Continuous Quality Improvement (CQI) Working Paper
Program Evaluation Basics
September, 2013

This document is one of eight working papers focusing on the components of a CQI system:

1. Leadership and Making the Business Case,
2. Managing Data to Support CQI,
3. Qualitative Case Review Processes,
4. Turning Data into Information,
5. Action Planning,
6. Program Evaluation Basics,
7. Building the Capacity of the CQI Workforce and
8. CQI Structure, Teams and Communication.

The purpose of these documents is to define and describe the range of specific strategies within each component to implement a high functioning CQI system in a jurisdiction. While recognizing that the evolution of CQI is iterative and requires time to refine and implement strategies, each working paper is intended to stimulate thinking about a range of strategies, to identify possible barriers to implementation and to identify solutions and recommendations.

Each document includes specific citations and suggestions for additional background publications, information and materials, but all the working papers derive critical background information from three key sources. First, the CQI Framework helped identify the 8 components and key strategies across the working papers.[i] Second, the Administration for Children and Families Information Memorandum on CQI helped inform many of the concepts in these papers.[ii] Third, and perhaps most important, the National Child Welfare Resource Center for Organizational Improvement (NRCOI), in collaboration with the National Resource Center for Child Welfare Data and Technology (NRC-CWDT), convened 75 representatives from 23 states and numerous organizations to attend a National CQI Working Meeting on August 29-30, 2012. Participants reviewed draft working papers and worked intensively to refine their content during and after the working meeting.[iii]

I.   Definition and Background

Program evaluation is a *systematic approach to gathering and using both quantitative and qualitative data to improve program processes and to test the effects of a program on the desired results*. Program evaluation asks the question: How do or will we know if **_it_** (the "program" as defined below) is working? Specifically, is there credible evidence that the program contributed to achieving the desired results?

NRCOI
National Child Welfare Resource Center for Organizational Improvement
A service of the Children's Bureau, a member of the T/TA Network

For the purposes of this working paper, the term "program" will be used to designate a range of new and existing interventions including for example practice changes, special initiatives, practice models, policy changes, action plans, changes in front line practice or a specific service model or program.
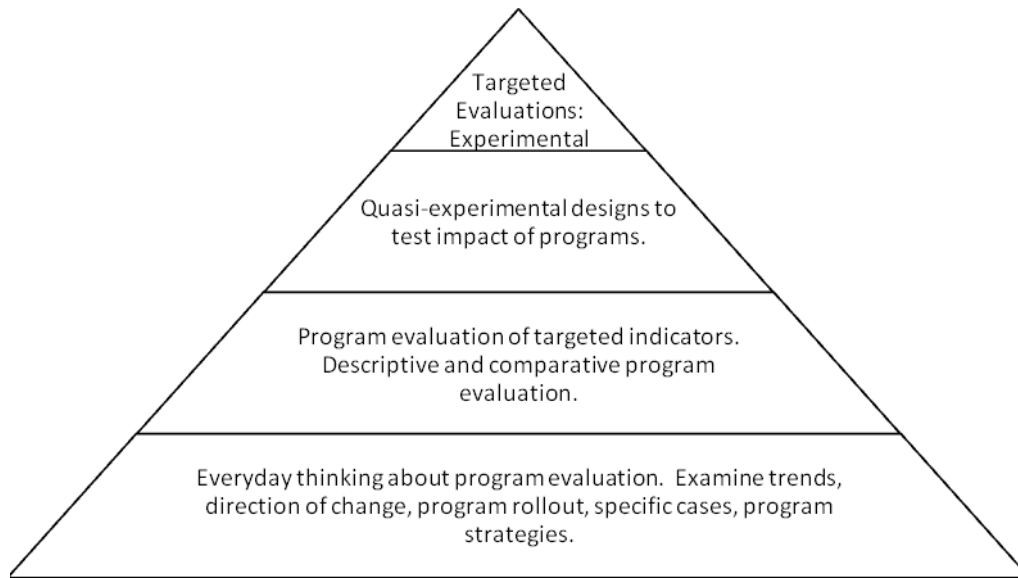
In the context of CQI, program evaluation is a way of thinking, collecting information and providing results on the program that can be used for decision making on questions that are important to the agency. Program evaluation tests program processes and effectiveness in typical conditions (real life practice), making it amendable to applied observational and quasi-experimental methods. CQI program evaluation can be applied to new or existing programs at multiple levels such as the team, local, tribal, county, regional or state level.

Program evaluation is an essential component of the CQI process that aligns with the Children's Bureau's five components of CQI (ACYF-CB-IM-12-07) as a *feedback* mechanism to stakeholders and decision makers for making adjustments to programs and processes. At the most fundamental level, program evaluation is a way of thinking that focuses on results and evidence rather than compliance and tradition. At the most complex level, program evaluation is a rich array of research designs, statistical methodology, and knowledge building strategies.

Program evaluation is a way of thinking about linkages of practices to results that undergirds all aspects of CQI. The other core components of CQI (e.g., leadership, data, analysis capacity, structure and well trained staff) compliment and undergird program evaluation. As these other aspects grow stronger they support more sophisticated program evaluation designs. CQI program evaluation adds three unique and important processes to the CQI system: the use of logic models, testing of program penetration/fidelity and testing of impact.

CQI program evaluation offers a robust system to test what is working and inform decisions, including cost decisions. Internal to the agency capacity, program evaluation matches the agency's needs, focuses on issues important to the agency and tests programs in typical practice. It moves the agency from a focus on compliance (what doing?) toward a focus on outcomes (why doing?). CQI processes can identify successes and highlight practices associated with success. It provides the agency with a system of accountability and transparency to consumers and partners and may provide answers to challenges. Program evaluation is at once fundamental to CQI but also a more advanced aspect of CQI. If in place at some level of sophistication, having the capacity to do program evaluation is a characteristic of a high functioning CQI program.

Program evaluation is scalable from an everyday way of thinking about trends to occasional experimental testing of program outcomes. The evidence for any program's effectiveness may be built from the bottom up through increasingly rigorous and sophisticated tests of impact (as illustrated in the diagram below). Or, programs with experimental evidence of effectiveness implemented into typical practice may be tested to confirm that the desired results are achieved when implemented in typical practice within specific jurisdictions. The bottom tiers of the diagram overlap with data analysis and managing data for CQI processes, but differ in the way of thinking that links programs with results. The top two tiers refer to the unique contributions of program evaluation to the CQI process.

Targeted
Evaluations:
Experimental

Quasi-experimental designs to
test impact of programs.

Program evaluation of targeted indicators.
Descriptive and comparative program
evaluation.

Everyday thinking about program evaluation. Examine trends,
direction of change, program rollout, specific cases, program
strategies.

II.   Key Strategies

Design programs and evaluations concurrently. Ideally, programs should be developed or initiated with the end or result in mind. That is, the goal or results are the first consideration and the program is then designed to improve the result or achieve the goal based on ideas on how change will occur. Program evaluation and program design should be developed in tandem.

Involve internal and external stakeholders. The basics of program evaluation can be mastered and practiced by everyone in the agency, from leadership through front line staff. To achieve improvement, everyone must understand the desired results, apply the program with skill and commitment, and then monitor how or if the results are achieved. Engaging the entire agency, consumers, and partners in the program evaluation process reinforces the notion that child welfare is a community effort.

Have an internal agency CQI person manage program evaluation functions and coordinate efforts with external evaluators. Program evaluation requires research design, analysis, and presentation skills. Investing skilled personnel rather than relying solely on external evaluators is necessary to spread the program evaluation skills and ways of thinking across the agency.

Build and adjust data systems to support both outcomes measurement and the linking of practices or processes to outcomes. This will build the capacity to link credible data on practices to inform program evaluation questions. For example, creating links between administrative and case review data opens up opportunities for expanded data analyses to examine the impact of specific practices on child and family outcomes.

Include cost related analysis or cost considerations in CQI program evaluation. How much does it cost to serve families, how many more families need the service, what are the cost offsets? Cost information is critical as jurisdictions decide whether and how to expand, support and modify programs based on program evaluation results.

NRCOI

National Child Welfare Resource Center for Organizational Improvement

A service of the Children's Bureau, a member of the T/TA Network

Program evaluation may consider costs of the program and determine whether cost avoidance or savings has occurred as a result of a program. This is an important aspect of program evaluation in the context of agency performance and decision making; the methodology and data to support this are evolving.

Program evaluation also may consider how many families, children, clients, workers etc. might be included if the program were to expand. There may be needs for targeted roll out of a program or cost considerations that are informed by program evaluation results.

Utilize logic models. A logic model is a statement of how programs are linked to results and a statement of the rationale for change. These models vary in formality and complexity, but understanding the logic of any program is essential to CQI.

A simple logic statement essential for action planning is: "If we (the agency) do _____ (program, practice, or policy) then we expect to achieve _____ (result or change) because _____ (rationale for change)". For example, if we implement CQI, then we expect better client results because CQI provides knowledge and engages everyone in designing solutions.

More formal and sophisticated logic models may be laid out in tables that illustrate the links between programs, activities, outputs and outcomes. The rationale for change is often included separately. For CQI as a program, a simple logic model might looks as follows:

Inputs	Activities/outputs	Proximal Results	Middle Results	Long Term Results
Time, effort, and resources dedicated to implementing core components of CQI.	Core components of CQI and related strategies are implemented with increasing fidelity.	Agency staff and community partners use data and participate in action planning.	Agency leadership and staff make proactive decisions that address long-term, adaptive and sustainable solutions.	The agency's results or outcomes improve in response to the solutions.

CQI is thought to improve agency results because it empowers staff to understand the people they serve and their practice, to self-evaluate, and make proactive decisions based on evidence and data. Agency work is linked to the results for families. The agency's practices and systems are being continually improved because of the feedback loop from front line through leadership.

Based on such a logic model and its rationale for change, program evaluation might explore the following results:
- Agency personnel will implement CQI strategies with increasing fidelity.
- Agency personnel will develop action plans and use data to monitor results and guide decisions.
- Agency personnel will focus on designing systemic solutions to problems or barriers that are proactive, adaptive, and sustainable and address long-term outcomes.
- The results achieved by the agency will improve, especially in the targeted areas.

Test program fidelity and consistency. To produce results, a program must first be implemented with potency. Potency refers to the fidelity of using program strategies as designed, the program's penetration or consistency of application, and the time needed to impact change.

A method is needed to monitor whether programs are implemented as planned (both practice frequency and quality of practice) and consistently across the organization. Sample methods might include:
- Calendar or other work logs or supervisor monitoring at the team or local level.
- Focus groups or presentations of work that demonstrate the practice.
- Case review processes.
- Administrative data reports that monitor fidelity with important practices.
- Special data collection strategies such as case reviews or confirming phone calls to customers might test if the strategies are implemented.
- Administrative data analysis that identify consistency and penetration across geographic or demographic groups.

CQI program evaluation processes need to include feedback on program fidelity and consistency with action planning to improve program implementation. Without adequate fidelity to the program model, program evaluation of results is usually premature. Nonetheless, differences in program implementation are opportunities to modify practices in response to questions such as:
- What are agency staff and community partners' perceptions of barriers to implementation?
- Are differences in program implementation based on geographic or demographic variables? Do differences exist between groups that do and do not receive the program? What underlies these differences in practice?
- If groups that receive or do not receive the program are reasonably similar, are results for the groups different?

Test program impact. Tests of program impact intend to demonstrate whether the program produced the desired results and to rule out other explanations of that change. Testing the impact of the program on results requires a comparison group that is scalable on a continuum of rigor.

All designs, including randomized control trials, have limitations and no design is perfect. It is important to design the comparison to best address the program qualities and agency needs and capacity.

Non-experimental or non-quasi-experimental designs may show results, but have weak evidence to support the program as the reason for those results. However, these designs still have utility for the agency and may guide thinking and build knowledge and action planning. While such designs may have no comparison or no baseline assessments of performance, they may produce informative results through:

- Pre-test to post-test changes on measures.
- Trends in the data before and after the program.

Quasi-experimental designs set up a comparison group that helps to rule out other explanations of program improvement. The more the intervention group and comparison group are matched at the baseline (before the program in implemented), the stronger the design. Stronger designs produce more credible evidence that the program, rather than other factors, made the difference. These include:

- Unmatched historical comparison (e.g., comparison now to former level of performance within another agency, in literature results, or results with a different group).
- Aggregate historical comparison group (e.g., comparison to past performance as an overall indicator).
- Matched historical comparison group (e.g., comparison to past performance with the same group or population).
- Same time comparison group not necessarily equivalent (e.g., comparing one county with the program to another without the program).
- Same time matched population comparison means results achieved for a nearly identical population at the same time (e.g., comparing rates in the program to an identical population that did not get the program).
- Same time matched comparison (e.g., comparing results for cases that are matched on demographics and time).

To qualify as an experimental design, there must be random assignment of persons to one condition (e.g., the program) or the other condition (e.g., no program or another alternative program).

Although overall tests of program impact are important in program evaluation, in the CQI context, program evaluation may examine subgroups within child welfare to describe whether the program works better for some groups than others. The program evaluation results may change over time due to other factors. Program evaluation in the CQI context might identify when changes to the program are needed to respond to new directives, evidence or other factors influencing the agency.

III.   Implementation Barriers

Deciding when to use external evaluators. Some may be opposed to having program evaluation capacity built into the child welfare agency, arguing that such evidence is too weak for action or that external evaluation is required for objectivity. Child welfare agencies might consider the following factors in deciding when to use external evaluators:

- Internal capacity. Quasi-experimental designs require a solid knowledge base in research methodology and statistical analysis. Conducting randomized control trials requires a higher level of expertise than is available in most agencies.
- Funding. Grant funds may require external evaluation or use of research designs beyond the internal capacity of the agency.
- Questions on dissemination of a program. Agencies may want experimental testing or more rigorous testing of particular program before rolling out a program to a wider audience or to have defensible data on a program's results.
- Partnership. An agency may want to strengthen a partnership with an external evaluator for a variety of political, financial, or collaborative reasons.
- Strength of the external partner. External partners vary in their desire to serve the agency's needs and respond to the agency's agenda.
- Mandates. Lawsuits or adversarial or complex situations may be best addressed through external evaluation procedures.
- CQI process results. CQI processes may be unable to improve agency results or achieve targets; external evaluation and consultation might provide a fresh or more sophisticated approach to analysis and action planning.

Transparency tension. There is tension about transparency and agencies' comfort with being transparent given external scrutiny they usually face. Some considerations for agencies include:
- Reliability of data. High data quality and reliable trends over time produce more understanding.
- Ability to produce information. Results indicators alone are less helpful than knowledge building around subgroups, demographics or agency practices that influence results.
- Partnerships. Partners who are engaged with the information may share accountability and responsibility with the public child welfare agencies.
- Confidentiality. Challenges due to concerns over confidentiality may or may not be warranted. Legal collaboration with agencies on what information is allowable, to whom, with what releases may help.
- Strength of support. Support from legislators, leadership, and other partners.

VI. Background Information and Materials

Chinman, M., Imm, P., & Wandersman, A. (2004). *Getting to outcomes 2004: Promoting Accountability through Methods and Tools for Planning, Implementation, and Evaluation.* Santa Monica CA: Rand Corporation. Downloaded 3/2/2012 from http://www.rand.org/content/dam/rand/pubs/technical_reports/2004/RAND_TR101.pdf

Fetterman, D.M. (2000). *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A synthesis of the literature.* Tampa, FL: University of South Florida, The Louis de la Parte Florida Mental Health Institute, Department of Child and Family Studies.

Guo, S., & Fraser, M.W. (2010). *Propensity score analysis: Statistical methods and applications*. Thousand Oaks, CA: Sage.

Hess, P. M., & Mullen, E.J. (Eds.) (1995). *Practitioner-researcher partnerships: Building knowledge from, in and for practice*. Washington, DC: NASW press.

NIATx. http://www.niatx.net/Home/Home.aspx?CategorySelected=HOME.

Preskill, H., & Beer, T. (2012). *Evaluating Social Innovation*. Available at: http://www.fsg.org/Portals/0/Uploads/Documents/PDF/Evaluating_Social_Innovation.pdf.

Schorr, L.B. (2003). *Determining "what works" in social programs and social policies: Toward a more inclusive knowledge base*. Washington, DC: Brookings Institution.

Testa, M.F., & Poertner, J. (2010). *Fostering Accountability Using Evidence to Guide and Improve Child Welfare Policy.* Oxford University Press.

U.S. Government Accountability Office. (2011). *Experienced Agencies Follow a Similar Model for Prioritizing Research.* GAO-11-176. Available from: http://www.gao.gov/products/GAO-11-176.

U.S. Government Accountability Office. (2009). *Program evaluation: A variety of rigorous methods can help identify effective interventions* (Report No. GAO-10-30). Washington, DC: Author. Retrieved from http://www.gao.gov/new.items/d1030.pdf.

Yates, B.T. (1996). Analyzing costs, procedures, processes, and outcomes in human services: An introduction. Thousand Oaks, CA: Sage.

---

[i] Using Continuous Quality Improvement to Improve Child Welfare Practice – A Framework for Implementation, Casey Family Programs and the National Child Welfare Resource Center for Organizational Improvement, May 2005. http://muskie.usm.maine.edu/helpkids/rcpdfs/CQIFramework.pdf

[ii] Information Memorandum: Continuous Quality Improvement in Title IV-B and IV-E Programs. ACYF-CB-IM-12-07. http://www.acf.hhs.gov/programs/cb/resource/im1207

iii This was an intense working meeting that detailed the current needs and successes among child welfare jurisdictions in the core implementation components of CQI. Prior to the meeting participants received and reviewed draft working papers developed by the NRCs on 8 CQI core components. Participants with shared expertise worked in groups during the meeting and focused on refining the working papers on the content and execution of CQI core components. Large group sessions focused on the links between these components and the technical assistance (TA) needs of jurisdictions. As key stakeholders in the process, participants defined their needs, shared their successes and struggles, and thought creatively to further refine a CQI framework to advance the work in child welfare. The NRCOI and NRC-CWDT thank them for these efforts.