Best Practices in Program Evaluation and Learning

May 28, 2015

Organized by the Federal Reserve Bank of Dallas

The ***Connecting Communities®*** audio conference series is a Federal Reserve System initiative intended to provide timely information on emerging and important community and economic development topics with a national audience. The audio conference series complements existing Federal Reserve Community Development outreach initiatives that are conducted through our regional Reserve Bank offices and at the Federal Reserve Board of Governors in Washington, D.C.

Welcome

Thank you for attending today's session. How to join (remember, this session is being recorded)

- Call-in number: 888-625-5230
- Conference code: 35331544
- Webinar link: https://www.webcaster4.com/Webcast/Page/584/8330

You have two options to ask questions today:
**These questions may be part of the recorded archive for this session.*

- Email us at: communities@stls.frb.org.
- Type your question into the chat box of the webinar.

An archived recording of this session will be available on the ***Connecting Communities®*** web site shortly after the session ends. Please visit www.stlouisfed.org/connectingcommunities/

In connection with this session, you can find a variety of additional resources on this topic , available at **www.fedcommunities.org**. We encourage you to browse through this site and to contact your regional office if you would like additional information on any of these items.

Legal Notices and Disclaimer

The information, analyses, and conclusions set forth are those of the presenters and do not necessarily indicate concurrence by the Board of Governors of the Federal Reserve System, the Federal Reserve Banks, or members of their staffs.

Community Development

The Community Development (CD) function within the Federal Reserve System – consisting of individual departments at each of the twelve Federal Reserve Banks as well as at the Board of Governors – promotes economic growth and financial stability for low- and moderate-income (LMI) communities and individuals through a range of activities, including:

- **Convening stakeholders**, including practitioners, financial institutions, nonprofits, governmental agencies, and the philanthropic and private sectors,
- **Conducting and sharing research** to examine economic challenges facing low- and moderate-income communities and attendant policy implications; and,
- **Identifying emerging issues**.

Today's Presenters

- **Daniel L. Millimet, Ph.D.**
  Professor of Economics, Southern Methodist University
- **Moira Inkelas, Ph.D. , MPH.**
  Associate Professor, Fielding School of Public Health, Department of Health Policy and Management, University of California at Los Angeles

Today's Agenda

- How does one measure the causal impact of a program?
- Tools for successful evaluation
- How an improvement approach contributes to community-wide change
- How measurement can support ongoing alignment of change efforts and actions.

Research Conference on Program Evaluation

November 16-17

Hosted by the Federal Reserve Bank of Dallas
in Partnership with Southern Methodist University

Visit the Dallas Fed's website for more information and research paper submission guidelines.

Evaluation of Program Success

Dr. Daniel L. Millimet
Professor of Economics
Southern Methodist University
millimet@smu.edu

Introduction

- Policymakers and community organizations often undertake programs in order to hopefully achieve some desired outcome
- Examples:
  - Job training program or job search assistance to lead to employment and raise earnings
  - Preventative health care to improve future health
  - Financial capacity to improve household hunger/food security
- When one's goals center on creation and implementation of such programs, *ex post* evaluation of a program's *causal effect* is vital to
  1. Judge "success"
  2. Determine the efficient usage of limited financial resources moving forward

Introduction, cont'd

- Terminology
  - Following terms are all synonyms for our purposes
    1. Program
    2. Policy
    3. Intervention
    4. Treatment
  - *Treatment group* refers to individuals exposed to or whom participate in the program in question
  - *Control group* refers to individuals not exposed or whom do not participate in the program in question
  - *Causal effect of a program* is the expected change in outcome for a given individual when participating versus not participating in the program
    - E.g., What is the difference between the actual weekly earnings of an individual after participating in a job training program and what it would have been had the individual not participated?

Introduction, cont'd

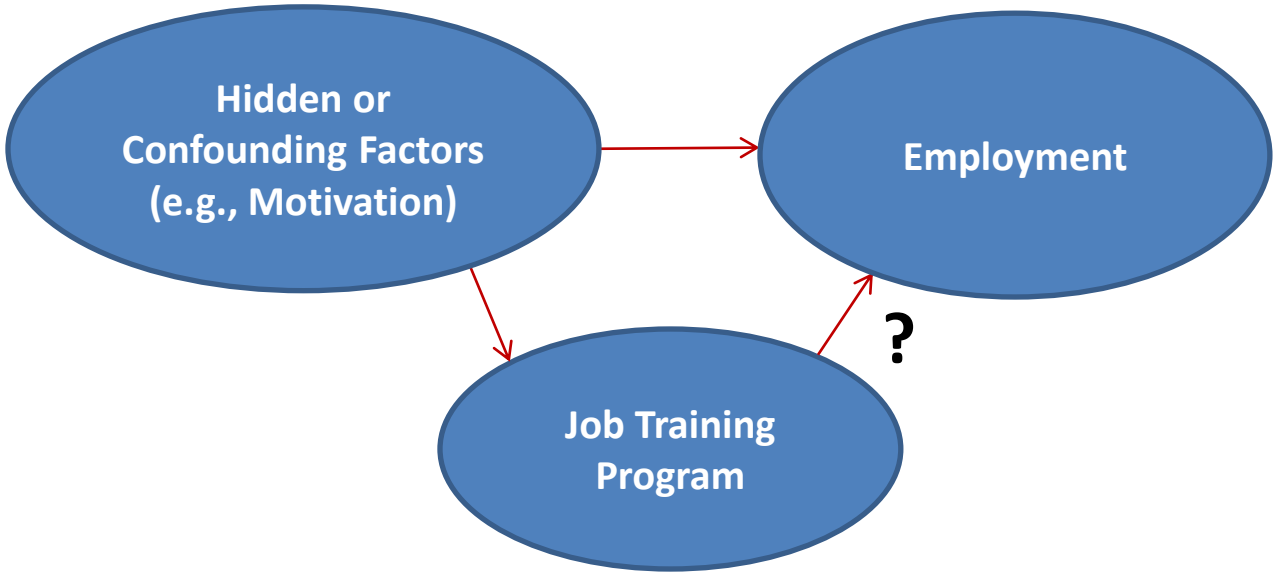
- Estimation of the *causal effect of a program* confronts several challenges
- Perhaps the most important and difficult to overcome is

self-selection

- Self-selection refers to differences between individuals participating in the program and individuals not participating that
  - Affect one's participation in the program, and
  - Are associated with the outcome of interest

Introduction, cont'd

- With self-selection, correlation between program participation and outcomes does not imply causation



- How does one isolate the causal effect?

Image source: Author's original work.

Randomized Control Trials (RCTs)

- RCTs or experiments circumvent the self-selection issue by ensuring that participating individuals (treatment group) and non-participating individuals (control group) are *identical on average*

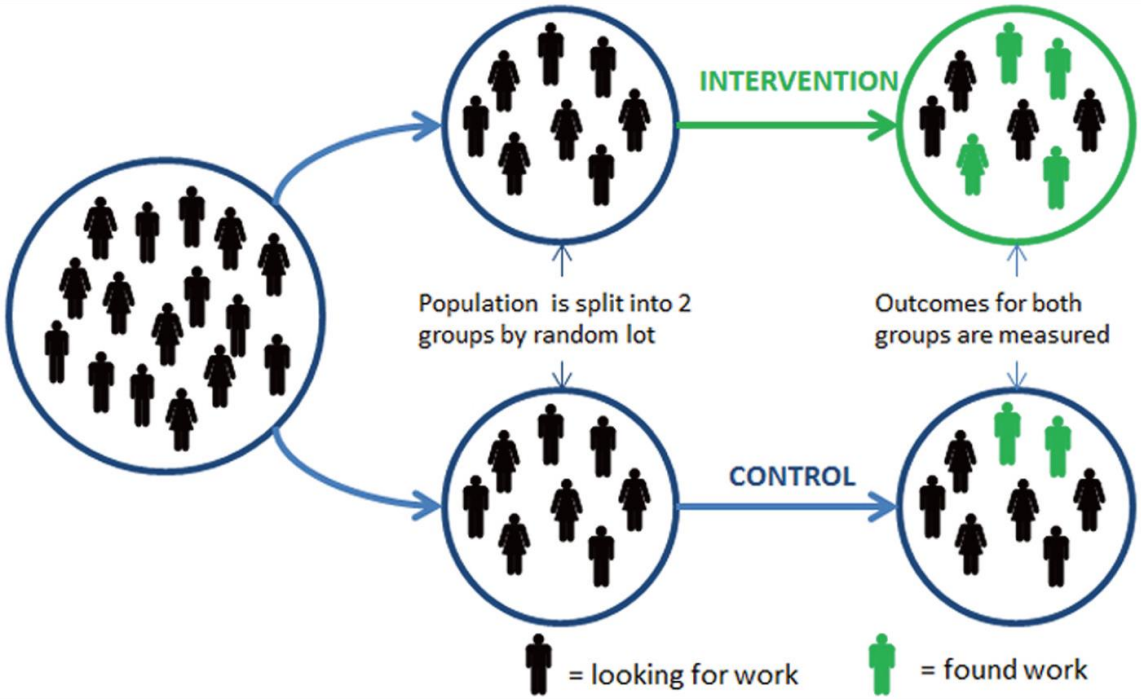


Image source: iddgblog.wordpress.com

RCTs, cont'd

- While often touted as the "gold standard" for evaluation, RCTs are not always possible
  - Resource-intensive:
    - Time, money
  - Politically infeasible:
    - Objections to randomization
- Then what?

Image source: Freshspectrum.com

Evaluation with Observational Data

- In the absence of RCTs, one must resort to evaluation methods utilizing *observational data*
- Refers to survey data collected from individuals who self-select into (and out of) the program

Image source: everythingmaths.co.za

Evaluation with Observational Data, cont'd

- With RCTs, individuals in the treatment and control group are identical on average
- With observational data, this need no longer be true
- Overcoming *self-selection* with *observational data* can be difficult, but is not without hope and is much more useful than correlations
- Requirements
  1. Data
  2. Statistical strategy

Image source: www.coursera.org/course/introstats

Evaluation with Observational Data
Data Requirements

1. Information on treated and untreated individuals
   - May require construction of a control group
2. Proper measurement of program participation and outcomes
   - Measurement error can mask causal effects
   - Self-reported vs. verified
3. Detailed individual attributes
   - Depth and breadth may be dictated by statistical strategy
4. Sample of sufficient size
   - Affects precision of estimates
5. Sample should be representative of the population of interest
   - Internal vs. external validity

Evaluation with Observational Data
Statistical Strategy

- With RCTs, there is no need to control for other attributes of individuals since randomization ensures that these are identical (on average) across the treatment and control groups
- Without randomization, one must use additional data to circumvent self-selection bias
- Strategies
  1. Cross-sectional approaches
  2. Longitudinal approaches
  3. Instrumental variables

Evaluation with Observational Data
Cross-Sectional Approaches

- Cross-sectional data approaches require data from a *single* point in time after the program
- The causal effect of the program is estimated by
  - Comparing average outcomes across the treatment and control groups, after
  - Eliminating self-selection bias by controlling for all differences between treated and non-treated individuals that are
    - Associated with program participation, and
    - Associated with the outcome of interest
- Strong data requirements
- However, with sufficient data, controlling for differences may be accomplished via
  1. Regression analysis
  2. Matching

Evaluation with Observational Data
Longitudinal Approaches

- Longitudinal approaches require data from *multiple* points in time, including both before and after program participation
- The causal effect of the program is estimated by
  - Comparing the change in average outcomes across the treatment and control groups, after
  - Eliminating self-selection bias by controlling for all time-varying differences between treated and non-treated individuals that are
    - Associated with program participation, and
    - Associated with the outcome of interest
- Strong data requirements
  - In contrast to cross-sectional approaches, multiple rounds of data collection are needed, but
  - Time invariant attributes of individuals are implicitly removed by examining changes in outcomes
- Estimation again may be accomplished via
  1. Regression analysis
  2. Matching

Evaluation with Observational Data
Instrumental Variables

- Cross-sectional approaches require all individual attributes – time-varying or time invariant – that are associated with program participation and outcomes be measured in the data
- Longitudinal approaches require all time-varying individual attributes that are associated with program participation and outcomes be measured in the data
- Often, both may be unrealistic
  1. Measurement difficulty: motivation, innate ability
  2. Data costs: time, money

Evaluation with Observational Data
Instrumental Variables, cont'd

- Instrumental variables refers to a statistical technique that is reasonably straightforward with either cross-sectional or longitudinal data that may circumvent self-selection bias
- Poses different data requirements: Requires an attribute of individuals that is
  1. Associated with program participation, but
  2. Not associated with the outcome of interest

- Intuition: Attribute (or instrument) acts as a randomizing tool similar to the randomizing agent in RCTs
- Best to understand by way of examples…

Evaluation with Observational Data Instrumental Variables, cont'd

- Example 1:

Image source: Author's original work.

Evaluation with Observational Data Instrumental Variables, cont'd

- Example 2:

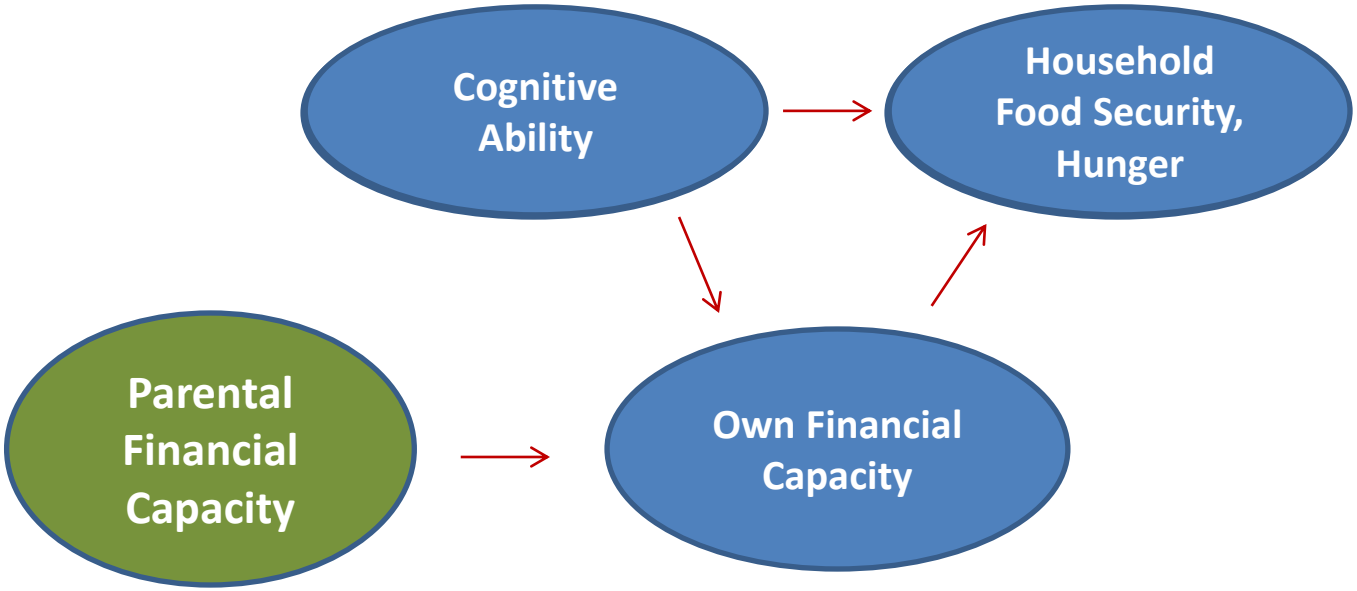


Image source: Author's original work.

Conclusion

- Successful evaluation of programs is critical for sound decision-making and efficient utilization of resources
- In the absence of RCTs, careful
  - Data collection, and
  - Statistical analysis

  can provide invaluable estimates of causal effects
- Requires
  1. A willingness to devote resources to data collection and analysis
  2. An appreciation of the significant value of *causal estimates* versus *simple correlations*

Measurement to Improve Community Well-Being

Moira Inkelas, PhD, MPH
Associate Professor
UCLA Fielding School of Public Health, Department of Health Policy and Management
minkelas@ucla.edu

Introduction

To achieve an outcome for a community population, we are seeking solutions that…

…work at **scale** *(do not break down when we try it for everyone)*

…will **spread** to others *(all organizations implement the change, not just the most "enlightened" organization)*

…are **sustained** over time *(do not degrade as attention turns to other topics)*

A Systems Approach

Definition of a System

"A system is an interdependent group of items, people, or processes working together toward a common purpose."

Associates in Process Improvement, *Quality as a Business Strategy*, 1987

Improving a System

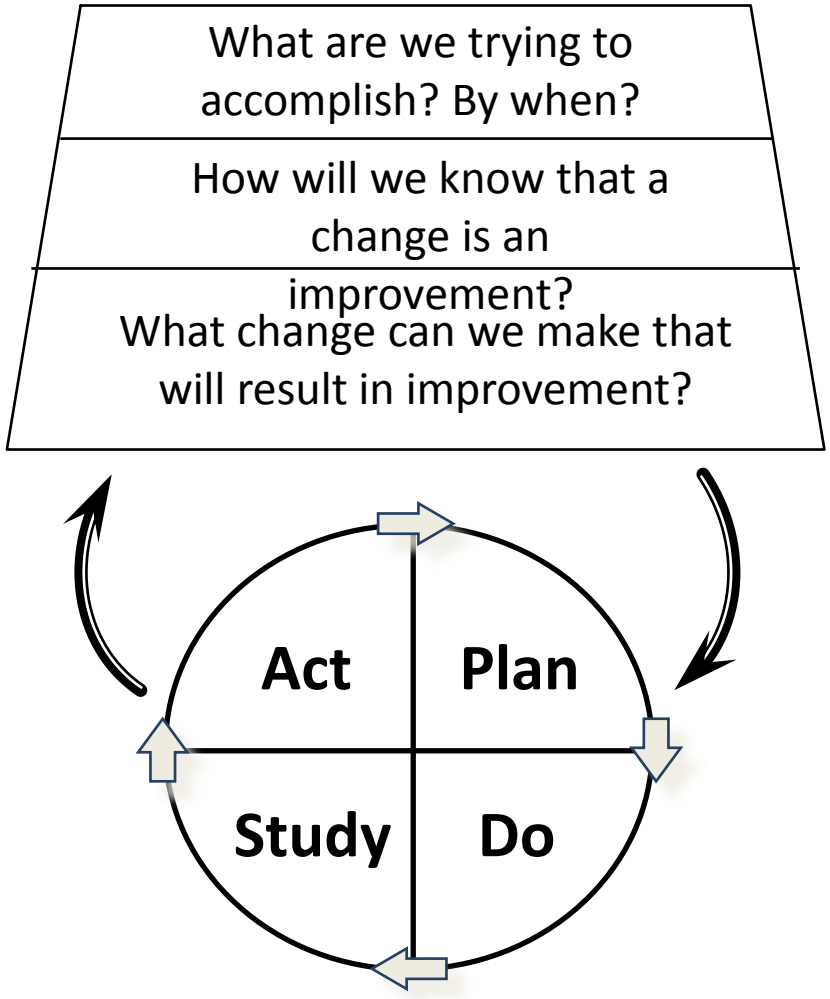
- How is improving a system different from improving a program?
  - —Programs can be planned, implemented and evaluated.
  - —It is not possible to plan and specify each of the detailed actions necessary for a system to produce better results.
  - —Optimizing one part of a system does not optimize the overall system.
  - —Community systems are complex and are never permanently “fixed.”
- To change outcomes for a population, we need an approach that sets a heading but allows for adaptation and adjustment, using testing to learn its way forward.

The Importance of Ongoing Learning

TED
IDEASWORTHSPREADING

"Designers recognize prototyping as the essence of the iterative process. With each version, we get instant feedback about what works and what doesn't work. So the capacity to prototype is essential….Design truly is a contact sport. It demands that we bring all of our senses to the task and that we apply the very best of our thinking, our feeling, and our doing to the challenge that we have."

The Model for Improvement



Aims

Measures

Changes

Source: Provost L. Model for improvement: Aims, measures, changes. Associates in Process Improvement. lprovost@apiweb.org

Change Concepts

Global Aim: Optimize early childhood development and school readiness outcomes by developing a community-wide system that increases family participation in care, improves sharing of information across sectors, promotes health and development for all children, and provides active management for families with problems or risks that is tailored to their support needs.

Goal Targets

SYSTEM CAPACITY

- 90% of families reached
- 10% annual increase in the inclusion of children's health & ECE providers
- 80% of partners have change activity >=3.0 (on a 0.5 - 5.0 scale)
- 10% annual increase in average "learning culture" scores of organizations

PROCESS

- 90% of parents receive family-centered care
- 90% of parents asked if they have child devt concerns, family stressors, depression
- 90% of parents linked to other resources with the offer of follow-up support
- Time to service initiation, for referrals
- 90% of families referred to specialized agencies complete the intake process
- 95% of families with medium or high support needs entered into a registry
- 95% of parents follow steps in care plans

OUTCOME

- % families without hardships in parenting, social, economic, and health goal areas
- % parents with depression
- 90% of children read to daily
- < 3% of children score as high risk and < 13% score as moderate risk on 36 month ASQ
- < 10% of children are developmentally vulnerable at kindergarten entry
- % proficient in third grade reading

Practice Drivers

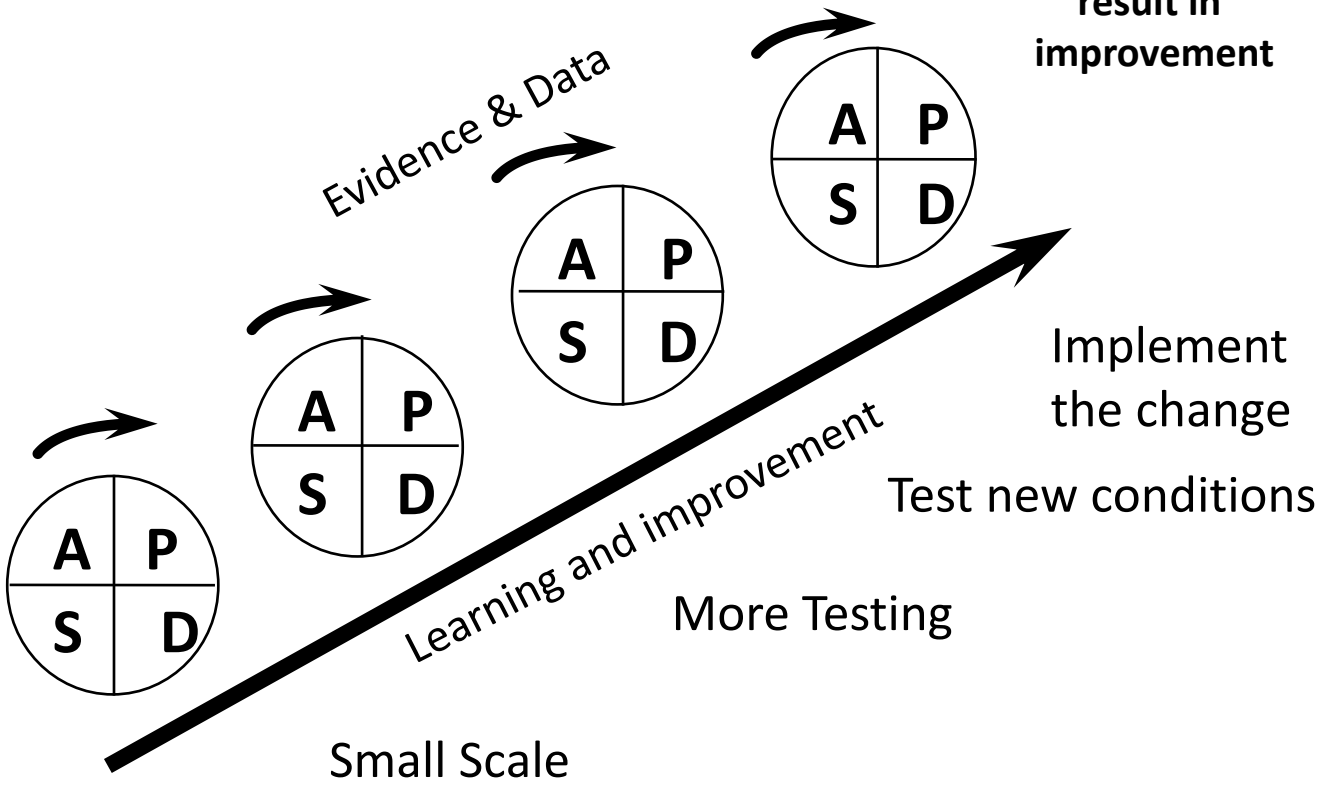
- Activate and empower parents to support their child's needs
- Develop care pathways across service sectors
- Improve flow to supports and services
- Increase capacity and capability of services & supports
- Actively manage care for a population

Improvement Drivers

- Cultivate accountable leadership focused on population outcomes
- Active participation in an organized process improvement effort
- Support the human element of change
- Measure & share data on how the system is working

Using Plan-Do-Study-Act (PDSA) Cycles for Sequential Building of Knowledge

Include a range of conditions in the sequence of tests, before implementing the change



Source: Associates in Process Improvement

The Goal of Measurement Is to Drive a Change

- Focus diverse partners on shared outcomes
- Establish shared accountability for reaching goal targets
- Shape an understanding of what matters and how to influence it
- Build and maintain enthusiasm for improvement
- Enable partners to think and work as a system
- Support improvement with frequent and real-time information

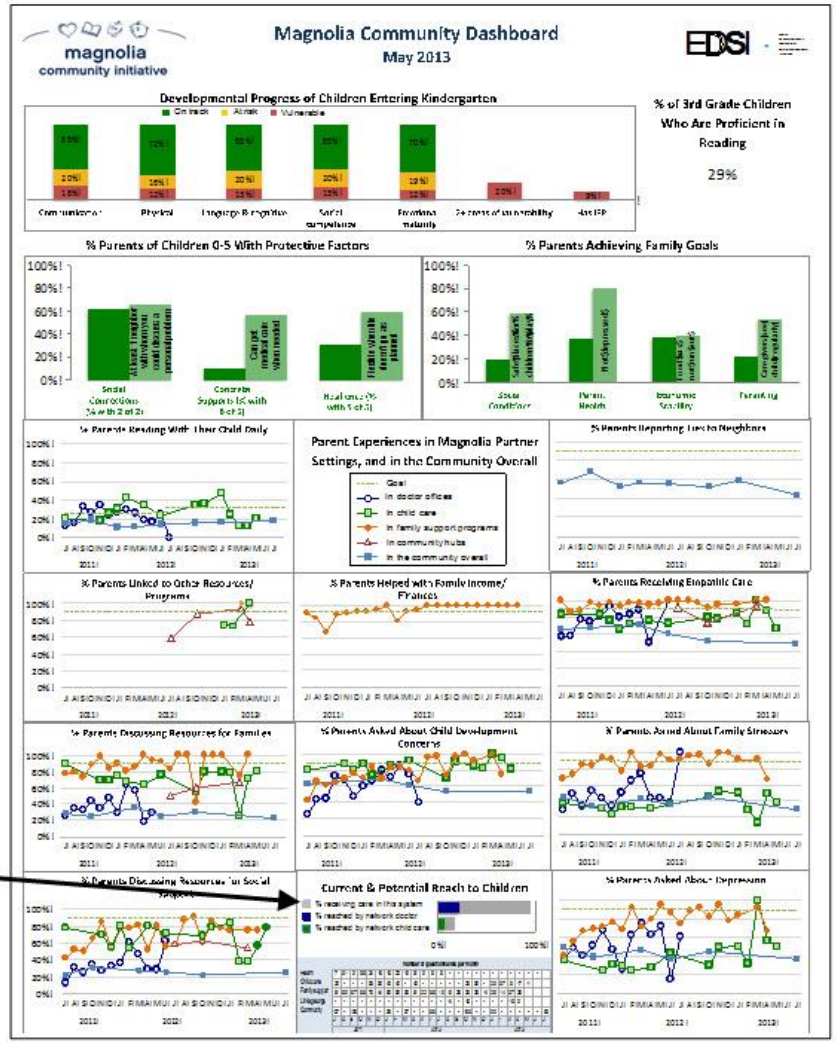
Community Dashboard

Developmental progress at school entry

Protective factors for families

Parent activities and behaviors

Potential and actual reach to children in the community

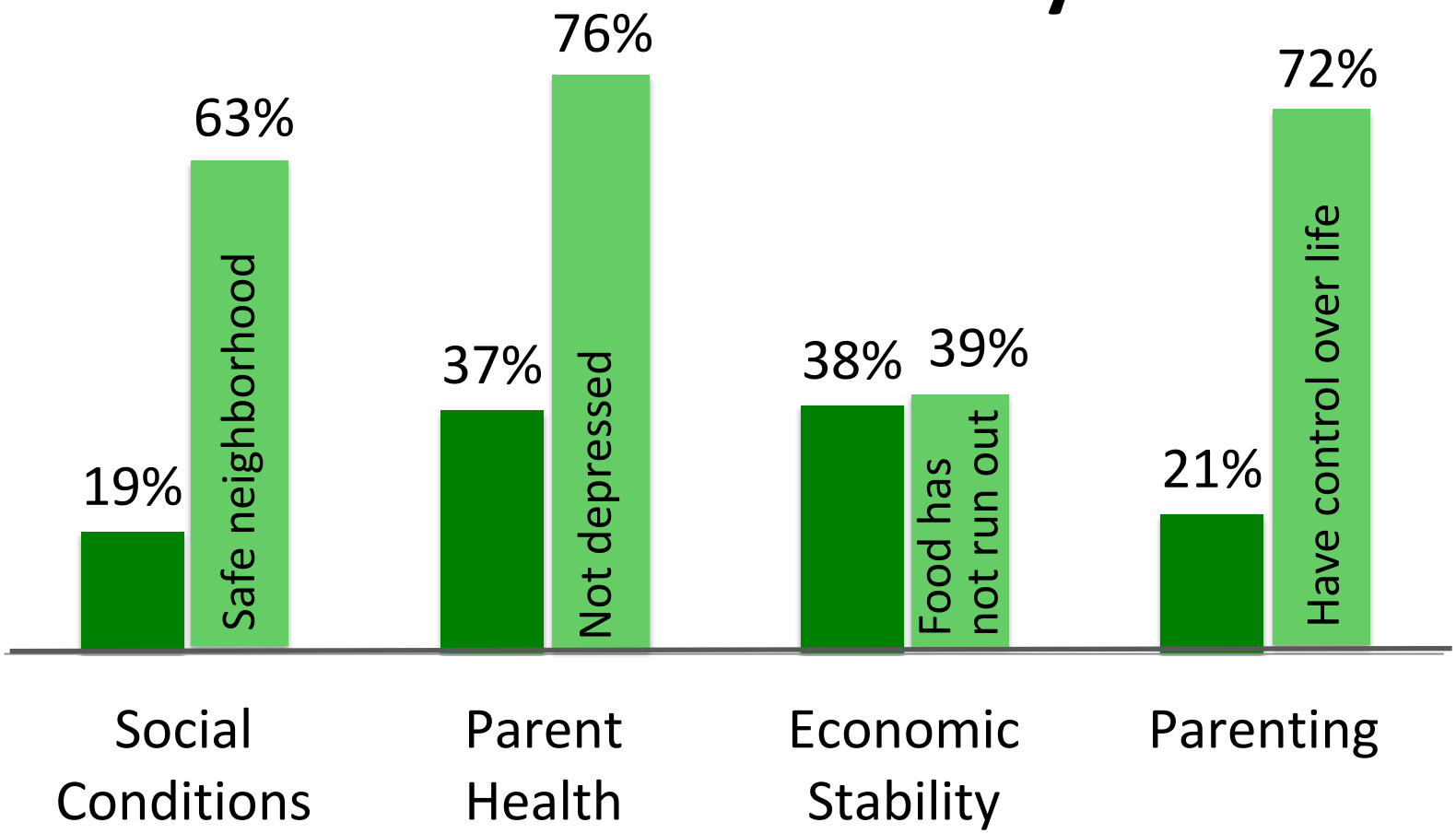


Reading proficiency, third grade

Family and community conditions

Measures of real-time improvement in services and supports

Understanding Family and Community Conditions



Measurement to Encourage "Systems Thinking"

Goal:

Improve flow to supports and services

Current parent conditions:

% of local parents with young children reporting social isolation: 41%

Current system practice:

% of parents linked with at least one network organization: 89%

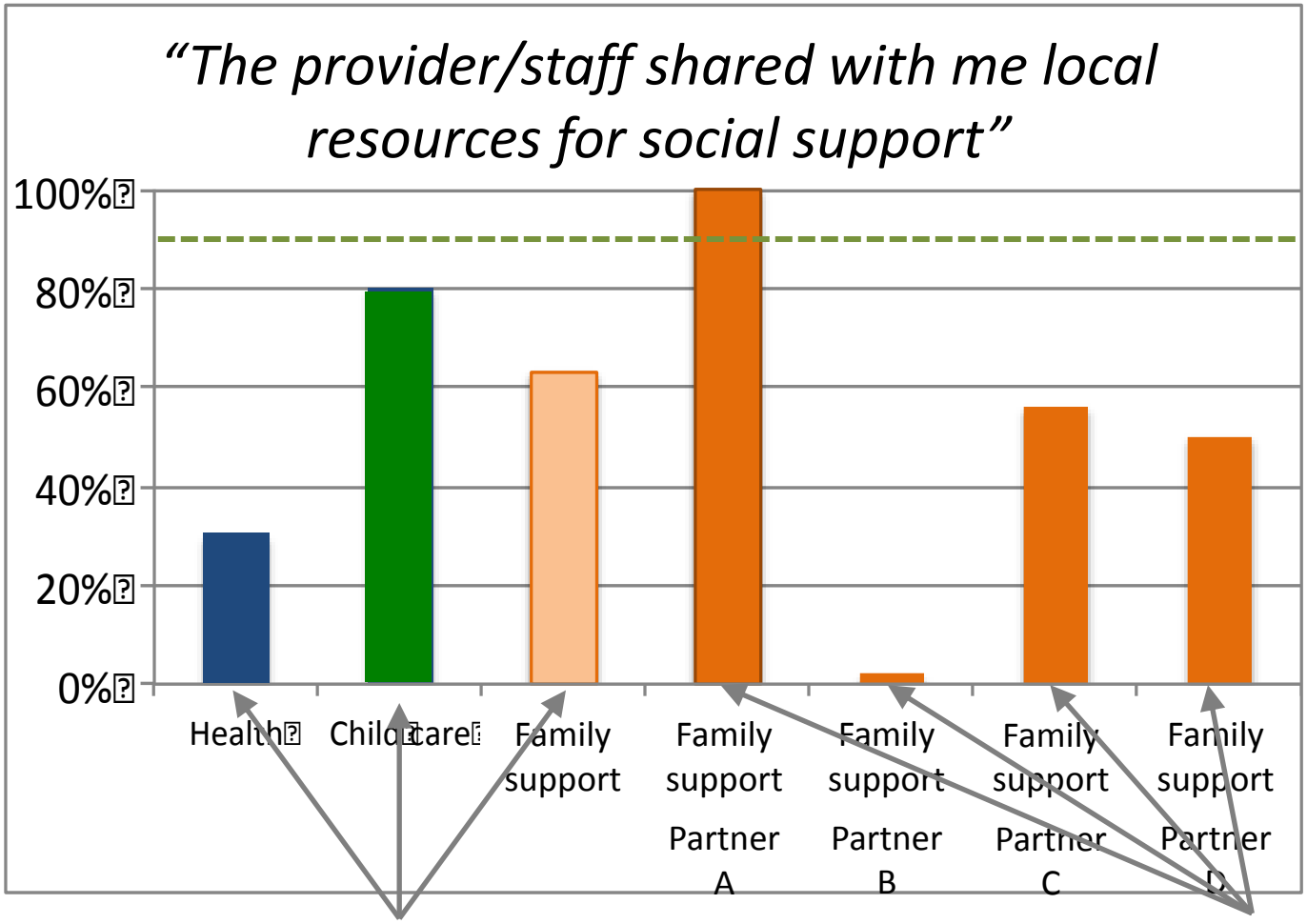
Average number (mode) of network organizations that parents are linked with: 1.0

We have good reach, poor depth.

Can we increase successful linkage of families to additional services/supports, across the network, by using

- a common set of prompting questions, and
- "warm handoffs"?

Measurement for Learning

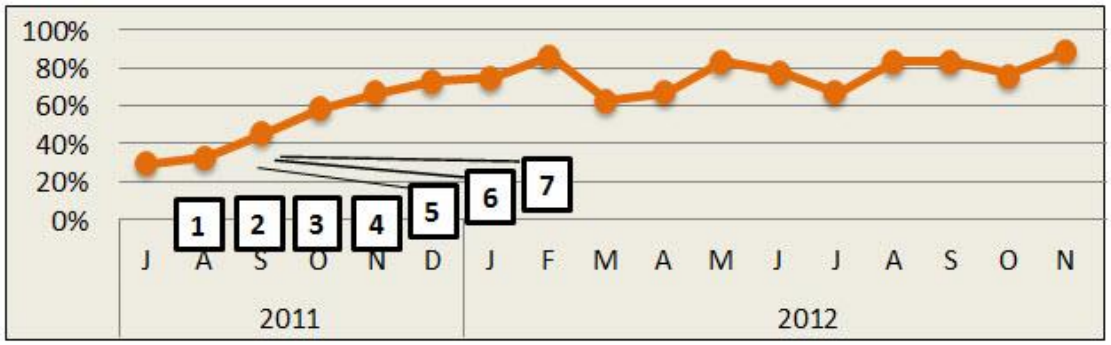


Goal target

Opportunities for learning <u>across</u> sectors

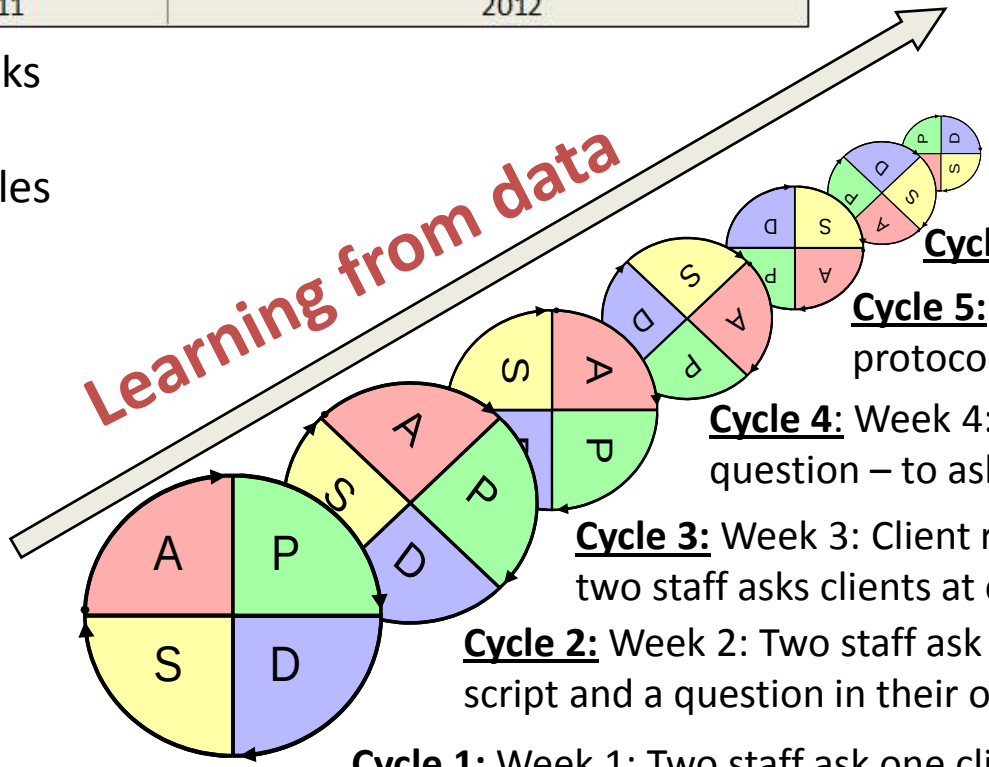
Opportunities for learning <u>within</u> sectors

Measurement for Improvement



Our measure tracks the impact of improvement cycles

90% of clients are asked about depression



Cycle 7: All staff begin the protocol; analyze failures

Cycle 6: All staff begin the protocol

Cycle 5: Week 5: Document the protocol (script, and question wording)

Cycle 4: Week 4: Test variation for asking the question – to ask in the same and "best" way

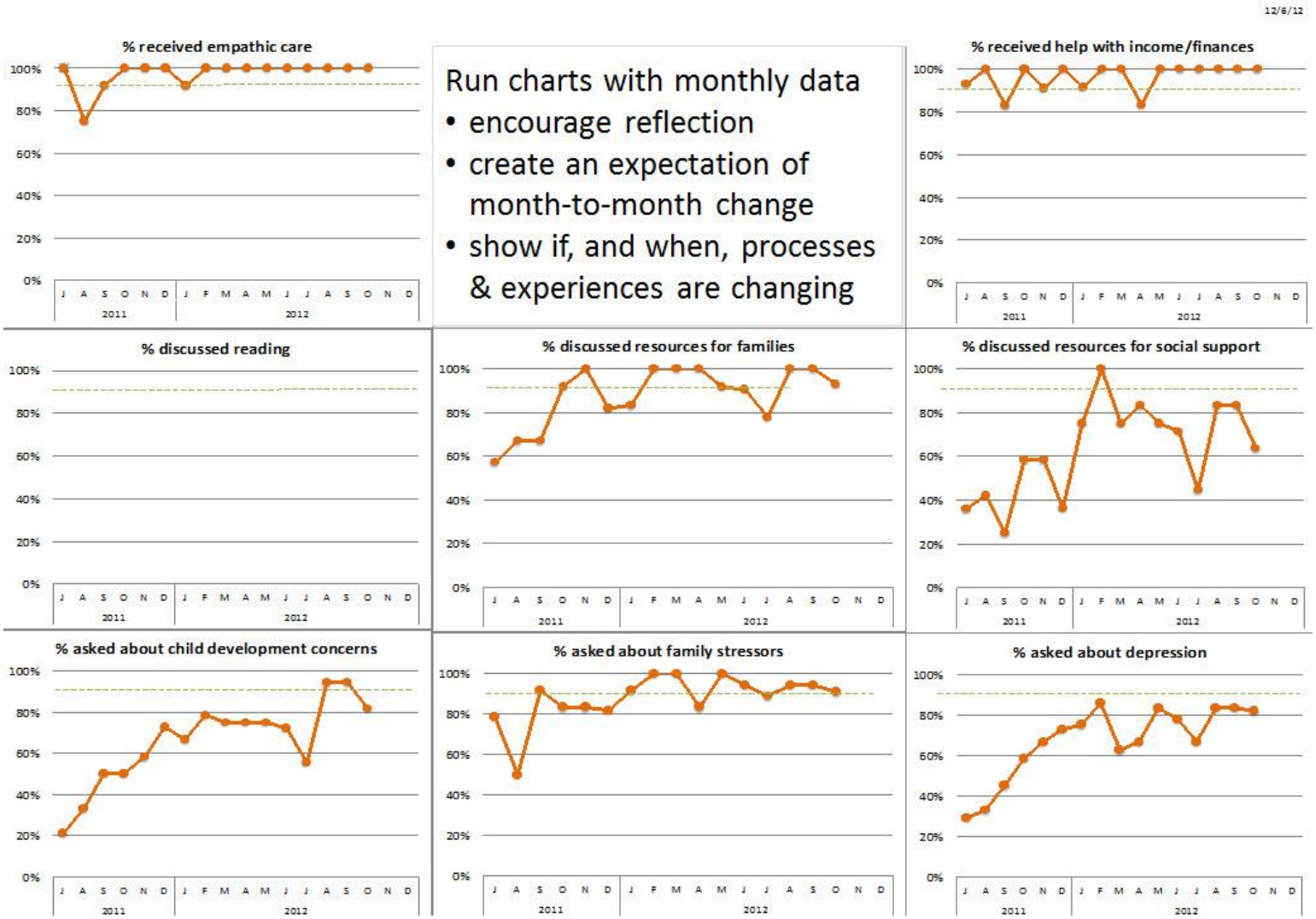
Cycle 3: Week 3: Client responds but flow isn't right; two staff asks clients at end of the visit

Cycle 2: Week 2: Two staff ask clients, using a 2-sentence script and a question in their own words

Cycle 1: Week 1: Two staff ask one client each, using PHQ-2 item

IDEA: Asking all clients about depression will identify people needing support

Frequent and Real Time Information



Effective Measurement

- A “family” of measures that represent the key influences on the outcome
- Measures that help partners reflect on their contributions and actions
- Measures that are feasible to collect – considering what can be scaled and spread
- Providing information about the system in “real time”
- Include expectations for change (numeric goal targets)
- Include measures reported by clients/residents, who are the “voice of the system”
- Embedded in a learning system

Avoiding Pitfalls of Measurement

- Too many measures
- Measures that cannot be changed by actions of the partners
- Measures that are too complicated to understand
- Collection and/or reporting takes too much time
- Display of measures does not quickly help tell the story
- Measures are not at the right level(s)

Summary

- No single program, or sector working alone can solve a population health problem
- An organized way of learning our way forward, using improvement methods, is a promising way of engaging many diverse stakeholders in solving complex problems

Conclusions

- Measurement can support learning in community systems by:
  - offering a visual display of what matters;
  - seeing the system we are trying to put in place;
  - showing expectations for change;
  - inspiring and tracking progress overall, and for each sector and partner.
- Measures need to be coupled with process improvement, to help diverse organizations move from planning to action

Questions?

You have two options to ask questions today:

1. Email us at: communities@stls.frb.org.
2. Type your question into the chat box of the webinar.

Wrap-up/Closing

Thank you to today's presenters and to all participants for joining this session.

Next steps:

- All session materials are available on our web site and in the next few days we will be posting an audio file of today's session.
- If you have topical suggestions for future sessions, or any questions about this program, please feel free to contact us at communities@stls.frb.org
- Information about future sessions will be posted on our website along with archived materials from past sessions: www.stlouisfed.org/connectingcommunities/

Related Resources

FedCommunities.org

Federal Reserve community development resources promoting people, place, practice and small business

In connection with this session, you can find a variety of additional resources on this topic , available at **www.fedcommunities.org**.

It provides an array of practical resources to help you in your role as a community development professional, whether you work involves supporting people, places, the practice of community development, or small business development.