SHARED MEASUREMENT |
THE WHY IS CLEAR, THE HOW CONTINUES TO DEVELOP

MARK CABAJ

This paper explores five practical shared measurement challenges that, if not handled well, can weaken a group's ability to evaluate and manage their Collective Impact effort.

PREFACE

The practice of shared measurement is one of the five conditions of Collective Impact (CI), an approach that guides diverse organizations to tackle tough social, economic and environmental challenges at scale.

The case for shared measurement is clear:

> *Developing a shared measurement system is essential to collective impact. Agreement on a common agenda is illusionary without agreement on the ways success will be measured and reported. Collecting data and measuring results consistently on a short list of indicators at the community level and across all participating organisations not only ensures that all efforts remain aligned, it also enables the participants to hold each other accountable and learn from each other's successes and failures (Kania & Kramer 2011).*

The practice of shared measurement is still developing.

The field of measurement is not new. In the 1920s, Charles Winslow Taylor rigorously measured the efficiency of different ways of organizing work in industrial enterprise, a practice that eventually heralded a measurement revolution in the private sector. In the 1930s, educators began to develop standardized tests to assess the progress of students in such areas as math and science. In World War II, government officials sought to assess the effectiveness of propaganda campaigns on the morale of the general population. In the 1960s and 1970s, researchers and evaluators created sophisticated measurement tools, such as randomized controlled trials, to assess the effectiveness of policies and programs that accompanied the growth of the welfare state (e.g., employment counseling, pre-kindergarten programs,

affordable housing). In the 1990s, executives and managers in governments, philanthropies and non-profit organizations widely embraced measurement as part of the "new managerialism" to public administration (Jayne 2009). It eventually led to large hospitals, food banks, and early childhood development centres learning how to fill out logic models (Cyton, Petit, Kingsley 2014).

Shared measurement is a new niche for the measurement field and it is infinitely more complex and demanding. To the long list of generic political, conceptual, methodological and logistical problems that come with trying to measure changes and operations in organizations and communities, shared measurement adds the challenges of:

Table 1: The Five Conditions of Collective Impact (Kania et al. 2011)	
Common Agenda	All participants share a vision for change including a common understanding of the problem and a joint approach to solving it, through agreed-upon actions.
Shared Measurement	Collecting data and measuring results consistently across all participants ensures efforts remain aligned and participants hold each other accountable.
Mutually Reinforcing Activities	Participant activities must be differentiated while still being coordinated through a mutually-reinforcing plan of action.
Continuous Communication	Consistent and open communication is needed across the many players to build trust, assure mutual objectives and create common motivation.
Backbone Support	Creating and managing collective impact requires one or more separate organizations with staff and a specific skill set to serve as the backbone for the entire initiative and coordinate participating organizations and agencies.

- Working with diverse organizations (sometimes in the hundreds)
- Working across multiple domains (e.g., health, education, housing), at multiple scales (e.g., programs, neighbourhoods, cities and regions)
- A focus on more complex phenomenon (e.g., poverty)

"Shared measurement," sighed the participant of a Tamarack workshop on evaluation, "is like traditional measurement on steroids."

Collective Impact (CI) advocates and participants are busily developing shared measurement systems and practices through a process of trial and error. They regularly document their ideas, case studies and methods and share them on websites, peer exchanges and at conferences. Happily, several organizations are devoted to tracking and assessing their progress. This includes, of course, Foundation Strategies Group (FSG) in the United States (Preskill, Parkhurst, Splansky-Juster 2014), the Federal Reserve Bank of San Francisco (Cyton et al. 2014), New Partnership Capital (NPC) in the United Kingdom (Ni Ogain, De Las Casa, Svistak 2013; Vliet, Kail, Whargon 2014), Social Venture Partners and the Centre for Social Impact (CSI) in Australia, to name a few. Many of these helpful cases and resources are located on the FSG-sponsored Collective Impact Forum, https://collectiveimpactforum.org.

TAMARACK
I N S T I T U T E

FIVE IDEAS

I have already expressed some of my reflections on shared measurement in other articles (Cabaj 2014, Auspos & Cabaj 2015). In short, I believe (a) that shared measurement is a necessary part of any community-wide effort to tackle a complex challenge; (b) that it must be embedded in a larger framework for strategic and learning and evaluation in order to be useful; and, (c) that the complexities of shared measurement work are significant enough that CI participants – and the funders and evaluators who support them – should proceed with caution.

This paper explores five practical shared measurement challenges that, if not handled well, can weaken a group's ability to evaluate and manage their CI effort:

1. Ensuring that shared measures are organized in a way that reflects the group's evolving strategy or theory of change
2. Distinguishing between shared outcomes, measures and measurements.
3. Creating good sense-making and decision-making processes
4. Employing a mix of "big design" and "agile" approaches to the development of shared measurement systems
5. Acknowledging, monitoring and responding to a variety of perverse behaviours that often emerge with measurement processes

This list is far from comprehensive. It simply captures some of the insights that have emerged for me as I have worked with – or studied – CI efforts across Canada, the United States, Australia and New Zealand in the last few years.

1. MATCHING SHARED MEASUREMENT SYSTEM TO THE EVOLUTION OF A GROUP'S "RESULTS FRAMEWORK"

Many CI participants and evaluators often rush to select the best measures for their CI effort before they have fully developed their common agenda and strategy.

It's usually a struggle. Typically, they can generate a list of possible indicators to track progress towards their long-term goal or common agenda (e.g., ending homelessness, graduating high school, avoiding the criminal justice system). However, they don't yet have a results framework, theory of change or outcome pathways to guide their selection of short- and medium-term outcomes and measures. As a result, their conversations often spin around such issues such as which indicators *could* they collect – rather than which *should* they collect.

Take, for example, the experience of a coalition of agencies in Surrey. They agreed to reduce the number of people who were working (at least) part-time, but were unable to pull together sufficient outcomes to get people off the streets. While network members were clear about how to measure their final goal – an annual community-wide homelessness count carried by volunteers – they were unclear about the strategies, outcomes and measures to get there.

TAMARACK INSTITUTE

Rather than try to develop a strategy in a boardroom, they decided to develop it by trying things out in the field. This involved following homeless labourers as they went about their day-to-day activities, identifying the various obstacles they encountered along the way, and developing solutions through a process of trial and error.

Their organic approach was productive. They discovered that homeless labourers had limited access to shelters because agency staff required proof of social assistance and closed their doors too early for workers cleaning up construction sites in the evening. So, the coalition convinced local shelters to extend the hours and criteria for admission. Then it became clear that temporary employment agencies hiring the homelessness to clean up construction sites for builders were charging the labourers an excessively high fee for the service. So, the coalition convinced local non-profit employment groups to perform a liaison function with builders free of charge. This resulted in a doubling of labourers' income. Another issue involved the ineligibility of people with no home address for a bank account. Companies complained that this prevented them from paying part-time workers through automatic deposit. So, the coalition worked with a local credit union to open bank accounts and savings plans.

Within 18 months, after multiple iterations of similar interventions, the group saw a significant reduction in the number of homeless labourers.

The example illustrates the range of strategies available to CI participants, as well as the implications for evaluation and shared measurement. In the early days of their efforts, the Surrey coalition avoided discussions on shared measures entirely, focusing instead on getting real-time feedback on specific initiatives and seeking out patterns in what did and did not work.

Over time, they began to see such patterns, and organized their work around expanding access to shelters and permanent housing, increasing earned income and growing financial savings. They then created a framework or umbrella strategy which focused on these outcome areas or pathways. Still, they allowed CI participants a great deal of flexibility in the activities and interventions their members could choose to contribute. Eventually, after more experimentation, the group refined their model and packaged their work into a formal strategy or model, which laid out the activities and processes in more detail. (See Table 2.)

The Surrey example illustrates how important it is that CI participants match their evaluation and measurement approach to their strategy. It is unhelpful to try to impose measures in emergent situations beyond those that monitor progress on the ultimate goal (e.g., ending homelessness). CI participants are simply not clear which additional outcomes merit measurement until they have tried things out on the ground. Once CI participants are confident enough to commit to umbrella and detailed strategies, they can commit to concrete measures and sophisticated measurement practices.[1]

---

[1] This is also the conclusion reached by researchers at NPC (Van Vliet, Kail, Wharton 2014).

TAMARACK INSTITUTE

Table 2: The Strategy Continuum[2]

	Emergent Strategy	**Umbrella Strategy**	**Detailed Strategy**
Strategy Type	CI participants agree on the ultimate goal, but aim to develop the pathways to that goal through experimentation.	CI participants agree on the ultimate goal and broad pathways to get there, but allow for a great deal of flexibility in terms of specific activities and initiatives.	CI Participants agree on the ultimate goal and broad pathways, and are able to lay out a detailed set of activities and initiatives to get there.
Evaluation & Shared Measures	Securing real-time feedback on experimental initiatives and seeking out patterns, to see which pathways are most promising.	Providing feedback on strategies and progress on a select number of measures, as well as processes to identify improved or new strategies.	Providing feedback on strategies and progress on a larger set of measures as well as processes to identify improved or new strategies.

2.

DISTINGUISHING BETWEEN SHARED OUTCOMES, SHARED MEASURES & SHARED MEASUREMENTS

CI participants often get confused about the distinction between shared outcomes, shared measures and shared measurement. As a result, they often invest more time and attention on shared measurement than is necessary.

In some cases, the three are completely aligned. Take, for instance, a typical CI effort to end homelessness (there are over 100 such efforts in North America). Their participants generally feel it is important to reduce the time someone spends in a city's network of shelters before they are placed in more permanent housing. This is a shared outcome. This can easily be turned into a measure, such as the number of days in the shelter system, often accompanied by a target of 7-14 days of registration. This is a shared measure. Finally, most groups conclude that their individual organizations' existing data are not aligned enough to track such things. They therefore are willing to develop a shared measurement system, even shared databases, to ensure that they are on the same page. This is shared measurement. (See Table 3.)

---

[2] See Auspos, Pat and Cabaj, Mark (2015).

TAMARACK I N S T I T U T E

Table 3: Shared Outcomes, Measures & Measurements

	Definition	Example: Ending Homelessness	Example: Increasing High School Graduation Rates
Shared Outcomes	Short- to long-term outcomes around which all CI participants agree to align their efforts to achieve that impact.	Rapidly re-housing people who become chronically homeless.	Assisting students' achievement of developmental milestones from pre-school through high school.
Shared Measures	Measures that CI participants feel are important to track to assess the overall progress of their work.	The length of time which people who are homeless spend in city shelters before moving on to permanent housing.	The percentage of students achieving a certain score on standardized grade three reading tests.
Shared Measurement	A set of techniques or processes that require diverse organizations to employ the same techniques for gathering, analyzing and reporting data.	Shelter agencies agree to use a shared database where they all enter the service data, including length of stay.	The school district already collects standardized data on this measure, so the other 20+ CI partners do not need to be involved in gathering this data.

In some instances, the pool of shared measures might be quite small. This is illustrated in the approach taken by a CI initiative seeking to improve high school graduation in a Canadian city. The participants agreed to a comprehensive framework organizing key outcomes and measures that students should achieve from pre-school to the end of high school. They then discovered that most of these measures – such as students' scores on standardized grade three reading test, a key predictor of later success – were already collected by the school district. As a result, the group asked, "Why would the other CI partners – e.g., mentoring organizations, counseling organizations, youth development programs – need to do anything more?"

Creating shared measurements just for the sake of having shared measurements is silly and inefficient. They can take a long time, and a lot of energy and resources to develop and they always require an "extra effort" from agencies to collect and report. Even well-meaning funder efforts to create a common measurement system can go too far. I interviewed an executive director on her experience of using a new measurement and reporting system developed by multiple funding organizations in a large city in the United States who sought to "streamline" reporting requirements for their grantees. She shared her appreciation of their efforts, but then confided:

> *It's good that many of us are tracking the same things, but it all went a bit overboard. We seem to track anything that moves – including the number of*

diapers we hand out, which is crazy. I think we just got caught up with the idea that we should all be tracking the same thing. It actually takes a lot of time to manage this all and I am pretty sure they only use a tiny bit of it.[3]

What is the ideal number of shared measures for a CI initiative? It's difficult to say. The advice offered in many publications is 12-14. This is unhelpful. Most initiatives to end homelessness, for example, employ many more than that in their sophisticated shared data systems, while the high school graduation initiative employed half that amount. The better response seems to be "as many as the CI participants feel are essential to tracking their progress and do not outstrip their collective capacity to manage and use effectively."

How much is that? Each group needs to find that out for itself. Remembering the difference between shared outcomes, measures and measurements will make it a little bit easier.

> **What is the ideal number of shared measures for a CI initiative? As many as the participants feel are essential to tracking their progress and do not outstrip their collective capacity to manage and use effectively.**

3. DEVELOP SENSE-MAKING & DECISIONS TOOLS

CI participants can become consumed with the production of data. In fact, they can become so consumed that they forget to pay attention to developing robust processes that (a) make sense of data, and (b) use it to help make decisions about the CI strategy and operations.

Take, for example, the reflections of Srilatha Batliwala on the results of weak interpretation and decision-making with respect to shared measurement data in a rural development project:

> *"I worked in a community health project in a rural area of western India in the 1970s. One of our goals was to eradicate diarrhoeal deaths of young children and thus bring down the level of child mortality. Village health workers were required to report the number of diarrhoea cases treated at a monthly meeting. Within a year, at one of these meetings, our director berated several health workers who had no diarrhoea cases to report from their villages. This was not taken as a sign that our strategy was working but of the poor performance of the workers in failing to report cases".*[4]

---

[3] This is from a personal conversation with an agency in Edmonton in 2015.
[4] See: http://www.alliancemagazine.org/analysis/measuring-social-change-assumptions-myths-and-realities/

In this case, the project director jumped to conclusions about the results, without pausing to dig deeper into the data. Are health workers tracking and reporting data differently? Is the difference in performance due to sloppy execution on part of the health workers ("execution failure"), or an imperfect program ("strategy failure")? Or is it because they work in very different villages, some of which mistrust outsiders and therefore are reluctant to participate in the program ('failure to customize strategies to context")? Do funders and management have unrealistic expectations about project outcomes? Without a deliberate process to interpret the data, draw conclusions and make data-informed decisions, data collection is often a waste of time and energy.

So many CI efforts are rich in data, but poor in sense-making and decision-making. In his book *Blink*, Malcolm Gladwell summarizes a lot of the research on this problem and uses fine examples to demonstrate how it leads to bad judgments (Gladwell 2005).

It can also paradoxically lead to more work. When CI participants are unsure what the data might mean, they often conclude they need more data, and initiate an expanded round of data collection, in the hope that some magical insight will emerge in the next round. Unfortunately, this usually adds to the "data noise," rather than helping to create useful feedback on a group's strategy. This may push some innovators to lose interest in evaluation and measurement entirely.

CI participants who are serious about being data- and evidence-informed must work hard to include sense-making and decision-making processes when designing their evaluation and shared measurement system. They can be quite inventive in this regard. Paul Born, an experienced CI pioneer, often advises CI participants to avoid collecting new data in the early days of the effort. Instead, he says, they should focus on getting a handle on the data that they already collect, use it for a year, and then consider the possibility of collecting new data. In addition to building their sense-making and decision-making "muscles," this often reveals that they may not need to collect as much new data as they originally thought.

Another technique is to create a prototype of a process to review, interpret and use data and see what does and does not work early in the shared measurement process. This involves creating a hypothetical set of data for the measures and questions requested by CI participants. They then simulate a "use" session to see what kinds of questions the data raises, the conclusions they would draw and decisions they would make.[5] This allows everyone to experiment with sense-making and decision-making processes. It also tests whether people are getting the data they feel that they need.

Whatever practices a CI group chooses to begin with, it should treat them as just that, a beginning. Interpreting data and making good decisions requires continuous experimentation and improvement. While writing this article, for example, I picked up the latest *Harvard Business Review*, entitled "The Leader's Guide to Problem Solving" (HBR 2017). It covers topics

---

[5] Tom Kelly, formerly the Director of Evaluation for the Annie E. Casey Foundation, and now Vice President of Knowledge, Evaluation and Learning, with the Hawaii Community Foundation, shared his "data rehearsal" during a Tamarack evaluation workshop in 2013.

published in the same magazine over 20 years ago, including "How to Tackle Your Toughest Decisions," "Outsmart Your Own Biases," and "Learning in the Thick of It." If executives of Fortune 500 companies are endlessly preoccupied with building their capacity in these areas, participants of CI efforts, who arguably are tackling even more complex challenges, should do the same.

4. COMBINING A BIG DESIGN & AN AGILE APPROACH

One of the most consistent pieces of advice offered by advocates of shared measurement systems is that organizations should develop their approach to evaluation and measurement through a traditional planning and design process. In other words, they should:

1. Discover the information needs and requirements of those requesting data.
2. Design one or more options to meet those needs.
3. Develop the design further through testing and consultation.
4. Implement the final design and refine the kinks as they emerge.

The strengths of this approach – which software designers call the Waterfall or Big Design Up Front (BDUF) – are clear. It is comprehensive, systematic and logical. It is also easy to communicate, manage and fund. Finally, it appeals to a group of diverse CI participants who may crave some level of predictability amidst the usual messiness of working together on tough social problems.

The limitations of this approach are equally apparent. It can be expensive and time-consuming. It can result in unwieldy processes that are difficult to adapt once the final design is complete and implementation begins. Annual surveys of database and management information projects across the world show that Waterfall-organized projects tend to arrive late, go over-budget and (eventually) get discarded (The Standish Group 2012). Why? Zaid Hassan, a veteran social innovator that has worked on a wide array of complex situations across the world, argues that these projects are based on the mistaken assumption that we can develop a perfect knowledge of the world, can get things right the first time, and that the context in which we work doesn't change (Hassan 2014).

> **The Agile approach typically results in lower costs and risks of development. It also makes it easier to continually adapt data-systems in response to the users' evolving "requirements."**

The larger the CI effort, the more pronounced these limitations become. Annie E. Casey's evaluation of its *Making Connections initiative*, a long-term and well-funded collection of CI-type initiatives designed to improve social and economic outcomes in vulnerable neighbourhoods, has been the subject of a first-class case study. It illustrates just how difficult it

TAMARACK
INSTITUTE

is to adapt a large-scale evaluation design that can keep up to an ever-evolving initiative. Midway through the initiative the Foundation's leadership realized that a comprehensive longitudinal survey of neighbourhood residents' wellbeing (its "big, expensive bet for capturing change") might no longer be the only or the best way to track progress (Fiester 2011).

Software designers are well aware of difficulties posed by Waterfall and BDUF. In 2001, a committed group came together to develop an alternative model: the Agile approach. It turns the Waterfall approach on its head: the Agile approach focuses on developing small elements of data-systems through multiple cycles of testing and adaptation (often called "sprints"). Data experts and data users work together to interconnect each of these small elements overtime.

The Agile approach typically results in lower costs and risks of development. It also makes it easier to continually adapt data-systems in response to the users' evolving "requirements." Agile practitioners fully anticipate the kinds of change experienced by the Making Connections participants and are prepared to quickly drop, add or upgrade parts of the data-systems, as required.

Waterfall & Agile approaches[6]

TAMARACK INSTITUTE
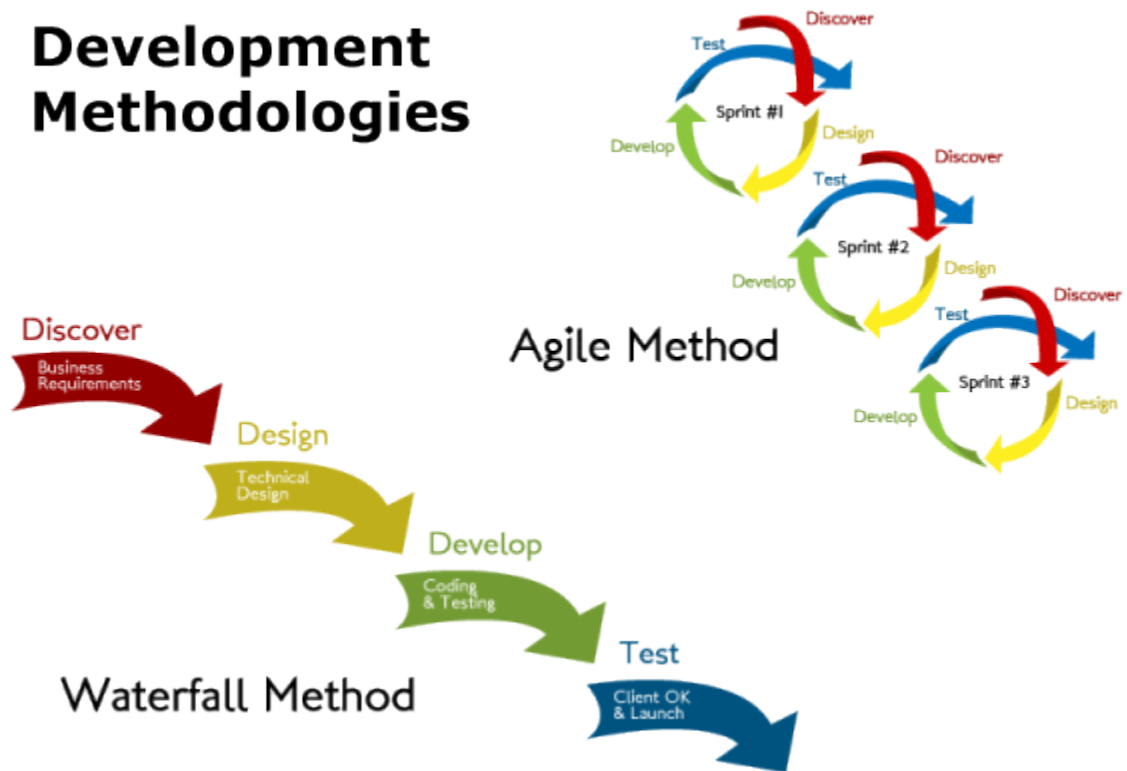
The Agile approach is well-suited to CI initiatives. Take, for example, the 24/7 Diversion initiative in Edmonton. This group of community agencies, police and health services agreed to work together to divert the large number of public calls taken by police services and hospital emergency wards that were not truly crises (e.g., intoxication, arguments in the street, sleeping in alleys). This would free up overtaxed police and health services staff. In addition, it would directly utilize social agencies to respond to short-term needs (e.g., emergency shelter, support of mental health worker, provision of food) and then to offer longer-term support to help prevent personal crises in the future (e.g., housing, education, income support programs).

In response to pressure to put diversion teams on the street quickly, the group decided to organize their work around 90-day cycles, each focused on developing 3-5 tangible "collaborative diversion practices." These included simple things: developing a protocol for communicating with the 211-dispatch line; creating a safety protocol for teams arriving on the scene of a reported crisis; figuring out how teams would transport people to emergency services; and developing messages for the public. After each 90-day cycle, the diverse group reflected on what they achieved, identified what had and had not worked and laid out a new set of practices to be invented or improved in the next 90 days.

Their agile work included the development of a "real time app," the centerpiece of their evaluation and measurement strategy. The purpose of this handheld device is to provide a way for all partners to measure and report key element processes and outcomes of their diversion model in "real time." This includes such things as the time to tackle a diversion (receive, respond, and resolve diversion calls), the patterns of crisis and client characteristics (e.g., intoxication, young male between 18 and 25), the various types of "hand off" after the incident (e.g., shelter, counselling, connection to family), and (eventually) the longer wellbeing of clients receiving post-crisis support.

The evaluation and measurement conversations at the 90-day review were spirited and curious. What can we do to make it easier for staff to input data into the device? Is there a way to aggregate that data more quickly? Are there any data that we are not using and perhaps should drop? Does our data on diverted cases align with the data held by police and medical services? Can we work with agencies that provide emergency and post-crisis supports to track over the longer term the wellbeing of people assisted by the service? After each review, the backbone group and evaluators had enough input to drive their next cycle of experimentation.

Within a year, the collaborative managed to establish a fully functioning diversion model, add several new diversion teams and expand from the downtown location to two other neighbourhoods. Most importantly, they developed ambitious targets for the number of diverted calls, a number they exceeded in the beginning of Year Two. "It's the most productive collaborative in which I've ever been involved," noted one veteran of inner-city coalitions. Two other partners were so impressed that they introduced the Agile approach into their own organizations.

The success of 24/7 Diversion reflects a larger pattern. Agile projects are far more likely to be rated as successful by their users than Waterfall projects. They are more apt to be delivered on

time and budget, to be easier to implement and adapt, and produce data that people find useful (The Standish Group 2012).
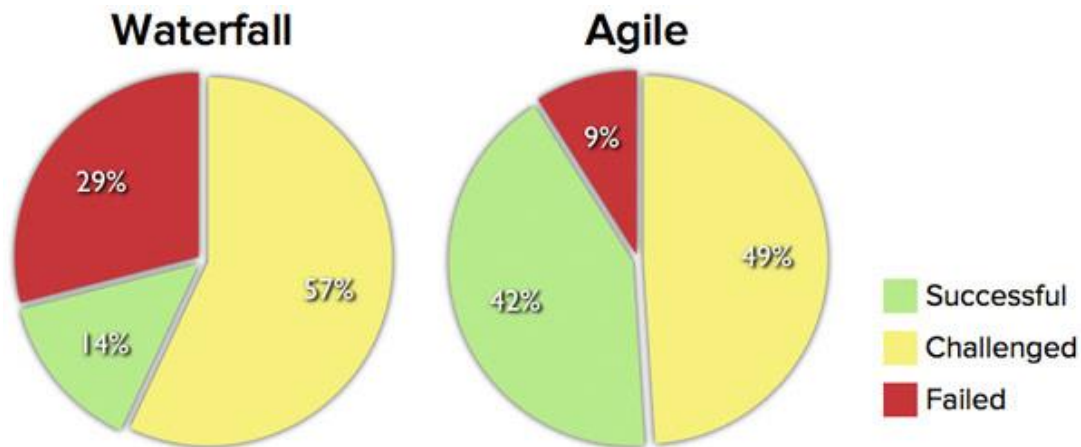
Source: The CHAOS Manifesto, The Standish Group, 2012.

In practice, CI practitioners often benefit from employing a blended approach to developing shared measurement processes. This includes first "zooming out" to complete a broad-as-possible discovery of the evaluation questions and measurement demands of CI participants; then "zooming in" to develop the first iterations of the system using an Agile approach; and periodically "zooming out" again to ensure that overall system matches the need for the CI effort. The success of 24/7 Diversion was in part due to its ability to build on a comprehensive requirement document that emerged from an earlier, ultimately discontinued, Waterfall project.

While the Agile approach appears to be superior, its widespread adoption is limited by two things. First, many in the CI field continue to promote the Waterfall approach. Even the thoughtful review of emerging shared measurement processes by the UK-based NPC group prescribed a development process based on Waterfall thinking (Ni Ogain, Svistak, De Las Casas 2013). Next, many CI participants continue to use a Waterfall approach to develop their CI initiatives, taking many months or years to develop their strategy and partnerships. This essentially forces data teams to employ the same plodding approach, regardless of its limitations.

If CI participants are going to realize the full benefits of shared measurement, we need to roll up our sleeves and change these comfortable, yet unproductive, habits.

A blended approach:

1. "Zooming out" to complete a broad-as-possible discovery of the evaluation questions and measurement demands of CI participants.
2. "Zooming in" to develop the first iterations of the system using an Agile approach.
3. Periodically "zooming out" again to ensure that overall system matches the need for the CI effort.

5. BEWARE OF PERVERSE CONSEQUENCES

Although advocates and practitioners of shared measurement are clear about the possible benefits of the practice, they often are unaware of the perverse consequences that may accompany it.

In his landmark study of the measurement experiment in the British public services in the mid-1990s, Ralph Smith identified types of perverse behaviour that appeared after such systems were introduced across diverse agencies and departments (Pidd 2005; Smith 1995).

Table 4: Perverse Behaviours in Response to Performance Measurement

Tunnel Vision	Organizations, faced with many different targets, choose the ones that are easiest to measure and/or offer rewards, then ignore the rest.
Sub-Optimizations	Organizations choose to operate in ways that serve their own operation well but damage the performance of the overall system.
Myopia	Organizations focus their efforts on short-term targets at the expense of longer-term objectives.
Measure Fixation	When outcomes are difficult to measure, there is a natural tendency to use measures based on measureable outputs, which replace the desired outcome as the organization's major focus.
Misrepresentation	Organizations misreport or distort performance measures to create a good impression.
Misinterpretation	Organizations use or analyze information in a way that is misleading and/or difficult to interpret.
Gaming	Organizations deliberately under-achieve in order to secure a lower target in the round of activity.
Ossification	Organizations cannot be bothered to revise or remove measures that are past their "sell-by" date and/or have lost their purpose.

It is possible to see multiple behaviours at the same time. For example, British health administrators seeking to reduce unacceptably long patient wait-times in hospitals were happy to report that their laser-sharp focus on this problem eventually led to measurable and significant improvements. However, it also resulted in doctors routinely ignoring other clinical priorities that were not measured (e.g., goal displacement); in frontline staff who refused to admit patients until they judged that these people could be seen within the agreed-upon performance target of four hours (e.g., sub-optimization); and administrators who manipulated

TAMARACK INSTITUTE

the data on waiting lists to show a rosier picture of their results than was the case (e.g., misrepresentation). A public inquiry into the spike in the number of "excess deaths" at a Birmingham hospital concluded that negative performance measure-induced behaviours were partly to blame (Smith 2013).

CI initiatives are not immune to these temptations. This is particularly true when participants of different initiatives attempt to compare their results. Benchmarking is notoriously difficult, even in hierarchical systems where very similar organizations share the same outcomes, programs and administrative environments. Educators, for example, warn how misleading it is to compare schools' performances through standardized tests because the schools operate in different neighbourhoods, with different socio-economic and demographic groups, and offer a different mix of services. "It's comparing apples and oranges," is the common refrain.

> " **CI participants need to (a) beware that perverse behaviours are likely to emerge, (b) vigilantly monitor their work so that these problems can be spotted early, and (c) having discovered them, take whatever remedial action seems appropriate.** "

The complexity of benchmarking is amplified in CI efforts which tend to involve organizations with different missions, target populations and activities. Take, for instance, the case of a group of agencies which organize summer recreation programs for a city's rapidly growing population of immigrant youth, many of them refugees struggling to find their footing in Canada. They agreed to work under the umbrella of the larger initiative in order to increase the rate of high school graduation across the city. But assisting youth to succeed academically is just one of their broader goals, which also include (a) building resilient youth (e.g., developing new relationships, facilitating the development of a healthy cross-cultural identity, increased leadership skills, etc.); and (b) increasing the capacity of emerging ethno-cultural organizations to manage projects, engage their communities and secure grant funds.

Simply asking these organizations to collect and report on measures that capture only one small part of their work was enough to threaten the group's impressive, albeit fragile, community development model:

> *We feel conflicted. We want to be a good partner in this initiative. We are also aware that funders are trying to align their fund to achieve these outcomes. That makes sense too. Yet we fear that if we reorganize our programs to focus only on academic results – say, for instance, turning our recreation programs into homework clubs – we'll only make a very modest contribution to school success, and we'll lose so much more in the process. It does not feel good.*[7]

---

[7] Personal communication in 2014.

It's nearly impossible to avoid situations like these because they are an inevitable part of any measurement exercise. CI participants simply need to:

- Be aware that they are likely to emerge
- Vigilantly monitor their work so that these problems can be spotted early
- Having discovered them, take whatever remedial action seems appropriate

If followed, these simple rules can be effective. In response to the growing concern over the widespread use of measurement in hospital care, the British government added measures to capture the quality of patient care, dropped some measures entirely, and limited how many measures get reported to the general public. Similarly, in the case of the CI effort to increase high school graduation, the initiative funders agreed that the network of summer program providers should be exempt from reporting on measures that are not fully relevant to their work and decided to find other sources of funding for it.

The ability of CI to get to the next level of practice and outcomes depends on the willingness and ability of advocates to acknowledge the shadow side of shared measurement and encourage practitioners to manage it better.

CONCLUSION

The organizations and residents in Lancaster County, Pennsylvania, are pioneers in Collective Impact. Like many of their peers across North America, they decided to end homelessness in their community. Eight years after organizing their formal coalition, they managed to end chronic homelessness. The 360 plus people who ended up on the streets this year were placed into stable housing within a relatively short period time. The community appears to be the first county in the United States to accomplish this. The group is now turning its attention to preventing people from ending up on the street in the first place.

Emboldened by the progress of their homelessness work, the United Way of Lancaster County embraced the CI approach. In 2015, they shifted all their resources to support 17 CI initiatives, each contributing to four "Bold Goals" in the area of early childhood development, education, poverty and medical care.

Evaluation and shared measurement form a cornerstone of their strategy. They track county-level data for the major indicators related to their Bold Goals (e.g., math and reading scores, monthly homeowner costs) and have developed a common intake form for service agencies, including measures to track clients' progress on a self-sufficiency scale. They complement these outcome measures with ongoing assessment of their CI process. These include network surveys to explore the changing working relationships amongst CI participants, and a collaborative self-assessment tool to assess the health of these relationships. They have even employed After-Action-Reviews following key meetings and events to encourage real-time learning. The group has one of the most coherent evaluation and measurement systems for an early stage group that I have ever seen.

They've also wrestled with the five challenges described above:

1. Ensuring that shared measures are organized in a way that reflects the group's evolving strategy or theory of change.
2. Distinguishing between shared outcomes, measures and measurements.
3. Creating good sense-making and decision-making processes.
4. Employing a mix of "big design" and "agile" approaches to the development of shared measurement systems.
5. Acknowledging, monitoring and responding to a variety of perverse behaviours that often emerge with measurement processes.

They are navigating these challenges well for a number of reasons. They are committed to tracking progress, both to inform their evolving strategy and to create public support for their new approach. Funders are willing to invest sufficient funding and resources to ensure the work is done well. And they have partnered with a very reputable research firm (Franklin & Marshall College, Center for Opinion Research). Their CI team has excellent technical expertise, a sensitivity to the complexities of community change work and a commitment to working in partnership with CI participants.

The most important factor of all, however, is that the group is prepared to relentlessly invent, develop and adapt their evaluation frameworks, measures and methods. The partnership leaders, backbone staff and evaluators are putting together a short list of "upgrades" to their evaluation tools and processes in preparation to guide the second 3-year cycle of the collective initiatives.

CI participants seem eager and ready for the next chapter of evaluation. As one participant at the county's most recent Collective Impact Summit noted:

> *It's a pain in the butt at times because it's messy, and takes way, way more time and effort than I thought. But it's worth it: thanks to these Collective Impact Partnerships, our community can do things now that we could not do before, and we can see the difference it makes for kids. So, if we say we are going to work together better, then we have to measure better together. We just have to figure out how to do it. Count me in for the long haul.[8]*

Spoken like a true Collective Impact pioneer.

---

[8] Personal communication with a CI participant at the Collective Impact Summit in Lancaster, Pennsylvania, September 13, 2017.

TAMARACK INSTITUTE

ABOUT MARK CABAJ

Mark is President of the consulting company From Here to There and an Associate of Tamarack. Mark has first-hand knowledge of using evaluation as a policy maker, philanthropist, and activist, and has played a big role in promoting the merging practice of developmental evaluation in Canada. Mark is currently focused on how diverse organizations and communities work together to tackle complex issues, on social innovation as a "sub-scene" of community change work, and on strategic learning and evaluation.

REFERENCES

Auspos, Pat, Cabaj, Mark. (2014). Complexity and Community Change: Managing Adaptively to Improve Effectiveness. Washington, DC: The Aspen Institute Roundtable on Community Change.

Cabaj, Mark. (2014). Evaluating Collective Impact: Five Simple Rules. The Philanthropist. Issue 1, Volume 26. Retrieved from: https://thephilanthropist.ca/2014/07/evaluating-collective-impact-five-simple-rules/

Cabaj, Mark, Weaver, Liz. (2016). Collective Impact 3.0: An Evolution in the Revolution. Waterloo, ON: The Tamarack Institute.

Cytron, Naomi, Petit, L.S. Kathryn, Kingsley, G. Thomas. Editors. (2014). What Counts: Harnessing Data for America's Communities. San Francisco, CA: Federal Reserve Bank of San Francisco & The Urban Institute. http://www.whatcountsforamerica.org/portfolio/shared-measurement-advancing-evaluation-of-community-development-outcomes/,

Fiester, Leila. (2014). Measuring Change While Changing Measures: Learning In, and From, The Evaluation of Making Connections. Baltimore, MA: The Annie E. Casey Foundation.

Gladwell, Malcolm. (2005). Blink: The Power of Thinking Without Thinking. New York, New York: Little, Brown.

Harvard Business Review. (2017). The Leaders Guide to Problem Solving. Fall.

Hassan, Zaid. (2014). The Social Labs Revolution: A New Approach to Solving our Most Complex Challenges. San Francisco, CA: Berrett-Koehler Publishers.

Hayne, Philip. Managing Complexity in the Public Services. (2009). Berkshire, England: Open University Press.

Kania, John & Kramer, Mark. (2011). Collective Impact. Stanford Social Innovation Review. Winter.

Van Vliet, Alex, Kail, Anglea, Sharon, Rachel. (2014). The Future of Shared Measurement: A Guide to Assessing Individual Sector Readiness. United Kingdom: NPC.

TAMARACK INSTITUTE

Ni Ogain, Eibhlin, Svistak, Marina and de Las Casas, Lucy. (2013). Blueprint for Shared Measurement: Developing, Designing and Implementing Shared Approaches to Impact Measurement. United Kingdom: NPC. Retrieved from: file:///C:/Users/Mark%20Cabaj/Downloads/Blueprint-for-shared-measurement1%20(1).pdf

McGinnis, Simon and Elhandy, Kasturi. (2012). Unintended Behaviour Consequences of Publishing Performance Data: Is More Always Better? The Journal of Community Informatics. Volume 8, No. 2.

Preskill, Hallie, Parkhurst, Marcie, Splansky-Juster, Jennifer. A Guide to Evaluating Collective Impact: Learning & Evaluation in the Collective Impact Context. Boston, MA: The Foundation Strategies Group.

Pidd, M. (2005). Perversity in Public Service Performance Measurement. International Journal of Productivity and Performance Management. Volume 54, No.5/6: 482-493.

Smith, Peter C. (2013). Strategic Targets for Public Services: Lessons for Canada from the English Experience. Tansley Lecture: Retrieved from: https://www.schoolofpublicpolicy.sk.ca/documents/research/archived-publications/tansley-publications/2013_Tansley%20Publication.pdf

Smith, Peter C. (1995). "On the Unintended Consequences of Publishing Performance Data in the Public Sector." In International Journal of Public Administration. Volume 18. No. 2/3: 277-310.

The Standish Group. (2012). The Chaos Report: The Year of the Executive Sponsor. Retrieved from: https://cs.calvin.edu/courses/cs/262/kvlinden/resources/CHAOSManifesto2012.pdf