Improvement Science Cheat Sheet

Focus Collective Efforts

Understand the Problem and the System that Produces It

Generate Ideas for Change

WHAT NEXT?

Spread and Scale

Test and Build Evidence
**Improvement Science Cheat Sheet**

- Understand the Problem and the System that Produces It
- Focus Collective Efforts
- Generate Ideas for Change
- Spread and Scale
- Test and Build Evidence

*WHAT NEXT?*
Improvement Science Cheat Sheet

Focus Collective Efforts

Understand the Problem and the System that Produces It

Generate Ideas for Change

WHAT NEXT?

Spread and Scale

Test and Build Evidence
1. The Purpose of Measurement: Explore how measurement for improvement is different than measurement used for accountability or research.

2. System of Measures: Motivate the need for a “system of measures” for improvement that includes outcome, driver, process, and balancing measures.

3. Case: Show how a measurement system supports improvement efforts.
The Three Faces of Performance Measurement: Improvement, Accountability, and Research

David Yeager
University of Texas at Austin

Anthony Bryk
Jane Muhich
Hannah Hausman
Lawrence Morales

Practical Measurement

Connections Between Quality Measurement and Improvement

Donald M. Berwick, MD,* Brent James, MD, MStat,† and Molly Joel Coyer, MD‡

BACKGROUND. Measurement is necessary but not sufficient for quality improvement. Because the purpose of the national quality measurement and reporting system (NQMS) is to improve quality, a discussion of the link between measurement and improvement is critical for ensuring an appropriate system of metrics.

OBJECTIVES. To classify metrics and discuss the shortcomings of existing systems and potential solutions.

RESULTS. Neither the dynamics of selecting the dynamics of improvement work or the barriers to embedding the metrics. The metrics are not just in the tool kit, but also in the tool box and individuals actors.

CONCLUSIONS. Neither the dynamics of selecting the dynamics of improvement work or the barriers to embedding the metrics are included in the tool box and individuals actors.
“We are increasingly realizing how critical measurement is for the improvement we seek, yet how counterproductive it can sometimes be to mix measurement for accountability with measurement for improvement. Considered one by one, measurement for each purpose can be good and very important. If done poorly, it can be bad. If the measurements are mixed together in inappropriate ways, they can indeed be harmful or destructive, with the mixed purposes interfering with one another.”

Leif I. Solberg, MD
To provide usable information for improvement, we need to consider:

**What is measured:**
Needs to be closely aligned to the actual work and specific to the processes and outcomes you hope to change.

**How & when it is measured:**
Needs to be embedded in the daily workflow.
Must produce data accessible in a timely manner.

**Social processes shaping use:**
Requires transparency, trust, low stakes, and the safety to take risks.

**Language** needs to be meaningful to those doing the work.
<table>
<thead>
<tr>
<th>ASPECT</th>
<th>IMPROVEMENT</th>
<th>ACCOUNTABILITY</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why?</td>
<td>Develop and evaluate</td>
<td>Identify exemplary or</td>
<td>Develop and test theories about the</td>
</tr>
<tr>
<td></td>
<td>changes in practice</td>
<td>problematic performers</td>
<td>relationships between conceptual</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(teachers, schools,</td>
<td>variables</td>
</tr>
<tr>
<td></td>
<td></td>
<td>districts)</td>
<td></td>
</tr>
<tr>
<td>What?</td>
<td>Outcomes and processes</td>
<td>End of the line outcomes</td>
<td>Latent variables</td>
</tr>
<tr>
<td></td>
<td>that are the object of</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>change</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lief Solberg, Gordon Mosser and Sharon McDonald *Journal on Quality Improvement* vol. 23, no. 3, (March 1997), 135-147.
<table>
<thead>
<tr>
<th>ASPECT</th>
<th>IMPROVEMENT</th>
<th>ACCOUNTABILITY</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why?</td>
<td>Develop and evaluate changes in practice</td>
<td>Identify exemplary or problematic performers (teachers, schools, districts)</td>
<td>Develop and test theories about the relationships between conceptual variables</td>
</tr>
<tr>
<td>What?</td>
<td>Outcomes and processes that are the object of change</td>
<td>End of the line outcomes</td>
<td>Latent variables</td>
</tr>
<tr>
<td>How often?</td>
<td>Frequently as practice occurs</td>
<td>Usually collected once a year (after the fact)</td>
<td>Typically once or twice per study (after the fact)</td>
</tr>
<tr>
<td>Testing your theory</td>
<td>Sequential tests</td>
<td>No theory to test</td>
<td>One large test</td>
</tr>
<tr>
<td>Sample size</td>
<td>“Just enough” data, small sequential samples</td>
<td>Obtain 100% of available, relevant data</td>
<td>“Just in case” data</td>
</tr>
</tbody>
</table>

Lief Solberg, Gordon Mosser and Sharon McDonald *Journal on Quality Improvement* vol. 23, no. 3, (March 1997), 135-147.
<table>
<thead>
<tr>
<th>ASPECT</th>
<th>IMPROVEMENT</th>
<th>ACCOUNTABILITY</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Why?</strong></td>
<td>Develop and evaluate changes in practice</td>
<td>Identify exemplary or problematic performers (teachers, schools, districts)</td>
<td>Develop and test theories about the relationships between conceptual variables</td>
</tr>
<tr>
<td><strong>What?</strong></td>
<td>Outcomes and processes that are the object of change</td>
<td>End of the line outcomes</td>
<td>Latent variables</td>
</tr>
<tr>
<td><strong>How often?</strong></td>
<td>Frequently as practice occurs</td>
<td>Usually collected once a year (after the fact)</td>
<td>Typically once or twice per study (after the fact)</td>
</tr>
<tr>
<td><strong>Testing your theory</strong></td>
<td>Sequential tests</td>
<td>No theory to test</td>
<td>One large test</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>“Just enough” data, small sequential samples</td>
<td>Obtain 100% of available, relevant data</td>
<td>“Just in case” data</td>
</tr>
<tr>
<td><strong>Social Conditions of Use?</strong></td>
<td>Data shared in a low-stakes, safe environment conducive to change.</td>
<td>Publically available. Formal collection process to assure appearances of neutrality and objectivity.</td>
<td>Meets scientific standards that are held in the field.</td>
</tr>
</tbody>
</table>

Lief Solberg, Gordon Mosser and Sharon McDonald *Journal on Quality Improvement* vol. 23, no. 3, (March 1997), 135-147.
<table>
<thead>
<tr>
<th>ASPECT</th>
<th>IMPROVEMENT</th>
<th>ACCOUNTABILITY</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why?</td>
<td>Develop and evaluate changes in practice</td>
<td>Identify exemplary or problematic performers (teachers, schools, districts)</td>
<td>Develop and test theories about the relationships between conceptual variables</td>
</tr>
<tr>
<td>What?</td>
<td>Outcomes and processes that are the object of change</td>
<td>End of the line outcomes</td>
<td>Latent variables</td>
</tr>
<tr>
<td>How often?</td>
<td>Frequently as practice occurs</td>
<td>Usually collected once a year (after the fact)</td>
<td>Typically once or twice per study (after the fact)</td>
</tr>
<tr>
<td>Testing your theory</td>
<td>Sequential tests No theory to test</td>
<td>No theory to test</td>
<td>One large test</td>
</tr>
<tr>
<td>Sample size</td>
<td>&quot;Just enough&quot; data, small sequential samples</td>
<td>Obtain 100% of available, relevant data</td>
<td>“Just in case” data</td>
</tr>
<tr>
<td>Social Conditions of Use?</td>
<td>Data shared in a low-stakes, safe environment conducive to change.</td>
<td>Publically available. Formal collection process to assure appearances of neutrality and objectivity.</td>
<td>Meets scientific standards that are held in the field.</td>
</tr>
</tbody>
</table>

Key limitation for improvement: Does not illuminate WHY the outcomes occur or what should be done to change them.

Lief Solberg, Gordon Mosser and Sharon McDonald *Journal on Quality Improvement* vol. 23, no. 3, (March 1997), 135-147.
<table>
<thead>
<tr>
<th>ASPECT</th>
<th>IMPROVEMENT</th>
<th>ACCOUNTABILITY</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Why?</strong></td>
<td>Develop and evaluate changes in practice</td>
<td>Identify exemplary or problematic performers (teachers, schools, districts)</td>
<td>Develop and test theories about the relationships between conceptual variables</td>
</tr>
<tr>
<td><strong>What?</strong></td>
<td>Outcomes and processes that are the object of change</td>
<td>Latent variables</td>
<td>Latent variables</td>
</tr>
<tr>
<td><strong>How often?</strong></td>
<td>Frequently as practice occurs, typically once a year (after the fact)</td>
<td>Typically once or twice per study (after the fact)</td>
<td>One large test</td>
</tr>
<tr>
<td><strong>Testing your theory</strong></td>
<td>Sequential tests</td>
<td>One large test</td>
<td>“Just in case” data</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>“Just enough” data</td>
<td>One large test</td>
<td>“Just in case” data</td>
</tr>
<tr>
<td><strong>Social Conditions of Use?</strong></td>
<td>Data shared in a low-stakes, safe environment conducive to change.</td>
<td>Publicly available. Formal collection process to assure appearances of neutrality and objectivity.</td>
<td>Meets scientific standards that are held in the field.</td>
</tr>
</tbody>
</table>

**Key limitation for improvement:**
**IMPRactical to administer; not designed to inform changes in practice**

<table>
<thead>
<tr>
<th>ASPECT</th>
<th>IMPROVEMENT</th>
<th>ACCOUNTABILITY</th>
<th>RESEARCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Why?</td>
<td>Develop and evaluate changes in practice</td>
<td>Identify exemplary or problematic performers (teachers, schools, districts)</td>
<td>Develop and test theories about the relationships between conceptual variables</td>
</tr>
<tr>
<td>What?</td>
<td>Outcomes and processes that are the object of change</td>
<td>End of the line outcomes Latent variables</td>
<td></td>
</tr>
<tr>
<td>How often?</td>
<td>Frequently as practice occurs</td>
<td>Usually collected once a year (after the fact)</td>
<td>Typically once or twice per study (after the fact)</td>
</tr>
<tr>
<td>Testing your theory</td>
<td>Sequential tests</td>
<td>No theory to test</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>“Just enough” data, small sequential samples</td>
<td>Obtain 100% of available, relevant data</td>
<td></td>
</tr>
<tr>
<td>Social Conditions of Use?</td>
<td>Data shared in a low-stakes, safe environment conducive to change.</td>
<td>Publically available. Formal collection process to assure appearances of neutrality and objectivity.</td>
<td>Meets scientific standards that are held in the field.</td>
</tr>
</tbody>
</table>

**Measurement for improvement supports the ongoing refinement of knowing what works for whom and under what conditions**

Lief Solberg, Gordon Mosser and Sharon McDonald *Journal on Quality Improvement* vol. 23, no. 3, (March 1997), 135-147.
A System of Measures to Inform Improvement
Schools That Lead –
Shue-Medill – Milford Central Academy

SCHOOLS THAT LEAD:
Advancing Powerful Student Learning

- Dana Diesel Wallace
- Sofi Frankowski
- Michele Savage
- Nancy Carnvale
Proportion of Graduates Retained in Grade 9 (2015)

- Retained in Gr 9
- Successful Gr 9

MCA
- 17% Retained in Gr 9
- 83% Successful Gr 9

Shue-Medill
- 20% Retained in Gr 9
- 80% Successful Gr 9
System of Feedback

Measure Types

O = Outcome Measures
D = Driver Measures
P = Process Measures
B = Balancing Measures

PDSA = Learning Cycle Measure

Visual adapted from Improvement Science Consulting
System of Feedback

Yearly

O

How is the system performing? What is the result?

Measure Types

O = Outcome Measures

Visual adapted from Improvement Science Consulting
Proportion of Graduates Retained in Grade 9 (2015)

MCA
- Retained in Gr 9: 17%
- Successful Gr 9: 83%

Shue-Medill
- Retained in Gr 9: 20%
- Successful Gr 9: 80%
System of Feedback: STL-Delaware Case

Aim:

O: % of Graduates Repeating Grade 9

Measure Types

O = Outcome Measures
System of Feedback: STL-Delaware Case

Aim:

O (Lagging): % of Graduates Repeating Grade 9
O (Leading): % of Students on Watch Lists

Measure Types

O = Outcome Measures
Leading Outcome Measures: 4 Early Warning Indicators

Percent Students on Math Watch List, 15-16 and 16-17

- 2015-2016: Q1 = 10.1%, Q2 = 8.3%, Q3 = 13.0%, Q4 = 12.2%
- 2016-2017: Q1 = 10.1%, Q2 = 10.1%, Q3 = 9.7%

Percent Students on ELA Watch List, 15-16 and 16-17

- 2015-2016: Q1 = 12.1%, Q2 = 11.2%, Q3 = 13.7%, Q4 = 12.9%
- 2016-2017: Q1 = 12.1%, Q2 = 11.2%, Q3 = 13.7%, Q4 = 12.9%

Percent Students on Attendance Watch List, 15-16 and 16-17

- 2015-2016: MP1 = 3.4%, MP2 = 9.4%, MP3 = 10.7%, MP4 = 8.9%
- 2016-2017: MP1 = 3.0%, MP2 = 9.4%, MP3 = 10.3%, MP4 = 8.9%

Percent Students on Discipline Watch List, 15-16 and 16-17

- 2015-2016: MP1 = 1.4%, MP2 = 3.4%, MP3 = 4.0%, MP4 = 1.1%
- 2016-2017: MP1 = 1.1%, MP2 = 1.3%, MP3 = 1.3%, MP4 = 1.1%
System of Feedback

Measure Types

O = Outcome Measures

D = Driver Measures

Are we making progress on the intermediate outcomes?

Visual adapted from Improvement Science Consulting
System of Feedback: STL-Delaware Case

Aim:

- **O** (Lagging): % of Graduates Repeating Grade 9
- **O** (Leading): % of Students on Watch Lists

**Measure Types**

- **O** = Outcome Measures
- **D** = Driver Measures

**D1**: Instructional Practices

**D2**: Student Mindsets
Being a "math person" or not is something about you that you really can't change. Some people are good at math and other people aren't.

### Driver Measure:
**Growth Mindset Student Survey Responses in Class A, September and January**

- **Sept:** 84.6% Agree, 15.4% Disagree
- **Jan:** 65.7% Agree, 34.3% Disagree
System of Feedback

Measure Types

O = Outcome Measures
D = Driver Measures
P = Process Measures
B = Balance Measures
PDSA = Learning Cycle

Are the parts of your system performing as planned?

Visual adapted from Improvement Science Consulting
System of Feedback: STL-Delaware Case

Aim:

O (Lagging): % of Graduates Repeating Grade 9

O (Leading): % of Students on Watch Lists

D1: Instructional Practices

D2: Student Mindsets

Retesting Process → P2 → P3

Measure Types

O = Outcome Measures

D = Driver Measures

P = Process Measures
Process Measure: Percent of Targeted Students Revising Tests, Brandy Cooper, Grade 6 Math

Maximum Value: 57%
Median: 17%

Tests (Number of Targeted Students):
- T1 (16)
- T2 (11)
- T3 (23)
- T4 (8)
- T5 (10)
- T6 (19)
- T7 (13)
- T8 (21)

Request to Retest
Process Measure:
Average Original and Revised Grades for Students Revising Tests, Brandy Cooper, Grade 6 Math
System of Feedback

Measure Types

O = Outcome Measures
D = Driver Measures
P = Process Measures
B = Balance Measures

PDSA = Learning Cycle Measure

Adapted from Improvement Science Consulting
System of Feedback: STL-Delaware Case

Aim:

O (Lagging): % of Graduates Repeating Grade 9

O (Leading): % of Students on Watch Lists

D1: Instructional Practices

D2: Student Mindsets

Retesting Process → P2 → P3

Measure Types

O = Outcome Measures

D = Driver Measures

P = Process Measures

PDSA = Learning Cycle

Measure

Reasons why students don’t use form

1. Will students use a form to make revisions?

2. How can I get more students to use form?

# Forms picked up

© 2018 Carnegie Foundation for the Advancement of Teaching
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
https://creativecommons.org/licenses/by-nc-sa/4.0/
What’s happening to the parts of the system we aren’t currently focused on?

Measure Types

O = Outcome Measures
D = Driver Measures
P = Process Measures
B = Balance Measures

PDSA = Learning Cycle Measure

Visual adapted from Improvement Science Consulting
System of Feedback: STL-Delaware Case

Aim:

O (Lagging): % of Graduates Repeating Grade 9

O (Leading): % of Students on Watch Lists

Measure Types

O = Outcome Measures
D = Driver Measures
P = Process Measures
B = Balance Measures

PDSA = Learning Cycle Measure

Adapted from Schools That Lead
Outcome Measure(s)

Get Movement Here by

Driver Measures

Get movement here by

Balancing Measure(s)

Hold these constant or improve them

Process Measures

Getting movement here
AIM: By May 2017, halve the proportion (from 1 in 5 to 1 in 10) of students who get held back in 9th grade.

Leading Outcome

Lagging Outcome

Primary Drivers

Secondary Drivers

Change Ideas

Driving & Assessing

Instructional Practices & Mindsets

Grading Policies

Teacher Instructional Practices & Mindsets

Student Behavioral Supports

Others…

Driver

Process

Request to Retest Form

Etc.

Balancing

Quality of First Attempt

© 2018 Carnegie Foundation for the Advancement of Teaching
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
https://creativecommons.org/licenses/by-nc-sa/4.0/
AIM:
By May 2017, halve the proportion (from 1 in 5 to 1 in 10) of students who get held back in 9th grade.

Started tracking early 2016

Student Behavioral Supports

Systematic collection Dec 2016

Common network measure. First survey September 2015

Quality of First Attempt

In the Future?

Secondary Drivers

Change Ideas

Request to Retest Form

Etc.

Instructional Practices & Mindsets

Chart.
Summer 2015

Student Mindsets That Promote Resilience

Revising Work & Assessments

Feedback on Student Work

Grading Policies

Etc.

Revised work & Assessments

Adult Instructional Practices & Mindsets

AIM:
By May 2017, halve the proportion (from 1 in 5 to 1 in 10) of students who get held back in 9th grade.

Started tracking early 2016

Student Behavioral Supports

Systematic collection Dec 2016

Common network measure. First survey September 2015

Quality of First Attempt

In the Future?

Secondary Drivers

Change Ideas

Request to Retest Form

Etc.

Instructional Practices & Mindsets

Chart.
Summer 2015

Student Mindsets That Promote Resilience

Revising Work & Assessments

Feedback on Student Work

Grading Policies

Etc.

Revised work & Assessments

Adult Instructional Practices & Mindsets

AIM:
By May 2017, halve the proportion (from 1 in 5 to 1 in 10) of students who get held back in 9th grade.

Started tracking early 2016

Student Behavioral Supports

Systematic collection Dec 2016

Common network measure. First survey September 2015

Quality of First Attempt

In the Future?
Activity
**Scenario:** You are part of an improvement team at an elementary school. The students at your school are struggling with fractions. You launch an improvement project with your school team to tackle this problem.

**Goal:** The goal of the project is to increase students’ understanding of fractions.

**Your Task:** Develop a system of measures for this project:
- Outcome Measures: 1-2 measures
- Process or Driver Measures: 2-3 measures
- Balancing Measures: 1 measure
FRACTIONS MEASUREMENT SYSTEM

Measure Types

O = Outcome Measures
D = Driver Measures
P = Process Measures
B = Balance Measures

Students’ Understanding of Fractions

Yearly

Prof Dev for Ts

Monthly

Instructional Materials

Interactions w Math Coach

Use of Math Manipulatives

© 2018 Carnegie Foundation for the Advancement of Teaching
This work is licensed under a Creative Commons Attribution NonCommercial-ShareAlike 4.0 International License
https://creativecommons.org/licenses/by-nc-sa/4.0/

Visual adapted from Improvement Science Consulting
FRACTIONS MEASUREMENT SYSTEM

Measure Types
- O = Outcome Measures
- D = Driver Measures
- P = Process Measures
- B = Balance Measures

Students’ Understanding of Fractions

Prof Dev for Ts

Instructional Materials

Interactions w Math Coach

Use of Math Manipulatives

Visual adapted from Improvement Science Consulting

© 2018 Carnegie Foundation for the Advancement of Teaching
This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.
https://creativecommons.org/licenses/by-nc-sa/4.0/
Key Takeaways

- Data help us understand our systems and whether changes we make lead to improvement

- Measurement for improvement is different than measurement for accountability and research

- A family or system of measures, which includes outcome, driver, process, and balancing measures is needed to guide improvement efforts