality	Criterion	Check
Valid	<ul style="list-style-type: none"> <li>✓ Strongly aligned to the research question.</li> <li>✓ Captures the most important aspects of change idea, drivers, and outcomes.</li> <li>✓ Items are written with enough variation in the responses to detect important differences and changes.</li> <li>✓ When combined across instruments, data will help the team answer each research question and discover connections between change idea implementation, drivers, and outcomes.</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Reliable	<ul style="list-style-type: none"> <li>✓ Items (for example, survey items, checklist items, rating scales, and log directions) are specific, clearly written, and well defined.</li> <li>✓ Items intended for students are age appropriate.</li> <li>✓ Rating protocol scale points are mutually exclusive. Scale point descriptions are distinct from each other so that every observation can fall into only one category.</li> <li>✓ Rating protocol scale points are collectively exhaustive. Scale points describe all possible variations on the dimension.</li> <li>✓ Directions for how and when to complete the instrument are clear and consistent across classrooms.</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Practical	<ul style="list-style-type: none"> <li>✓ Instruments are not too burdensome.</li> <li>✓ The timing and directions for the instrument can easily be embedded in classroom activities and routines.</li> <li>✓ Directions for how and when to complete the instrument are easy to understand.</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>

## Tool 4A. Template: Reviewer feedback instructions

### Directions for drafting reviewer feedback instructions

#### General introduction

Depending on how well the reviewer knows the team, write an appropriate introduction presenting the team and its mission. Name the instrument and its general purpose. Thank the reviewer for agreeing to review the instrument and specify the date by which you would like to receive comments.

#### Instrument details

Provide brief details about how, and how often, you plan to administer the instrument and how the team will use the resulting data. Explain how the items were developed or their source. Identify the characteristics of the people who will be filling out the instrument and the research question(s) the data will help answer.

#### Instructions

Explain what you hope to learn from the reviewer. Write a set of questions for the reviewer to consider that are relevant to the type of instrument to be reviewed. Questions can guide the reviewer to provide information helpful to improving the content, clarity, and feasibility of the instrument.

To confirm the validity of an instrument, you can ask the reviewer to consider how completely it covers essential content. For example:

- Are there important features about the quality of the lesson plan we are looking for that are missing from this rating protocol?

You can ask for feedback about how clear and understandable the instrument would be for intended respondents to help ensure the reliability of the resulting data. For example:

- Is the survey wording clear and appropriate for students in grade 3 and 4?
- Are the directions about how to complete the rating protocol easy to understand?

Reviewers can confirm that each response option is sufficiently distinct from other options. For example:

- Is there a clear basis for choosing one response (or rating) over another?

Questions can elicit feedback about how practical the instrument would be. For example:

- How long would it take you to answer this teacher survey?
- Do you foresee any problems with students completing this kind of exit ticket on their way out of class?
- Do you have any suggestions for how we could improve this?



## Tool 4B. Activity: Learn about cognitive interviews

### Learning activity: Cognitive interviews

#### Cognitive interview plans for survey items about safety

This cognitive interview plan is for two survey items about school safety.

- If students hear about a threat to school or student safety, they would report it to someone in authority.  
*strongly agree, agree, disagree, strongly disagree*

Planned probes:

- What did “someone in authority” mean to you? Who would students report it to?
- Can you tell me in your own words what this whole sentence is about?

- I felt safe at school all year.  
*strongly agree, agree, disagree, strongly disagree*

Planned probe:

- What you were thinking about when you were choosing your answer?

#### Example interview script

Read through the interview script. Then, discuss with the team what effective strategies the interviewer used, and what issues were revealed by the student’s responses to the items. In practice, you would want to do multiple interviews before deciding whether to make revisions, and you might want to perform multiple rounds of cognitive interviews depending on how much time is available and how extensively you revised the instrument. This activity will illustrate how you might use what you learn during an interview to identify potential areas of confusion in the survey items that might need to be improved.

**Interviewer:** *Thanks for helping me out today. We are planning to give a survey to students like you to find out what they think about things that happen at school. We just want to try out these sentences to make sure they are easy to understand.*

*I’ll be taking notes, and I’ll ask you to stop sometimes so that we can talk about what you were thinking as you respond to these statements.*

*Please read each statement and the answers out loud. Remember, you are doing this activity to help me understand if these are good questions for a survey or if they are confusing. This is not a test for you, and there are no right or wrong answers. I am just interested in what you are thinking as you read and respond to these statements so that I can help make them better if I need to. Your answers are private—I will not include your name with your answers.*

*Do you have any questions for me before we begin?*

**Student:** No.

**Interviewer:** *Okay—go ahead and read the first statement out loud.*

- If students hear about a threat to school or student safety, they would report it to someone in authority.  
*strongly agree, agree, disagree, or strongly disagree*

**Student:** I would say strongly agree—this definitely would happen.

**Interviewer:** *Okay, thank you. Now what did “someone in authority” mean to you? Who do you think students would report it to?*

**Student:** I was thinking we would tell the teacher or the principal if students were fighting or doing something dangerous like skateboarding on the stairs.

**Interviewer:** *Okay, can you tell me in your own words what this whole sentence is asking about?*

**Student:** It’s asking if students would tell if they see something dangerous.

**Interviewer:** *Okay, great, this is very helpful. Can you tell me what the word “threat” means to you in this sentence?*

**Student:** Well, I was thinking about safety. Sometimes students aren’t being safe, and we should tell the teacher. Did I get this one wrong?

**Interviewer:** *You are doing great. Thank you. Let’s move on. Go ahead and read the next statement.*

- I felt safe at school all year.  
*strongly agree, agree, disagree, or strongly disagree*

**Student:** Well, do you mean all this school year since I’ve been in fourth grade, or all of 2022?

**Interviewer:** *I’m just interested in what you think this means. How would you answer this one?*

**Student:** Okay—hmmmm. (Long pause.) I think I would say agree—well maybe disagree—but okay, agree.

**Interviewer:** *Can you tell me what you were thinking about when you were choosing your answer?*

**Student:** Well, I was thinking just about this school year because I think that’s what you meant by year. But I was thinking I feel safe in my classroom and in the school building, but sometimes I don’t feel safe on the school bus, so I wasn’t sure if I should think about that or not. I ended up answering about how I feel just in the school. Was I right?

**Interviewer:** *Thank you very much for sharing with me what you were thinking about these sentences. You did a great job helping me.*

### Learning activity: Cognitive interviews

#### Reflection

##### *Effective interviewer technique*

1. What words did the opening script include to help put the student at ease?
2. What words did the opening script include to convey the purpose of the task?
3. In item 1, why do you think the interviewer followed up with the unplanned prompt, “Can you tell me what ‘threat’ meant to you in this sentence?”
4. The student asked three questions. Explain why the interviewer’s responses were appropriate.

##### *Potential problems with the items*

5. What did you learn from this student about her understanding of the phrase “someone in authority” and “threat?”
6. What potential problems did this interview reveal about item 1? How might you reword the item to address this problem?
7. What two potential problems did this interview reveal about item 2? How might you reword it to address these problems?

#### Reflection answer key

##### *Effective interviewer technique*

1. What words did the opening script include to help put the student at ease?
  - The introduction explains what will happen during the activity to prepare the student for what to expect. “I’ll be taking notes, and I’ll ask you to stop sometimes so that we can talk about what you were thinking as you answer these questions.”
  - The introduction reminds the student that this is not a test. “This is not a test for you, and there are no right or wrong answers.”
2. What words did the opening script include to convey the purpose of the task?
  - The introduction describes the purpose of the activity twice so the student will understand why they are participating and what is expected of them. “We are planning to give a survey to students like you to find out what they think about things that happen at school. We just want to try out these sentences to make sure they are easy to understand.” “Remember, you are doing this activity to help me understand if these are good questions for a survey or if they are confusing.”
3. In item 1, why do you think the interviewer followed up with the unplanned prompt, “Can you tell me what ‘threat’ meant to you in this sentence?”
  - The student’s prior answers indicated that she was thinking about students’ unsafe behavior more than a possible outside threat to school or student safety. This unplanned follow-up prompt was asked to further explore the student’s understanding of the word “threat.”
4. The student asked three questions. Explain why the interviewer’s responses were appropriate.
  - When the student asked, “Did I get this one wrong?” the interviewer did not explain the possible misinterpretation, but simply said, “You are doing great. Thank you. Let’s move on to the next sentence.” There is nothing to be gained in that moment to explain where the respondent may have misunderstood the intended meaning of an item. Remember, it is the instrument being tested not the student. It is best to make note of the issue and keep the interview moving along with a positive tone.
  - When the student asked, “Well, do you mean all this school year since I’ve been in fourth grade, or all of 2022?” the interviewer did not reveal the intended meaning. The reply, “I’m just interested in what you think this means. How would you answer this one?” was an appropriate response to encourage the student to provide her own interpretation.
  - When the student asked, “Was I right?” after explaining her interpretation of the last item, the interviewer gave a positive response about the student’s helpfulness rather than respond about her answer. “Thank you very much for sharing with me what you were thinking about these sentences. You did a great job helping me.”

##### *Potential problems with the items*

5. What did you learn from this student about her understanding of the phrase “someone in authority” and “threat?”
  - The student appeared to understand the phrase “someone in authority” because she mentioned she would tell a teacher or the principal.
  - Her answers revealed that her understanding of the word “threat” in the context of the sentence focused on students acting in an unsafe way, which was not what the item developers had intended.
6. What potential problems did this interview reveal about item 1? How might you reword the item to address this problem?
  - The student did not interpret the word “threat” as intended, so the item should be revised with more direct language, such as, “If students hear about someone who might be planning to hurt people in your school, they would report it to someone in authority.”
7. What two potential problems did this interview reveal about item 2? How might you reword it to address these problems?
  - The student was not sure what “all year” meant in the sentence.
  - The student was not sure whether to include her time on the bus in how she felt “at school.”
  - To address these issues, the item could be reworded as “I felt safe traveling to and from school and in school this whole school year.” Or two separate items could be written to ask about traveling to school and being in school.

## Tool 4C. Worksheet: Evaluate interrater reliability

The goal of this activity is to examine interrater agreement for a group of raters who will be using a rating protocol in the context of a Plan-Do-Study-Act (PDSA) cycle. This activity could be used with any group of raters, such as teachers who will be rating student work, or classroom observers who will be rating teacher practices. The activity should be done after the raters have been introduced to the rating protocol but before the protocol is finalized and used in the PDSA cycle.

Why evaluate interrater agreement?

- Identify areas in the protocol that need revision or greater clarity.
- Identify training needs to help raters use the protocol more accurately.
- Provide useful feedback to raters.
- Generate evidence that the raters can use the rating protocol to provide reliable data during the PDSA cycles.

### Team leader prework

The team leader collects or creates a wide variety of samples, such as student written work or videos of teacher practices, using authentic examples when possible. Each sample is assigned a letter name. Copies are created, as needed.

### Activity setup

The team leader assigns each rater an ID number. Raters are paired, and the team leader assigns each pair the same five samples. Other rater pairs may receive the same samples depending on how many samples you have. Raters work independently (not in pairs) to review and rate each assigned work sample using the protocol.

### Individual raters review samples and record rating scores

Each rater puts their ID number at the top of the rating column and records their rating scores using whatever values are described by the protocol (for example, 1, 2, 3, or 4 for a 4-point scale).

Table to record individual rating scores

Sample	Rater ID # _____
	Rating
A	
B	
C	
D	
E	

## Appendix A. Tools

If the rating protocol has multiple dimensions (like a rubric), add rows in order to include separate ratings for each dimension. For example, if a rating protocol for student written work has one scale for content and another scale for writing conventions, you will want to include a row for each of these two dimensions:

**Table to record individual rating scores for rubrics with two dimensions**

Sample	Dimension	Rater ID # _____	
		Rating	
A	1	_____	_____
	2	_____	_____
B	1	_____	_____
	2	_____	_____
C	1	_____	_____
	2	_____	_____
D	1	_____	_____
	2	_____	_____
E	1	_____	_____
	2	_____	_____

## Calculate interrater agreement

When raters are done rating work samples, they will meet in pairs. They will create a table with both sets of ratings side by side. In the Agreement column, they will mark a “1” if the two ratings agree and a “0” if they do not. Add up the 1s and calculate the percentage of agreement out of the total possible agreements. This percentage is your interrater reliability.

**Table to compare rating scores for two raters**

Sample	Rater ID # _____	Rater ID # _____	Agreement (1 = yes, 0 = no)
	Rating	Rating	
A	_____	_____	
B	_____	_____	
C	_____	_____	
D	_____	_____	
E	_____	_____	
Total agreement points (number of agreements over total possible agreements)			__ / 5
Percentage agreement			

Note: This example is for a rating protocol with one dimension. If the protocol has ratings for multiple dimensions per sample, adjust the table accordingly.

**Table to compare rating scores for two raters, with example data**

Sample	Rater ID # _____ Rating	Rater ID # _____ Rating	Agreement (1 = yes, 0 = no)
A	3	3	1
B	4	5	0
C	2	2	1
D	1	1	1
E	5	4	0
Total agreement points (number of agreements over total possible agreements)			3/5
Percentage agreement			60%

## Discuss sources of disagreement in pairs

In this step, paired raters explore areas of disagreement and the reasons why. This discussion will reveal any rater misunderstandings or areas for needed improvement in the rubric.

### Discussion points

What was the percentage of agreement? If not 100 percent, where did we give different ratings for the same sample?

Why did we give different ratings? Was there a difference in our understanding of what the dimensions mean or in how to approach the rating task? What kind of training or preparation could have helped?

Why did we give different ratings? Did the sample seem to fit the descriptions for both ratings because of unclear descriptors or inadequate examples? What revisions could solve this?

Are there important features of the sample that are not measured by the protocol? Should we consider additional dimensions?

## Whole-group sharing and discussion

Have the pairs come together in the full group to share their findings and discussions. If multiple pairs rated the same samples, examine the agreement across all raters of common samples. Wherever there is less than 100 percent agreement, consider how rater training or rating protocol revisions could address the source of disagreement. As a whole group, discuss the following:

- How could the training or preparation to use the protocol be improved to increase inter-rater agreement?
- What revisions to the protocol could improve interrater agreement?
- If revisions are extensive, should we try this activity with a new set of samples to test the revised protocol?
- How confident are we that we will reliably collect data using this protocol?

## Tool 5. Template: Measurement plan

Developing practical, regular measurement routines helps a networked improvement community (NIC) team understand and plan for the required steps to collect data within Plan-Do-Study-Act (PDSA) cycles. The following template allows NICs to identify essential measurement elements and plot out when they will occur. Include instruments (such as surveys or rating protocols) and available data sources (such as attendance records) that were identified in the data map in [Tool 1C](#).

### Measurement plan

Add the instrument or available data source identified in the data element map (Tool 1C) in the first column. For each, identify:

- Related research question(s).
- Person or people implementing the measure.
- Action steps to ensure smooth coordination and correct implementation of the instrument. Plan to get consent for surveys, if needed. In addition, indicate what kind of identifiers are needed on the instruments, so that linking can occur as needed (for example, student surveys have a teacher ID number on them so student responses can be aggregated at the teacher level).
- When/how frequently measurement will occur (for example, daily, at the end of each week, every two weeks). If multiple PDSA cycles are planned, identify the cycles to be used.
- Identify any resources needed for implementation. For example, if the instrument data source is an online benchmark assessment tool, teachers implementing that measure will need to check out a class set of Chromebooks.

Add rows to the table as needed.

Instrument/data source	Related research question(s)	Who implements the measure	Action steps to prepare for measurement	When measurement occurs	Resources needed
1.					
2.					
3.					

## Tool 6A. Worksheet: Plan for data displays

### Considerations for data displays

Organize your data by the research questions that you have identified within each of the four areas in your driver diagram: change idea implementation, teacher attitudes and beliefs, student attitudes and beliefs, and student academic outcomes. Add rows to the table below, as needed, to match the number of questions you have for each area.

### Data sources

Within each area, map the data sources (instruments and item numbers) for each research question and enter these in the second column. Some questions may be answered by just one item at one timepoint, and others may be answered by multiple items within a measure, across measures, or across timepoints. For example, a question about how often the teacher implemented an element of the change idea might be examined by looking at data from a self-report teacher survey, a student survey, an observer's checklist, or a combination of the three. Data about student academic outcomes may come from student rubrics and one or more items on the teacher survey. Group all the data sources that relate to each question. You may update this table after each Plan-Do-Study-Act cycle as new data are available.

### Data display considerations

For each question and set of related data, make notes about how best to display the data to facilitate team members' exploration of the data. Consider what kinds of displays are appropriate. See [step 6A](#) in the main document about data displays. For example, if you are looking at change over time, graph the values with consistent scales so comparisons are easy to see. Include the exact wording of a survey item with the data, and use labels for response categories.

You may want to calculate an average value across a single item or across multiple items. Averages are appropriate only for items with response options that are ordered and can be assumed to have equal distance between each point. For example, an average response can be calculated for a survey item with points labeled strongly disagree, disagree, neutral, agree, and strongly agree (numbered 1-5) or for an open-ended question about how many days per week an educator implements a practice in their classroom. However, calculating an average response would not be appropriate for a survey item with points labeled never, a few times a year, about once a month, or about once a week.

Be sure to include the range of possible values and scale definition so that average scores are interpretable. Include the number of responses and total possible number of respondents so that it will be easy to see whether nonresponse might be a consideration when interpreting the data.



Appendix A. Tools

Map research questions to data sources and data display plans

Research question	Data sources (list instruments and item numbers)	Data display type (for example, bar chart or line graph)	How data will be displayed (for example, percentage distributions, averages, single time point, or longitudinally)
<i>Implementation</i>			
1.			
2.			
3.			
4.			
<i>Outcome</i>			
1.			
2.			
3.			
4.			

## Tool 6B. Worksheet: Plan for data inquiry

This handout will help you prepare to engage your team members to examine and use data that were collected during the Plan-Do-Study-Act cycles. It will be useful to have team members organize data around the questions you have about the change idea implementation and expected outcomes. You can use the table in [Tool 6A](#) to organize your data by research question and plan your data displays. Use the table in this current tool to prepare data discussion questions for each question and related displays.

### Plan data discussions

A structured inquiry process can guide team members to make predictions, explore and describe the data, and interpret the findings. This activity will also inform discussion about how to proceed with implementation of the change idea.

There are three steps to this process: orient, explore, and interpret.

**Orient.** Before displaying the data, prepare the team by sharing the research question(s) and related measures and items. Have the team discuss what they hope to learn and then make predictions for what they expect to find. For example:

- What would we expect to see if the change idea was implemented exactly as planned?
- What would we expect to see if improvements occurred as anticipated?
- What are you most curious about?
- What do you predict the data will tell us?

**Explore.** Display the related data, and ask probing questions to facilitate analysis of the data and description of the findings. For example:

- What stands out to you about these findings?
- What are the most important findings?
- What trends do you see over time (if multiple timepoints)?
- Do multiple sources have similar or different findings (if multiple respondents)?
- Do findings differ for certain subgroups of teachers or subgroups of students?

**Interpret.** Ask questions to elicit discussion about what the findings mean and how they can inform decisions about next steps.

- How can we summarize the findings considering our research question(s)?

## Appendix A. Tools

- Do any of these findings surprise you?
- Do we have evidence that we implemented the change idea as planned?
- What do any subgroup differences tell us about conditions or factors that might help or hinder implementation of the change idea?
- Do we have evidence that the change idea had the intended impact?
- What do any subgroup differences tell us about characteristics that might be associated with the outcomes?
- What might explain why we did not see the improvement we expected? Is there evidence that this might explain why we did not see the improvement we expected?
- What are the implications for next steps in implementing the change idea?
- What additional data should we collect for our next cycle?

You might find it useful to loop through the three steps separately for each research question and related data, building on what you learn with each new set of analyses.

Adapt the following table to include your research questions and tailored discussion questions for your team members. Use the final column to summarize team responses to the questions.

### Plan data discussion questions

Step	Discussion questions	Team responses
<i>Implementation research question</i>		
Orient		
Explore		
Interpret		
<i>Implementation research question</i>		
Orient		
Explore		
Interpret		
<i>Outcome research question</i>		
Orient		
Explore		
Interpret		

# APPENDIX B. EXAMPLES FROM OKLAHOMA EXCEL NETWORKED IMPROVEMENT COMMUNITIES

---

## Southwest Networked Improvement Communities Research Partnership

Since February 2019, the Oklahoma State Department of Education (OSDE) has engaged districts and schools across the state through Oklahoma Excel (OK Excel), an innovative professional development program. Teams of educators form networked improvement communities (NICs) to implement and test promising instructional practices through rapid improvement cycles known as Plan-Do-Study-Act (PDSA) cycles. This state program seeks to identify, promote, and celebrate innovative and effective programs in school districts across Oklahoma.

Learn more about the  
Southwest Networked  
Improvement  
Communities  
Research Partnership

- [OK Excel Blog](#)

Through content-area NICs, OSDE staff and participating districts identify a common problem and select an evidence-based strategy to address that problem, applying the principles of improvement science to conduct a series of PDSA cycles to measure progress and scale promising solutions. Regional Educational Laboratory (REL) Southwest has worked with OSDE through its Southwest Networked Improvement Communities Research Partnership (SWNIC) providing coaching aimed at supporting implementation of the NICs and continuous improvement processes, improving the professional development that OSDE provides to participating districts, and supporting their development and use of practical measures. (See appendix C for additional information about the SWNIC.)

## Practical measures in the Southwest Networked Improvement Communities Research Partnership

OK Excel serves as the NICs' hub; it develops the aims, change ideas, driver diagrams, and measures for teachers in the district- and school-level NIC teams to implement. In the 2019/20 and 2020/21 school years, OK Excel instructional specialists engaged NIC teams in review, refinement, and consensus building on their driver diagrams and the selection or creation of appropriately aligned measures. After implementation of measures through PDSA cycles, NIC teams are given the opportunity to suggest improvements to the measures. This allows opportunities for NIC members to contribute to the NIC, building buy-in to the process and ownership of the results.

Between 2019 and 2020, REL Southwest worked with OK Excel to refine its NIC model and improve implementation through a series of coaching projects. In 2020, REL Southwest

## Appendix B. Examples from Oklahoma Excel networked improvement communities

and OK Excel agreed that OK Excel would benefit from a transition of focus from NIC implementation to measurement. OK Excel began working with REL Southwest to improve its approach to practical measures. In 2020 and 2021, REL Southwest provided analytics coaching to OK Excel. REL Southwest supported OK Excel in examining the data collected in the prior year and then provided coaching on how OK Excel could improve its measurement and data collection processes to yield better, more actionable data. In parallel, REL Southwest provided training to OK Excel on principles of high-quality measurement and best practices for selecting, developing, testing, and using practical measures for its 2021/22 NICs. Through that joint work, many of the tools were piloted and refined for inclusion in this toolkit.

Here, we provide a few examples of how the NICs for math, English language arts (ELA), and science used some of the tools to support their work. In the next section, we provide a fuller example of how the Early Childhood Education (ECE) NIC created and used practical measurement.

- The OK Excel math NIC focused on teachers' use of modeling to foster student discourse and ability to explain and justify their mathematical ideas. The math instructional specialist used [Tool 1C. Table: Data map](#) to organize the data elements needed to answer research questions about teacher comfort and ability to select and use modeling tasks, student perceptions of math instruction, and student skill in mathematical justification and reasoning skills. The data map was the guide for developing the teacher survey, teacher modeling task checklist, student survey, and rubric to score student work used to collect data for this NIC.
- The OK Excel ELA NIC's long-term aim was to improve reading comprehension through a series of change ideas, including teacher selection of complex texts and effective use of reading conferences. One of the primary drivers toward this aim was student mindsets and beliefs in their ability to read complex texts. The ELA specialist used [Tool 2A. Checklist: Considering existing instruments](#) to record information about existing measures and to support her selection and adaptation of items in the Secondary Reading Attitudes Assessment (Tullock-Rhody & Alexander, 1980) for the student survey, and two resources to inform the creation of a text complexity rubric: a [text complexity toolkit](#) developed by TeachingBooks® and the book, *Text Complexity: Stretching Readers with Texts and Tasks* (Fisher et al., 2016).
- The science instructional specialist used these tools to improve surveys, rubrics, and exit ticket measures as the science NIC pursued its aim of improving students' ability to generate and evaluate evidence while arguing science claims. The science instructional specialist participated in a learning activity described in [Tool 4B. Activity: Learn about cognitive interviews](#) and then used that example to design a cognitive interview protocol to test the student survey about their teacher's use of the targeted instructional strategies.

## Practical measures in the Oklahoma Excel Early Childhood Networked Improvement Community

Here we provide an extended example of how the ECE NIC progressed through the six steps of practical measurement described in the toolkit. The OK Excel ECE NIC instructional specialist worked with three district teams, each led by a district improvement fellow, and eight teachers. The NIC's long-term aim was to increase preK-grade 2 students' positive attitudes toward school and academic outcomes through playful learning experiences in the classroom. The existing driver diagram included primary drivers toward this aim, including teachers' confidence in their ability to provide high-quality play experiences and teachers' knowledge and skills for implementing different types of evidence-based strategies. The initial expected outcome was improved student engagement and perceptions of school, which was predicted to lead to better overall academic performance. The OK Excel ECE instructional specialist served as the NIC leader and provided professional development activities to NIC members to learn about the research on effective use of play and apply best practices for planning and implementing high-quality play experiences connected to learning standards.

**For the PDSA cycle described here, the NIC wanted to measure teachers' self-reported mindset, knowledge, and skills related to the different types of play-based learning strategies and how frequently they were implementing these strategies, as well as students' positive engagement with school.** With support from REL Southwest and using processes and tools in this toolkit, the NIC leader and three improvement fellows developed instruments to collect data to inform their efforts to increase teacher confidence and understanding of effective play-based learning strategies, increase classroom time spent on effective strategies, and increase students' positive attitudes about school.

### Instrument development steps

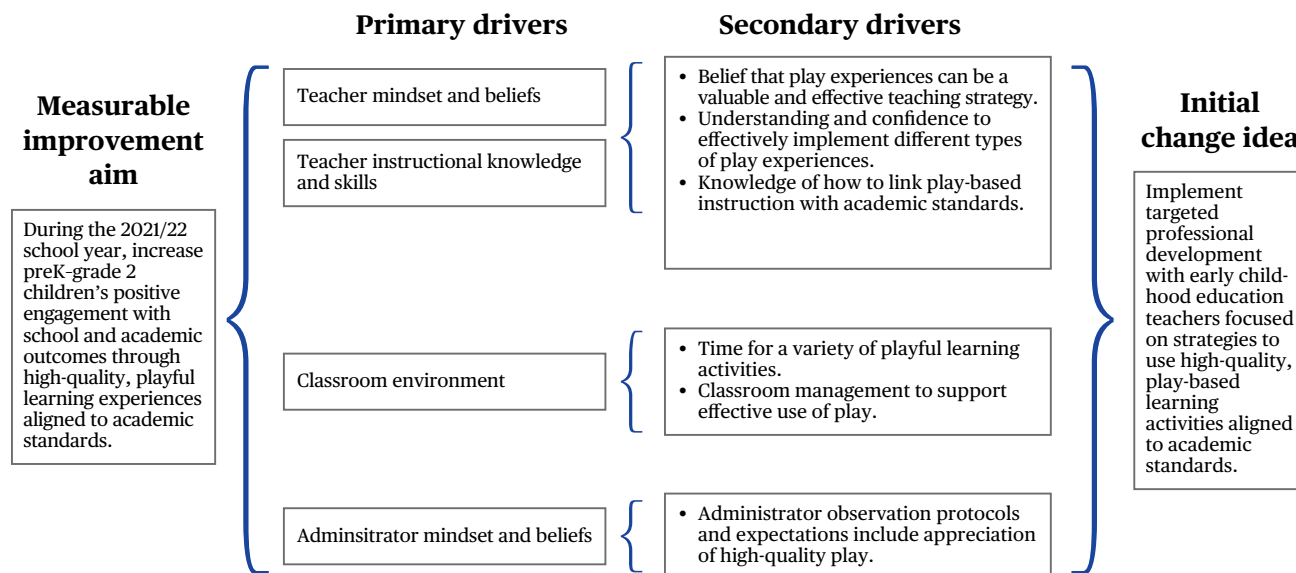
#### *Step 1. Identify what to measure*

The process of identifying what to measure started with the existing ECE NIC driver diagram (exhibit B1). Using the discussion questions in [Tool 1B: Discussion guide: Writing research questions](#), the NIC leader and team prioritized four key initial research questions related to teachers' growing understanding and implementation of different types of play-based learning activities, as well as children's perceptions of school (exhibit B2).<sup>4</sup>

---

4. Classroom management, administrator mindset and beliefs about the value of play-based learning, and student academic achievement were addressed in a separate PDSA cycle than is described in the rest of the examples in appendix B.

**Exhibit B1. Early Childhood Education Networked Improvement Community example from Tool 1A**



**Exhibit B2. Early Childhood Education Networked Improvement Community example from Tool 1B**

Draft research question	Implementation or outcome	Priority
1. How confident are teachers in their understanding and ability to implement effective play-based learning strategies?	Implementation	High
2. What challenges and successes do teachers have implementing play-based learning experiences, and what supports do they need?	Implementation	High
3. How much time are teachers spending on different kinds of play activities?	Implementation	High
4. Do students have positive attitudes about their learning experiences at school?	Outcome	High

Using [Tool 1C. Table: Data map](#), the ECE NIC leader identified the elements needed to explore these questions and four instruments to collect the data (exhibit B3).

**Exhibit B3. Early Childhood Education Networked Improvement Community example from Tool 1C**

Research question	Data element	Available data source	New data collection	Timing considerations
1. How confident are teachers in their understanding and ability to implement effective play-based learning strategies?	Understanding five types of play (free play, inquiry play, collaborative play, playful learning, learning through games)	na	Teacher survey	Measure at the beginning of cycle to help inform professional development and again at the end of the cycle to see if teacher confidence improves
	Creation of play opportunities that link to academic outcomes	na		
	Confidence facilitating children during play	na		
2. What challenges and successes do teachers have implementing play-based learning experiences, and what supports do they need?	Challenges	na	Teacher survey (open-ended questions) and classroom observation	Measure at least once in the beginning or middle of the cycle
	Successes	na		
	Needed resources	na		
3. How much time are teachers spending on different kinds of play activities?	Time for each of the five types of play	na	Teacher play audit	Measure at beginning, middle, and end of cycle to look at changes over time
4. Do students have positive attitudes about their learning experiences at school?	Attitude about school	na	Student survey	Measure at least at the beginning and end of the cycle
	Attitude about learning	na		
	Attitude about their teacher	na		

*Step 2. Consider existing instruments*

The NIC leader consulted recent literature on effective instruction to support academic achievement using play-based learning to inform the professional development provided to teachers and the content of the teacher survey. She looked at resources and research summaries about the benefits of play in early childhood from the Alliance for Childhood organization,<sup>5</sup> watched videos produced by REL Midwest related to play-based learning,<sup>6</sup> and looked through teacher and student surveys used in the national Early Childhood Longitudinal Studies<sup>7</sup> and other sources for possible survey items. For example, survey items used in a study about assessment in play-based kindergarten by Pyle and Deluca (2017) were borrowed in part for the development of the teacher survey.

5. <https://allianceforchildhood.org/publications-and-reports>.

6. <https://ies.ed.gov/ncee/rel/Products/Resource/10380>.

7. <https://nces.ed.gov/ecls/>.



### *Step 3. Draft instruments*

The ECE NIC leader learned about best practices for writing survey items and effective observation protocols and checklists through participation in REL Southwest training using materials included or referenced in this toolkit, including [Tool 3A. Worksheet: Activity to practice writing instrument items](#). The leaders of all four NICs supported by the OSDE NIC hub (ECE, math, ELA, and science) learned together and provided internal feedback to one another during this step to draft instruments using criteria included in [Tool 3B. Checklist: Criteria for high-quality practical measurement](#).

### *Step 4. Evaluate and refine instruments*

The ECE NIC leader participated in learning activities based on materials included in the toolkit related to evaluating instruments. The NIC leader conducted two types of pretesting to test the drafted instruments: reviewer feedback and cognitive interviewing. The draft student and teacher surveys were sent to early childhood professional staff at their districts and to teachers with targeted questions such as those in [Tool 4A. Template: Reviewer feedback instructions](#) (such as, “Are there important considerations about teachers’ ability to plan for play-based learning activities that are not included in this survey?”). In addition, the instructional specialist participated in the learning activity described in [Tool 4B. Activity: Learn about cognitive interviews](#). She then created a cognitive interview protocol, borrowing language from the exercise in the tool, and conducted cognitive interviews with four children (after getting required parental consent). The team reviewed the cognitive interview notes and discussed implications and revisions (exhibit B4).

## Exhibit B4. Early Childhood Education Networked Improvement Community example from Tool 4B

### Cognitive interview plans

Interviewer script: Thank you for helping me out today. We are planning to ask questions to students like you to find out what they think about school. I want to try out these sentences to make sure they are easy to understand. I'll be taking notes, and I'll ask you to stop sometimes so that we can talk about what you were thinking as you answer these questions. I will read each sentence to you, and then I want you to give me your answer about how often this is true for you by pointing to one of the faces. This smile face means "most of the time" (*interviewer points to the smile face and the words "most of the time"*); this face means sometimes (*interviewer points to the neutral face and the word "sometimes"*); and this face means never (*interviewer points to the sad face and the word "never"*). This is not a test for you, and there are no right or wrong answers. I am just interested in what you are thinking as you answer these so I can make the sentences better if I need to.

Are you ready?

1. I like to be at school.

*(most of the time, sometimes, never)*

2. I like to play at school.

*(most of the time, sometimes, never)*

Planned probe: Can you tell me what you were thinking about when you were answering this one?

3. I have fun when I learn at school.

*(most of the time, sometimes, never)*

Planned probe: Can you tell me why you gave this answer?

4. I have fun at school.

*(most of the time, sometimes, never)*

5. I learn a lot at school.

*(most of the time, sometimes, never)*

6. My teacher makes learning fun.

*(most of the time, sometimes, never)*

Planned probe: Can you give me an example of how your teacher does this? (Ask only if child answers most of the time or sometimes.)

7. I feel bored at school.

*(most of the time, sometimes, never)*

Planned probe: Can you tell me what "bored" means?

### Summary of cognitive interview notes and implications

- All four children's answers to the planned probes and unplanned probes indicated that they understood the sentences for items 1-7.

Implication: No items need to be reworded for clarity.

- Two of the children's answers included "yes" rather than "most of the time" when giving a positive answer and included "no" rather than "never" for a negative answer. One child asked about the difference between sometimes and most of the time.

Implication: Let's consider using yes/sometimes/no to simplify the response options.

- The use of face icons worked well for items 1-6 but not for item 7. For item 7, one child indicated that they are never bored at school and asked if they should point to the smile face for this one. One child who had been pointing to each answer hesitated on the item and said they were never bored at school but did not point to the answer. One child said "sometimes" and pointed to the neutral face. One child said, "I get bored when math is too easy" and pointed to the sometimes answer.

Implication: The face icons worked well in general, but because item 7 was worded negatively, the meaning of the answers did not match the corresponding faces. Let's drop item 7.

## Final instruments

### Early Childhood Education Networked Improvement Community teacher survey

To what extent do you agree with the statements? (*strongly disagree, disagree, agree, strongly agree*)

1. I feel confident that I understand the five different types of play.  
(*strongly disagree, disagree, agree, strongly agree*)
2. I feel like I don't have enough time to both do play and meet academic outcomes.  
(*strongly disagree, disagree, agree, strongly agree*)
3. I feel confident that I can design valuable play experiences for students.  
(*strongly disagree, disagree, agree, strongly agree*)
4. I can connect play experiences to academic standards and learning objectives.  
(*strongly disagree, disagree, agree, strongly agree*)
5. I feel confident facilitating students in their play.  
(*strongly disagree, disagree, agree, strongly agree*)
6. Identify at least one success in implementing play this week.  
\_\_\_\_\_
7. Identify at least one challenge in implementing play this week.  
\_\_\_\_\_
8. What other support do you need or questions do you have?  
\_\_\_\_\_

### Teacher play audit

1. How many minutes are student spending in free play each week on average? \_\_\_\_\_
2. How many minutes are student spending in inquiry play each week on average? \_\_\_\_\_
3. How many minutes are student spending in collaborative play each week on average?  
\_\_\_\_\_
4. How many minutes are student spending in playful learning each week on average? \_\_\_\_\_
5. How many minutes are student spending in learning through games each week on average? \_\_\_\_\_
6. Would you like to share anything else related to play that happened in your classroom this week? \_\_\_\_\_

## Appendix B. Examples from Oklahoma Excel networked improvement communities

### Teacher observation form (completed by instructional specialist)

Date: \_\_\_\_\_

Time start: \_\_\_\_\_

Time end: \_\_\_\_\_

For each type of play, note if it was observed.

Type of play	Observed (yes/no)
Free play	
Inquiry play	
Collaborative play	
Playful learning	
Learning through games	

Notes for follow-up discussion about successes, challenges, and needed support

---

### Student survey

1. I like to be at school.  
(no, sometimes, yes)
2. I like to play at school.  
(no, sometimes, yes)
3. I have fun when I learn at school.  
(no, sometimes, yes)
4. I have fun at school.  
(no, sometimes, yes)
5. I learn a lot at school.  
(no, sometimes, yes)
6. My teacher makes learning fun.  
(no, sometimes, yes)

*Step 5. Plan data collection routines*

The ECE NIC team used [Tool 5. Template: Measurement plan](#) to record plans for the timing and administration of the instruments.

**Exhibit B5. Early Childhood Education networked improvement community example from Tool 5**

Instrument/ data source	Related research question(s)	Who implements the measure	Action steps to prepare for measurement	When measurement occurs	Resources needed
Teacher survey	1	Teachers	<ul style="list-style-type: none"> <li>• Create and test online survey.</li> <li>• Test data file.</li> <li>• Make sure teachers are linked to districts.</li> </ul>	2x: beginning and end of the cycle	Online survey form
Teacher self-report play audit	2, 3	Teachers	<ul style="list-style-type: none"> <li>• Create and test online survey.</li> <li>• Test data file.</li> <li>• Make sure teachers are linked to districts.</li> </ul>	3x: beginning, middle, and end of the cycle	Online survey form
Observation checklist	3	Instructional specialist uses the checklist during teacher observations and follow-up coaching conversations	<ul style="list-style-type: none"> <li>• Schedule observations and follow-up coaching support for roughly the same time for each teacher.</li> </ul>	1x: middle of the cycle	Paper form
Student survey	4	The teachers read the surveys to the students	<ul style="list-style-type: none"> <li>• Teachers:</li> <li>• Print out paper forms and practice reading the surveys to the students.</li> <li>• Explore if consent is needed.</li> </ul>	2x: beginning and end of the cycle	Paper form

*Step 6. Plan for data discussions*

Using [Tool 6A. Worksheet: Plan for data displays](#), the ECE NIC team planned for the creation of data displays. Using [Tool 6B. Worksheet: Plan for data inquiry](#), the team created *orient*, *explore*, and *interpret* questions to guide the participants in their data discussions. The participants discussed the *orient* questions before looking at the data displays.

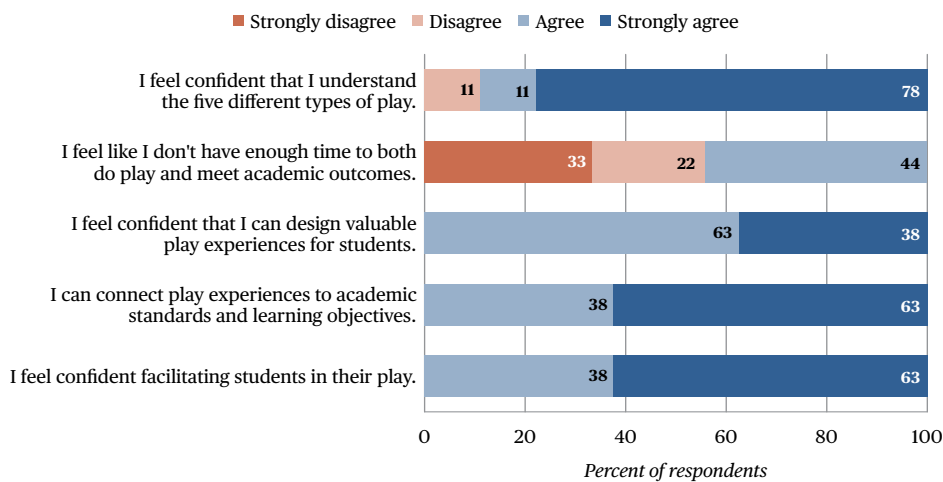
Below are some examples of the data displays and corresponding questions and responses for research questions 1, 3, and 4. (The team organized and reflected on answers to open-ended items from the teacher survey and observation checklist notes to explore research question 2 about successes and challenges; it did not create data displays.)

**Research question 1**

For the first example, the instructional specialist prepared the data display for the first teacher survey results after the first professional development session near the beginning of the cycle. She shared this with the three district team leaders to support their preparation for the remainder of the professional development work they were planning with the teachers on implementing effective play strategies. This data display includes survey results from all eight teachers from just the beginning of the cycle (exhibit B6).

**Exhibit B6. Early Childhood Education Networked Improvement Community example from Tool 6B**

*How confident are teachers in their understanding and ability to implement effective play-based learning strategies?*



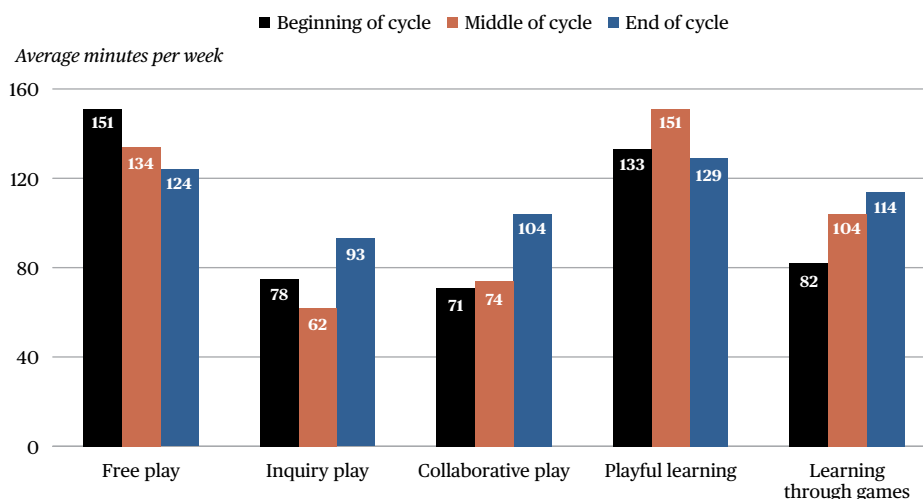
Steps	Discussion questions	Team responses
Orient	<ul style="list-style-type: none"> <li>At the start of the cycle, how confident do you think teachers are facilitating play and understanding the five types of play?</li> <li>What do you predict the data will tell us about teachers having enough time to both do play and meet academic standards?</li> </ul>	<ul style="list-style-type: none"> <li>Teachers are probably confident in facilitating students in their play.</li> <li>They may be least confident in connecting play to standards.</li> <li>Since we just started the professional development, they probably have less confidence in facilitating and understanding the five types of play.</li> <li>There may be some concern about not having time to do quality play and meet standards.</li> </ul>
Explore	<ul style="list-style-type: none"> <li>What are the most important findings?</li> <li>Were your predictions accurate?</li> </ul>	<ul style="list-style-type: none"> <li>Almost half agreed they did not have enough time to both do play and meet academic outcomes.</li> <li>Teachers felt more confident connecting play experiences to academic standards than we predicted.</li> </ul>
Interpret	<ul style="list-style-type: none"> <li>What are the implications for next steps in providing the professional development for our teachers?</li> </ul>	<ul style="list-style-type: none"> <li>Let's be sure to emphasize how the play-based learning can help meet academic outcomes, so there can be time for both simultaneously. Let's explore their feelings about this more.</li> </ul>

**Research question 3**

For the next example, the instructional specialist and three district team leaders prepared the data displays to share with all the teachers at the end of the cycle (exhibit B7).

**Exhibit B7. Early Childhood Education Networked Improvement Community example from Tool 6B**

*How much time are teachers spending on different kinds of play activities?*



Steps	Discussion questions	Team responses
Orient	<ul style="list-style-type: none"> <li>What will we expect to see if our change idea was implemented as planned?</li> </ul>	<ul style="list-style-type: none"> <li>The use of free play will decrease and the use of other types of play, particularly inquiry play, will increase.</li> </ul>
Explore	<ul style="list-style-type: none"> <li>What are the most important findings?</li> <li>Were your predictions accurate?</li> </ul>	<ul style="list-style-type: none"> <li>Free play declined by about half an hour per week from the beginning of the cycle to the end.</li> <li>Interesting that inquiry play and collaborative play increased in the second half of the cycle after we spent the most time learning and preparing for these strategies.</li> <li>Free play and playful learning remained relatively high.</li> </ul>
Interpret	<ul style="list-style-type: none"> <li>What are the implications for next steps in improving implementation of play-based learning strategies?</li> <li>Do we have evidence that we implemented the change idea as planned?</li> </ul>	<ul style="list-style-type: none"> <li>The postobservation conferences helped focus on inquiry play and collaborative play. We should continue to emphasize these two types of play.</li> <li>It looks as though we reported decreasing our free play and increasing other types of play, although might this be due partly to us being more informed about classifying our play activities into other types over time?</li> </ul>

One of the issues brought up by the NIC team during discussion of the implications is that the decline in the use of free play reported by teachers might be due in part to the teachers' increased understanding of how to classify other types of play that they might previously have identified as free play. This is an astute consideration relating to the potential accuracy of the data, given the context. If teachers understand terms and concepts

## Appendix B. Examples from Oklahoma Excel networked improvement communities

differently over time, the change in data related to those terms and concepts might be confounded and hard to interpret. Does the difference in data over time reflect an actual change in activities, or might this reflect a change in how the terms are understood? When issues like this come up, it is a good idea to incorporate that consideration into the conversation about how to interpret the data.

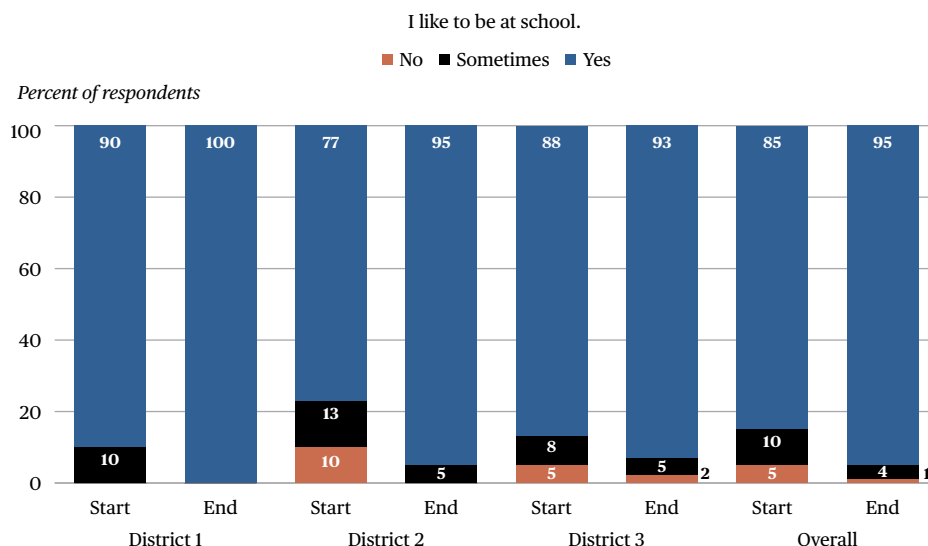
### **Research question 4**

For the next example, the instructional specialist and three district team leaders prepared the data displays about student perceptions of school to share with all the teachers at the end of the cycle. They looked at each survey item individually and at a display of all items together since all the items were measuring related concepts. This example is of the first item, “I like to be at school.” The district teams looked at their own data and data across all students (exhibit B8).



**Exhibit B8. Early Childhood Education Networked Improvement Community example from Tool 6B**

*Do students have positive attitudes about their learning experiences at school?*



Steps	Discussion questions	Team responses
Orient	<ul style="list-style-type: none"> <li>What are we expecting to see about students' perceptions of school?</li> </ul>	<ul style="list-style-type: none"> <li>Students will probably be more positive about school over time.</li> </ul>
Explore	<ul style="list-style-type: none"> <li>Were your predictions accurate?</li> <li>What stands out?</li> <li>What do you make of district differences?</li> </ul>	<ul style="list-style-type: none"> <li>More students said "yes" they like to be at school at the end of the cycle (10 percentage points more).</li> <li>Different demographics might play a role in district differences. Hard to tell.</li> </ul>
Interpret	<ul style="list-style-type: none"> <li>What are the implications for next steps in improving implementation of play-based learning strategies?</li> <li>Do we have evidence that we got an expected outcome?</li> <li>What implications does this have for measurement in the future related to student perceptions?</li> </ul>	<ul style="list-style-type: none"> <li>The students like being at school, but they were pretty positive at the start of our cycle too.</li> <li>The students showed more positivity, but it isn't clear if it is because we are getting better at providing effective play experiences. Could be because we had less virtual school (for some of us).</li> <li>Would be good to have a way to measure student engagement and positivity that would show more growth, such as a student engagement tracker or a daily student report-out activity.</li> <li>We should also look at amount of play in different districts to see if that explains district difference in student positivity.</li> <li>Next time look at student academic growth together with play experiences.</li> </ul>

The data interpretation conversations about student attitudes led to additional analyses. One of the reflections asked if the increase in positive attitudes might reflect a potentially confounding factor because some of the students were in a school that moved from some virtual instruction to all in-person instruction during this time and that alone could have accounted for the overall average increase in positive attitudes. To investigate this possibility, the analyses were done separately for students in that school and for students who had had all in-person instruction throughout the time period. Both groups had an increase

## Appendix B. Examples from Oklahoma Excel networked improvement communities

in positivity, so the team was able to determine that the average increase was not being driven solely by the students in the school that changed from some virtual instruction to all in-person instruction. Additionally, the team looked at the average student positivity increases for each district and compared those values with the total amount of play the teachers in that district reported. No clear pattern was detected, but this spurred conversation about plans to measure this potential association in the future.

The ECE NIC members talked a lot about their experiences with providing play-based learning activities throughout the data discussions, along with challenges and successes. Grounding the discussions in the data made these reflections richer, and the team had lots of discussion around ways to measure their use of play, students' perceptions, and how students grow academically in playful classrooms. The use of data to jump-start the reflections of their implementation of the change ideas—getting better at using play to promote learning—was a valuable addition to their processes and commitment for continuous improvement.

## References

- Pyle, A., & DeLuca, C. (2017). Assessment in play-based kindergarten classrooms: An empirical study of teacher perspectives and practices. *The Journal of Educational Research, 110*(5), 457-466.
- Fisher, D., Frey, N., and Lapp, D. (2016). *Text complexity: stretching readers with texts and tasks*. Corwin Literacy
- Tullock-Rhody, R., & Alexander, J. E. (1980). A scale for assessing attitudes toward reading in secondary schools. *Journal of Reading, 23*(7), 609-614.

# APPENDIX C. OVERVIEW OF THE SOUTHWEST NETWORKED IMPROVEMENT COMMUNITIES RESEARCH PARTNERSHIP



## Southwest Networked Improvement Communities Research Partnership (SWNIC)



### Partnership goals

Regional Educational Laboratory (REL) Southwest and the Oklahoma State Department of Education (OSDE) are partnering to support the use of **networked improvement communities (NICs)** and **improvement science** as part of the [Oklahoma Champions of Excellence Program](#). This new state program seeks to identify, promote, and celebrate innovative and effective programs in districts across Oklahoma. Through content-area NICs, REL Southwest researchers, OSDE staff, and participating districts will identify a common problem and then, by applying the principles of improvement science, will conduct a series of rapid minitrials to test and scale promising solutions.

The SWNIC partnership has the following goals:

- ▶ Build state capacity for implementing content-area NICs to test and scale innovative and effective school programs that support a well-rounded education, safe and healthy schools, and the effective use of technology.
- ▶ Increase state and district understanding of improvement science and the use of data for improvement in education settings.

### Projects

In partnership with REL Southwest, members engage in a learning cycle that includes training, coaching, technical support, applied research, and engagement activities to support the goals of the SWNIC partnership. This work reinforces member capacity to use research in solving high-leverage education challenges. To learn more about our current projects, please visit <https://ies.ed.gov/ncee/edlabs/regions/southwest/partnerships/swnic.aspx>.

**Coaching to support the use of NICs and improvement science.** Through the Oklahoma Champions of Excellence Program, selected Oklahoma public school districts will use OSDE-developed rubrics to rate promising programs across several content areas. To test and scale these programs, OSDE staff will form content-area NICs to apply the concepts of improvement science. REL Southwest is supporting these efforts by providing OSDE staff with in-depth coaching and consultation on establishing, managing, and sustaining NICs and on applying improvement science to educational settings. REL Southwest's work with OSDE has helped to strengthen their support for four Oklahoma districts as they pilot NICs focused on improvement ideas related to elementary science instruction and safe and healthy learning environments.

### Engaging our region

Engaging stakeholders is crucial to the REL mission of translating research into practice. Sharing research, learnings from coaching and training sessions, and insights from partnership members with stakeholders in the region supports Oklahoma's goal to promote and celebrate innovative and effective school programs. SWNIC engagement efforts include an [ongoing blog series](#) highlighting the partnership's work, a [roundup of REL resources](#) on NICs, and a research-to-practice [bridge event](#) in Oklahoma City on the fundamentals of NICs and improvement science.

Visit the [REL Southwest website](#) and follow us on [Twitter](#) to learn more about our work. You can also sign up for the [REL Southwest Spotlight newsletter](#) for regular updates and browse our upcoming [events](#).

Regional Educational Laboratory (REL) Southwest is part of a network of 10 RELs funded by the U.S. Department of Education's Institute of Education Sciences. Each REL serves a designated region of the country and helps states and districts use data and research to address policy and practice issues with the goal of improving student outcomes.

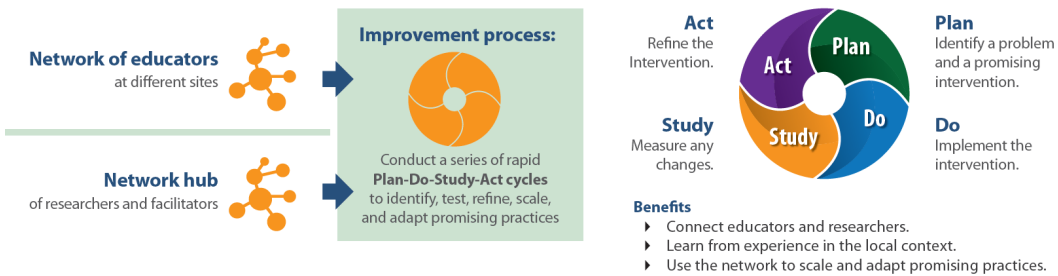
This work was funded by the U.S. Department of Education's Institute of Education Sciences (IES) under contract 91990018C0002, administered by American Institutes for Research. The content of the infographic does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

Southwest Networked Improvement Communities Research Partnership (SWNIC)

## Building networks for continuous improvement of the Champions of Excellence Program

Educators develop and use many promising practices that often are never shared or tested in other contexts.<sup>1</sup> Although research offers a valuable means to test and inform improvement efforts, formal studies may not always keep pace with real-world demands for evidence to inform practice.<sup>2</sup>

**Networked improvement communities** and **improvement science** accelerate the research process to spur innovation and identify **what works, when, for whom, and in what contexts**.<sup>3</sup>



## Oklahoma Champions of Excellence Program: Identify Excellence, Celebrate Dedication, Share Wisdom

REL Southwest and OSDE are implementing NICs as part of the state's **Champions of Excellence Program**. This program is using Title IV, Part A funds to identify, celebrate, and share innovative and effective programs that promote a **well-rounded education** and **safe and healthy schools**.<sup>4</sup>

**NIC structure and participants:** Content-area NICs, including:

- ▶ OSDE Curriculum and Instruction staff
- ▶ REL Southwest staff
- ▶ Public school districts

**Process:**

**1** Use content-area rubric to self-rate chosen school program [Bronze, Silver, Gold levels].

**2** Identify desired improvement and intervention.

**3** Work with NIC to conduct PDSA cycles.

**4** Celebrate and share results.

**NIC and rubric content areas:**

▶ Computer science	▶ Fine arts	▶ Science	▶ World languages
▶ English language arts	▶ Mathematics	▶ Social studies	▶ Safe and healthy schools

**References**

- <sup>1</sup> Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- <sup>2</sup> Carnegie Foundation for the Advancement of Teaching. (n.d.). Homepage. Retrieved March 18, 2019, from <https://www.carnegiefoundation.org>.
- <sup>3</sup> Oklahoma State Department of Education. (2018). *Title IV, Part A—Oklahoma Champions of Excellence Program*. Retrieved March 18, 2019, from <https://sde.ok.gov/TitleIVA>.
- <sup>4</sup> Proger, A. R., Bhatt, M. P., Cirks, V., & Gurke, D. (2017). *Establishing and sustaining networked improvement communities: Lessons from Michigan and Minnesota* (REL 2017–264). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved March 18, 2019, from <https://ies.ed.gov/ncee/edlabs/projects/project.asp?projectId=4556>.

4700 Mueller Boulevard  
Austin, TX 78723  
800.476.6861 | Twitter: @RELSouthwest | [relsouthwest@air.org](mailto:relsouthwest@air.org)  
<https://ies.ed.gov/ncee/edlabs/regions/southwest/default.aspx>

7725 03/19

### Acknowledgments

This toolkit was created through collaboration and contributions from our partners at the Oklahoma State Department of Education, Oklahoma Excel Team: Dawn Irons, Executive Director of Champions of Excellence and Networked Improvement Communities; Gena Barnhill, Mathematics Instructional Specialist; Andrea Chrisman, ELA Instructional Specialist; Magdalena Grifaldo, Early Childhood Instructional Specialist; and Susan Wray, Science Instructional Specialist. Our subject matter expert reviewer, Dr. Stephen Plank, Vice President, Research, Evaluation, Evidence and Data at The Annie E. Casey Foundation, also supported this work.

REL 2023-139

October 2022

This report was prepared for the Institute of Education Sciences (IES) under Contract 91990018C0002 by the Regional Educational Laboratory Southwest administered by American Institutes for Research. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

This REL report is in the public domain. While permission to reprint this publication is not necessary, it should be cited as:

Walston, J., & Conley, M. (2022). *Practical measurement for continuous improvement in the classroom: A toolkit for educators* (REL 2023-139). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. <https://ies.ed.gov/ncee/rel/Products/Publication/100841>.

This report is available on the Regional Educational Laboratory website at <https://ies.ed.gov/ncee/rel/>.