

ARE WE HELPING?

THE NEED TO (RE)CONSIDER “WHAT WORKS” IN EVALUATION IF WE ARE TO TRULY UNDERSTAND “WHAT WORKS” IN COMMUNITIES

Katya Fels Smyth
The Full Frame Initiative
katya@fullframeinitiative.org



Coordinated Community Responses to Domestic Violence: Are They Effective? 14th annual conference of the BISC-MI
Battle Creek, MI
November 6, 2009

WHY THIS MATTERS TO ALL OF US



WHAT IS OFTEN VALUED IS NOT WHAT OFTEN WORKS

BY POLICY MAKERS AND FUNDERS
(FRAME APPROACH)

(PRINCIPLES OF THE FULL

Issues and problems are discrete;
respond to the specific diagnosis using proven protocols
+
Friends and family are part of the problem
+
Don't allow people to become dependent on a program—
graduate them fully
+
People need to leave their communities to move forward
+
Transferability matters
+
Consistency and fidelity
+
Linear, rational models and systems
+
The model matters more than who practices it
+
Demonstrate efficacy
+
Management matters.



Life is messy;
respond to the entire reality flexibly and recognizing that
every person's situation is different
+
Friends and family matter
+
Through thick and thin and difference
+
Be a community within the community, not an alternative
to the larger community
+
Place and local specificity matter
+
Some of the best work happens in the gray areas
+
Change is good: adapt and evolve an approach
+
It only works with the right people working
+
Be accountable: focus on results that are sustained
+
Leadership matters.

3

© 2009 The Full Frame Initiative www.fullframeinitiative.org

THE ROADMAP

- ❖ Our responsibility
- ❖ How we're supposed to know if we're delivering on that responsibility, and what's wrong with this picture
- ❖ Where this leaves us (and many, many others)

4



I. OUR RESPONSIBILITY

OUR RESPONSIBILITY

Our responsibility is, first and foremost, to the communities we are in— to the people who count on us to help them and those they care about.

We'd better be as sure as we can be that we're helping.

We are best equipped to make this determination, and to continue doing good, if we have tools that help us learn and if we are in an environment that rewards learning and improvement as an outcome in and of itself.

THE STATE OF THINGS

Good intentions are not enough, nor is common wisdom, although both are important.

There is increasing interest on spreading “what works,” so we need some way of determining what works.



...e thinking: Government and philanthropy
...titioners have embraced the idea and the
...vidence-based practice” – programs that
...ven.”

7

SO WHAT? (OR, WHAT I’M GOING TO TALK ABOUT)

In designing initiatives, if we take our guidance only from “proven” programs, (programs that have been found “evidence based” through random assignment experiments)

- we’ll never get to some of the most promising complex interventions, with their many interactive, moving parts
- we won’t have the information that will allow us to adapt effective programs to local circumstances and how these change, and to lessons learned in the course of implementation
- we won’t be able to design new responses to unsolved problems

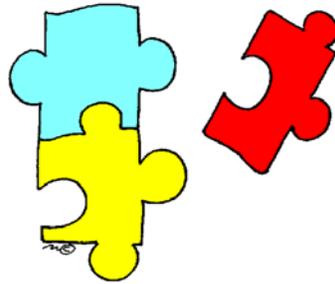
8

ACCOUNTABILITY MATTERS...

It's becoming clear that accountability is complicated.

All the "expert" knowledge about how to help in specific arenas sometimes doesn't add up to solutions that help marginalized people and communities in lasting and meaningful ways.

[How] can we judge whether we're making a difference?



9

**II. HOW WE'RE SUPPOSED TO
KNOW IF WE'RE DELIVERING
ON THAT RESPONSIBILITY, AND
WHAT'S WRONG WITH THIS
PICTURE**

ONCE UPON A TIME, THERE WERE TWO CHOICES

Anecdotes
Evidence of
good intentions
Ideology
Documentation of the
number and nature of
transactions

↓

“Unaccountable”

vs.

Evidence from
randomized controlled
trials

↓

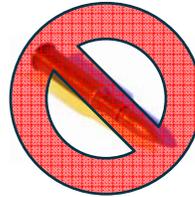
“Scientific validity”

11

THE ARGUMENTS AGAINST RCTS

Generally made:

- Too costly
- Time consuming
- Impractical
- Unethical
- Complex



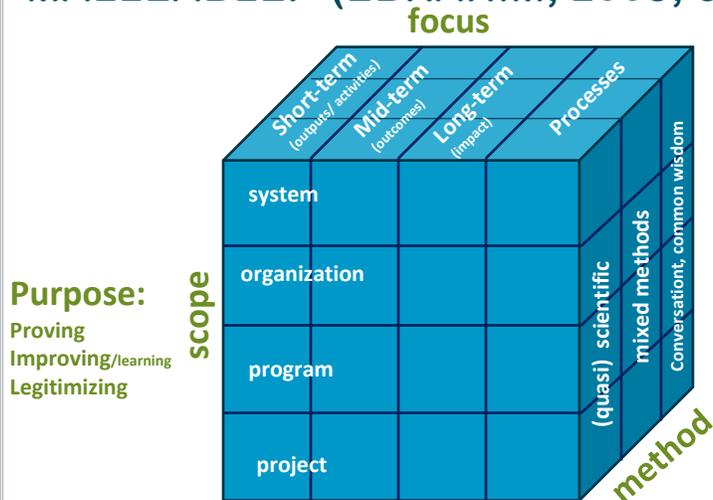
Extensive experience suggests:

RCTs alone do not further accountability, but must not be rejected. We need to understand efficacy *and* effectiveness.

“Evaluation practice [as it is currently construed] is essentially operating a scientific model in a political space.” (Louise Potvin, October 2009)

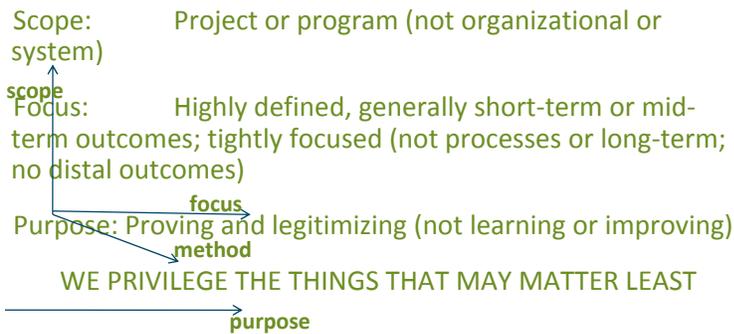
12

“ORGANIZATIONAL EFFECTIVENESS . . . CAN BE VIEWED AS HIGHLY MALLEABLE.” (EBRAHIM, 2005, 64).



Concepts of scope, focus, method matrix: after Murray and Cutt, 1998
Addition of purpose: Smyth, 2009

THE METHODOLOGY LIMITS SCOPE, FOCUS AND PURPOSE



Concepts of scope, focus, method matrix: after Murray and Cutt, 1998
Addition of purpose: Smyth, 2009

THE CURRENT FILTER: SAMPLE CHARACTERISTICS*

Program participants fit a clearly defined “profile” and all start at about the same baseline (and control when they don’t).

Setting: Context neutral -- what works in Boston works in Austin

Stasis: A model is refined and tested and then is not to be altered, adapted or modified without significant process and study.

Levels of impact: Individual and/or family change.

Power: expertise is largely held by professionals. Individual empowerment may be the goal of services; cohort, community empowerment generally left to community mobilization and organizing.

Key elements: a set pathway through program elements with clearly defined beginning and end.



“PROVEN”

** The presence of these characteristics is not a proxy for effectiveness, but unless we use a different filter, we won’t be able to differentiate those that work from those that don’t.*

15

WHAT DOESN’T FIT THROUGH THE CURRENT FILTER: EXAMPLES OF CHARACTERISTICS*

Population: multiply challenged, non homogeneous, potentially involved in multiple other interventions and systems; existing indigenous knowledge that is valued

Setting: community embeddedness, may engage in community change; community highly influenced by external forces, part of a larger system

Stasis: dynamic, evolving, emergent, demand-driven; significant front-line authority and flexibility

Levels of impact: multiple units of analysis (individual and community and system)

Power: participatory; flattened power dynamics; increasing capacity of a cohort or community, not just an individual; increasing voice

Key elements: a host of non-additive, individual program elements are drawn on, but without a specific recipe



“PROVEN”

** The presence of these characteristics is not a proxy for effectiveness, but unless we use a different filter, we won’t be able to differentiate those that work from those that don’t.*

16

WHAT'S AT STAKE IF WE DON'T USE A DIFFERENT FILTER



Potential for

- Distortion
- Irrelevance
- Missing what we're not looking for
- Not differentiating between necessary and sufficient
- Further marginalization of and harm to marginalized populations
- Reinforcing socially unjust policies and practices

17

CONSEQUENCE 1: DISTORTION

Of public policy:

- Conflation of elegance of evaluation with effectiveness
- Perpetuation of the "wild west" of unproven
- Privileging simplicity

Of program design and practice:

- Focus on what can be studied experimentally
- Focus on maintenance, not change; conformity, not innovation

18

CONSEQUENCE 2: IRRELEVANCE

Spending large sums to gather information that arrives too late.

Short-lived successes that don't add up.

Underestimating role of context and human involvement in studied program suggests a level of transferability that is rarely achieved.

Little to no understanding about what happens in the black box, so hard to know what are critical elements.

Getting caught completely off guard when the rules of the game change (what to do with an EBP when 50% of state funding is slashed?)

19

CONSEQUENCE 3: MISSING WHAT WE'RE NOT LOOKING FOR

Failure to see constraints and enhancements provided by setting and context.

Missing outcomes or side effects in domains beyond the scope of study (and that might ultimately tell us something about our theory of the problem).

Devaluation of ripple effects as "contamination."

Failure to recognize values and worldview as an important input.

Not examining, understanding, and valuing the role interventions may play in mitigating the undermining effects of issues such as poverty and violence.

20

CONSEQUENCE 4: NOT DIFFERENTIATING BETWEEN NECESSARY AND SUFFICIENT

Reinforcement of the myth of sole attribution
(myth: one program can change *everything!*)

Not being able to distinguish between failure that is a result of insufficient systemic supports and a program that is itself a failure.

Not understanding whether poor results in replication are a failure of the model, of implementation, of adaptation, of the system, or of using the model to leverage or combine with other necessary supports.

21

CONSEQUENCE 5: FURTHER MARGINALIZATION OF ALREADY MARGINALIZED POPULATIONS

Devaluation of local wisdom.



If people's situations, which are in some (perhaps large) part societally created, are of sufficient complexity that they cannot participate in any "proven" programs, or in RCT studies, those with the most to lose will lose the most.

"We have human service systems for people in need, not for people in greatest need." (Marya Hornbacher)

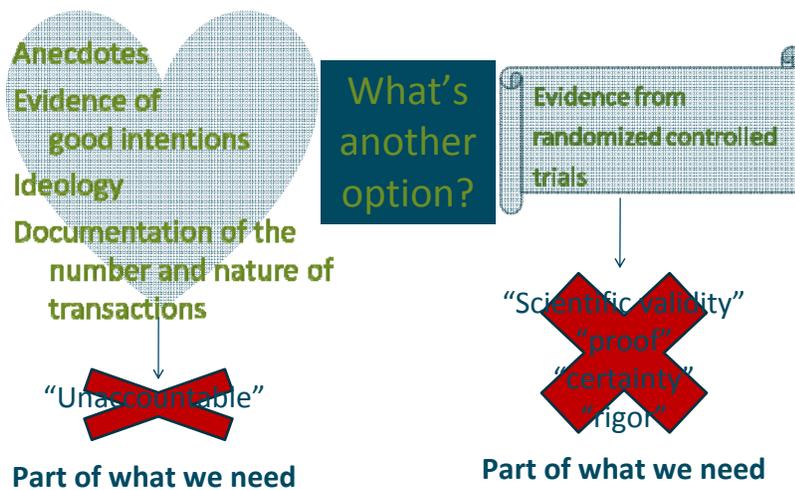
22

CONSEQUENCE 6: REINFORCING SOCIALLY UNJUST POLICIES AND PRACTICES

Privileging change at the individual and family level further obscures the role that social policies and practices play in perpetuating inequality and unequal access to resources, support and justice.



RETHINKING METHODOLOGY: IT'S NOT REALLY A CHOICE.



III. WHERE WE GO FROM HERE

A MORE COMPLETE APPROACH TO EVIDENCE AND EVALUATION

- Pays attention to systems
- Attends to interactions and amplifications
- Allows for complexity and change
- Uses mixed methods
- Incorporates multiple wisdoms
- Employs a philosophy of “enough”

PAYS ATTENTION TO SYSTEMS



Arrangements of parts– each necessary but not sufficient (some essential, some less than essential but enhancing).

Systems are more than this-- a system is the people in a given area interacting with the physical and policy environment and each other to secure and distribute resources.

No two systems are the same. Time, people, place, history, policies, resources– all combine.

All the elements may be right, but are not working together.

Research has intensively (over)studied how people change, and paid little attention to how communities and systems change.

Intervening on one level may generate change on another or elsewhere.

“The complexity of an unfolding system will always exceed what can be said about it.”
(Bonnie Duran, 2009)

27

ATTENDS TO INTERACTIONS AND AMPLIFICATIONS

- Making change ≠ sustaining change
- “Side effects” matter – a lot
- Monitor for where studied intervention may weaken the effect of another intervention
- Recognize that interventions are rarely neutral, and monitor for harm created for those who don’t find intervention beneficial

28

ALLOWS FOR COMPLEXITY AND CHANGE

- Recognize that the intervention is part of an ecosystem of interventions
- Privilege (instead of punish) mindful adaptation/ being demand-driven as a positive response
- Allow for different units and levels of analysis
- Understand progress isn't linear and unidirectional, that things don't get better at a constant rate, and may get worse before they get better.



29

USES MIXED METHODS

- Be participatory – every evaluation has biases and preconceptions.
- As appropriate, include:
 - ❖ theories of change,
 - ❖ experimental methods (RCTs have their place!),
 - ❖ case studies,
 - ❖ qualitative,
 - ❖ comparison of cohorts of participants' progress with community level data, their own baselines, other cohorts, or larger data sets.

N.B. This has huge implications for who hires whom for evaluation and how we select evaluation partners!

30

INCORPORATES MULTIPLE WISDOMS

- Practice knowledge is real knowledge (although flawed, as is any individual perspective)
- Draw on what is already known
- Define measurable outcomes that matter to those being asked to make the changes

Allow time necessary to achieve outcomes.

31

EMPLOYS A PHILOSOPHY OF ENOUGH

Generate enough evidence to make a robust determination of effectiveness, or lack thereof, quickly enough to allow for continuous improvements in program design and implementation, and in resource allocation decisions.

32

APPLYING A MORE COMPLETE APPROACH TO EVIDENCE MAKES POSSIBLE

- Rigorous evaluation of complex, evolving interventions
- Design of richer and more effective interventions.

33

A WELL STOCKED TOOL KIT

relevance + **rigor** → **actionable knowledge**

Is it the right thing to be assessing?

Are we assessing it right?

A clear understanding of what is known (and where the gaps are) to provide solid information for decision making

34

ACTIONABLE KNOWLEDGE MEANS:

Evaluation as a tool for learning and improving (imperfectly) and proving and sharing

Improving and proving are not sequential – they are concurrent

If there are funding partners willing to make this change with us, we have to be ready to go there.

35

IS IT THE RIGHT THING TO BE ASSESSING?

What we measure and pay attention to reflects underlying values and assumptions about the way the world works.

- Incremental changes (technological improvements): no changes in what we measure (seek better outcomes on same scales)
- Re-framing the problem (pressing on different levers, so monitor different outcomes) to things like systems, attitudes, relationships...

36

THIS BROADER APPROACH TO KNOWLEDGE IS A NEW MINDSET TO FIND WHAT WORKS

- It's not just about competing program models
e.g. Which home visiting program, which batterers' intervention program is most effective?
- It's not just about what's easy to count
e.g. In which programs do parents come away with more information about positive parenting practices?
- It's not just about isolated programs
e.g. Assessing programs without reference to how they work with one another, or how they are affected by their neighborhood and policy context .

37

ATTRIBUTION: WHAT ABOUT THE OUTCOMES THAT CANNOT BE ACHIEVED BY A SINGLE PROGRAM OR INSTITUTION?



Most of the significant outcomes for children, youth, families and communities are unlikely to be achieved by a single program or institution.

When many parts of the community contribute to achieving a certain outcome, it is impossible to attribute causation (i.e. credit or blame) to any one agency or program or family with certainty. However, *whether* outcomes are being achieved will be clear, and targeted data collection and keen observation will allow for informed judgments of whether all contributors are pulling their weight.

Being able to manage in circumstances where the role of each participating entity in bringing about success is not clearly distinguished will require a mind-shift and experimentation with new roles for funders and community-based interventions.

38

IMPLICATIONS OF THINKING ABOUT SYSTEMS:

We have to understand that not only can individual pieces fail to live up to their promise— we all have to look past the end of our noses so that edges blur a little for survivors and there aren't yawning cracks for people to fall into.

Redundancy is not the same as duplication— it's a safety feature. A system with some redundancy built in may ultimately be more efficient, because it never operates perfectly.

Barriers occur at the boundaries of programs far more than within them.

We have to honor that relationships matter: practitioner – participant, participant-participant, practitioner-practitioner, with government, funders, among programs— they all matter.

39

BECAUSE NO ONE PROGRAM OR EVEN SYSTEM CAN ACHIEVE VALUED OUTCOMES ALONE...

Requires hard-to-fund community capacity

- To measure, document and report progress *without insisting on definitive attribution*
- To encourage and facilitate work across systems and disciplines
- To perform cross-cutting monitoring and gap-filling functions
- To provide services that individual programs cannot undertake on their own (e.g. to help assure a quality workforce that is stable and committed)
- To help identify the need for mid-course corrections

40

WE ARE NOT ALONE

In Medicine

Roundtable on Evidence-Based Medicine of the Institute of Medicine:

“The current dependence on the randomized controlled clinical trial (RCT), is inadequate today and may be irrelevant tomorrow, as it seems to be useful only in increasingly limited circumstances (including a narrow range of illnesses and the absence of multiple problems in an individual patient).”

The Roundtable called for a re-examination of what constitutes evidence, and suggested that RCT should not continue to be considered the gold standard.

41

IN INTERNATIONAL DEVELOPMENT

Michael Woolcock, Research Director, Brooks World Poverty Institute, University of Manchester:

Acquiring needed knowledge will not be a product of simply deploying what some deem to be ‘gold standard’ evaluation protocols (e.g., randomised trials) per se but, rather, deep engagement with the contexts and processes within which all projects are embedded, and calling upon the full arsenal of research tools (qualitative, quantitative and historical) available to social scientists.”

... AND IN EDUCATION, CHILD WELFARE, INJURY PREVENTION, AND MORE...

42

OUR RESPONSIBILITY IS TO:

- work within and throughout our systems to help people who are struggling become better off.
- employ tools that tell us whether what we are doing is actually working, and make changes if it's not.
- be literate about the tools others are imposing on us, so that we can be part of informing whether these are the right tools (and what the questions are) before we see the results.

Being accountable to our communities requires that we actively seek out and employ rigorous and relevant means to assess impact and adjust our course.

43

QUESTIONS AND COMMENTS