The Integration of the Program Evaluation Standards into an Evaluation Toolkit for a Transformative Model of Care for Mental Health Service Delivery

M. Elizabeth Snow
Centre for Health Evaluation and Outcome Sciences (CHÉOS); School of Population & Public Health, University of British Columbia

Mai Berger
CHÉOS; School of Population & Public Health, University of British Columbia

Alexia Jaouich
Stepped Care Solutions

Mélanie Hood
Stepped Care Solutions

Amy Salmon
CHÉOS; School of Population & Public Health, University of British Columbia

Background: Stepped Care 2.0 (SC2.0) is a transformative model of mental health service delivery. This model was created by Stepped Care Solutions (SCS), a not-for-profit consultancy that collaborates with governments, public service organizations, and other institutions that wish to redesign their mental health and addictions systems of care. The SC2.0 model is based on 10 foundational principles and 9 core components that can be flexibly adapted to an organization's or community's needs. The model supports groups to reorganize and deliver mental health care in an evidence-informed, person-centric way. SCS partnered with evaluators from the Centre for Health Evaluation and Outcome Sciences (CHÉOS) to create a toolkit that provides evaluation guidance. The toolkit includes a theory of change, guidance on selecting evaluation questions and designs, and an evaluation matrix including suggested process and outcome metrics, all of which can be tailored to each unique implementation of the SC2.0 model. The objective of this resource is to support organizations and communities to conduct high-quality evaluations for the purpose of continuous improvement (a core component of the model of care) and to assess the model's impact.

Purpose: The purpose of this paper is to discuss the integration of the program evaluation standards (PES) into an evaluation toolkit for SC2.0.

Setting: In this paper, we describe the toolkit development, focusing on how the PES were embedded in the process and tools. We explore how the integration of the PES into the toolkit supports evaluators to enhance the quality of their evaluation planning, execution, and meta-evaluation.

Intervention: Not applicable

Research Design: Not applicable

Data Collection and Analysis: Not applicable

Findings: In this paper, we describe the toolkit development, focusing on how the PES were embedded in the process and tools. We explore how the integration of the PES into the toolkit supports evaluators to enhance the quality of their evaluation planning, execution, and meta-evaluation.

Keywords: program evaluation standards; evaluation; mental health.

Introduction

The program evaluation standards (PES) issued by the Joint Committee for Standards on Educational Evaluation (JCSEE) (Yarbrough et al., 2011) were integrated into an evaluation toolkit for the implementation of a transformative model of care for mental health service delivery, termed "Stepped Care 2.0" (herein SC2.0). The SC2.0 model was created by Stepped Care Solutions, a not-for-profit consultancy that collaborates with governments, public service organizations, and other institutions that wish to redesign their mental health and addictions systems of care to address the unique needs of their communities.

Stepped Care 2.0

Stepped Care 2.0© (SC2.0) is a transformative model for the delivery of mental health and addictions resources and services. It helps improve flexibility to meet the wide-ranging needs and preferences of the individuals it serves by recognizing that one service type will not work for everyone and by providing diverse options across a continuum of care. People will engage best with resources that meet their needs and preferences when they are ready to do so. The model aims to find the right balance of wellness promotion, illness prevention, different levels of treatment intensity, and management of risk. SC2.0 organizes care options around a nine-step framework where the intensity of services and investment (i.e., time, effort, cost) increases from Step 1 to Step 9:

- Step 1: Informational Self-Directed
- Step 2: Interactive Self-Directed
- Step 3: Peer Support
- Step 4: Workshops
- Step 5: Guided Self-Help
- Step 6: Intensive Group Programming
- Step 7: Flexible Intensive Individual Programming
- Step 8: Chronic Care & Specialist Consultation
- Step 9: Acute and Crisis Care

Stepped Care 2.0 is a flexible, open-access model that allows individuals to direct the level of care they receive based on their preference and readiness to engage. Care decision-making is collaborative, ensuring individuals are informed on what they can expect from each level of care and the investment required on their part to achieve desired outcomes. Decisions to step up or down are made by clients based in part on their readiness, according to their care preferences, and informed by routine measurement (Cornish, 2020). People can access services and supports at different step levels at the same time.

The 10 SC2.0 principles provide a foundation for the development and implementation of the SC2.0 model. These principles can open the door to new ways of working and collaborating, creating opportunities for addressing existing system challenges:

1. Social justice drives effective care system transformation and is an intervention in itself.
2. Multiple and diverse care options are required as one approach will not work for everyone.
3. All individuals and communities have strengths and capacity.
4. People engage with what they are ready to do; gold standard intervention is that which best fits the service user at any given time.
5. Professionals do not carry all the wisdom; people often know what is best for them.
6. Mental health literacy is required for people to make informed decisions.
7. An effective care system ensures people have access to care when and where it is needed.
8. The whole is greater than the sum of the parts: the strength of the system relies on multilevel collaboration.
9. Minimal interventions can produce powerful results.
10. There is no ideal solution; trial-and-error leads to growth and change. (Carey et al., 2021, p. 7)

The SC2.0 model has nine core components. Core components are essential features, or key ingredients, that directly contribute to the effectiveness of the model in producing positive outcomes (Carey et al., 2021). Core components can be implemented in many ways, as the ways the system is organized and what services are offered can vary from community to community. The core components are:

1. Key stakeholders are engaged throughout the co-design process.
2. Services populating the model align to a variety of step levels, reflect various intensities and include both formal and informal services.
3. Risk management is evidence-informed, distributed, and effectively addresses stigma inherent in the dominant risk paradigm.
4. Continuous service improvement is achieved through ongoing monitoring and improvement cycles.
5. Recovery-oriented practice is demonstrated clearly and consistently.

6. Clients have same day access to multiple levels of care.
7. Treatment is guided by single-session thinking.
8. Treatment planning is flexible and responsible by strategically reviewing client data and making data-informed adjustments as needed.
9. Treatment planning is client-centric. (Carey et al., 2021, p. 13)

Stepped Care Solutions

Stepped Care Solutions (SCS) is a Canadian not-for-profit consultancy group that aims to promote mental wellness by reimagining mental health, substance use, and addictions systems. SCS believes that communities are valuable informants on what is needed and what will work in their own context. SCS's implementation guide provides practice standards to help organizations, communities, and regions operationalize each of the core components in their own contexts (Carey et al., 2021), and consultants from SCS provide training and consultation to support the implementation of the SC2.0 model.

Different organizations and communities have varying levels of in-house evaluation resources and expertise. Noting this, SCS sought to provide additional evaluation support to the organizations and communities with whom they work. In spring 2021, SCS partnered with a team of evaluators from the Centre for Health Evaluation of Outcome Sciences (CHÉOS) to develop an evaluation toolkit that could be provided to organizations and communities implementing SC2.0 to support high-quality evaluations for the purpose of continuous improvement and to demonstrate impact.

The Development of the Evaluation Toolkit

Why Integrate the Program Evaluation Standards?

Given that the organizations and communities who might use the evaluation toolkit would have varying levels of evaluation capacity, there was a particular interest in providing resources and guidance to facilitate evaluation planning and implementation consistent with the model. To this end, the PES were used to guide the development of the evaluation toolkit. As noted in the preface to the third edition of the PES, "The standards identify and define evaluation quality and guide evaluators and evaluation users in the pursuit of evaluation quality" (Yarbrough et al., 2011, p. xxii). Although they are referred to as "program" evaluation standards, the definition of "program" used in the third edition of the PES is expansive enough to include a systems transformation such as SC2.0.

Process of Developing the Toolkit

The CHÉOS evaluation team proposed a set of tools to include in the evaluation toolkit based on their collective years of experience conducting evaluations across a broad range of programs, projects, and systems transformations. The intent was to build tools tailored to the SC2.0 model, with the potential to have some standardization across evaluations. Given that the SC2.0 model can be flexibly adapted to the needs and context of a given community or organization, the evaluation toolkit was also designed to be adaptable. Table 1 shows the contents of the toolkit.

Table 1. Contents of the SC2.0 Evaluation Toolkit

Toolkit component	Purpose
Evaluability checklist	To determine if the organization/community is ready for evaluation both in principle and in practice. If it is not, then the evaluability assessment will make recommendations to prepare for evaluation (Davies, 2013; Leviton et al., 1998).
The SC2.0 theory of change	To represent the theory of how implementing the SC2.0 model is expected to result in its intended outcomes. The theory of change includes a diagram (presented at three levels of increasing detail for different audiences) and an accompanying narrative document. The

Toolkit component	Purpose
	theory of change is designed to be adapted by an organization/community to fit their specific implementation of the model and can be used to help inform the evaluation of the implementation.
Evaluation plan templates	To create an evaluation plan that is adapted to the organization/community. It is used in conjunction with the evaluation guide.
Evaluation guide	To provide guidance on developing all aspects of the evaluation. The guide accompanies the evaluation plan templates, with detailed guidance to help organizations/communities build their evaluation plans.
Evaluation matrix	To provide suggestions for indicators that organizations/communities may want to include in their evaluation based on the inputs, activities, outcomes, impacts, assumptions, and mechanisms in the theory of change. This includes indicator definitions, potential data sources, data collection methods, and references to indicator sources.
Resource page	All the resources shown in the other tools are compiled in one place for easy access.

The first product developed for the evaluation toolkit was the SC2.0 theory of change (ToC). The intention was to construct a ToC that would present the key features of the SC2.0 model, and that could be tailored by communities or organizations to represent the theory supporting their specific implementation of SC2.0. The ToC was collaboratively constructed through a series of workshops, facilitated by the CHÉOS evaluation team, that engaged SCS staff and representatives from their partner communities and organizations (Dhillon & Vaca, 2018; Mason & Barnes, 2007). The ToC went through several iterations as it was strengthened through a comprehensive comparison with the SC2.0 implementation guide (Carey et al., 2021) and a targeted literature review to determine the level of empirical support for the mechanisms and assumptions. Diagrammatic representations of the ToC were created at three different levels for different audiences: a detailed (evaluation and research) version, a mid-level (planning and Implementation) version, and a high-level (communications) version. A narrative document was also created to accompany the three versions. A draft of the ToC was shared with a variety of collaborators to verify that it resonated with them. In addition, a graphic designer was engaged to assist with the visual communication of the ToC.

Next, the evaluation team created initial drafts of the remaining tools. These documents were drafted based on the experience of the evaluation team members in conducting evaluations and targeted searches of evaluation literature, combined with the understanding of the SC2.0 model that they developed through the process of co-constructing the SC2.0 ToC with SCS and their stakeholders. As the drafts began to take shape, the evaluation team reviewed each chapter of *The Program Evaluation Standards* (Yarbrough et al., 2011) to identify where and how each standard was included in the initial drafts of the tools and where the tools could be further developed to include the PES. In particular, the "Implementing" and "Hazards" sections of each chapter, as well as the scenarios, were particularly useful in that they provided concrete examples of how to apply the PES in evaluations and served as inspirations for how the evaluation tools could be further improved. Additionally, since the PES were published more than a decade ago, reviewing them prompted some further targeted literature searches to look for any recent advances in the literature.

Integrating the PES in the Toolkit

Through reviewing the PES, the team identified that some standards had already been included substantially and explicitly in the toolkit, and others had not. Among those included in the initial drafts of the toolkit were standards that naturally aligned with SC2.0's model of care and/or the evaluation team's common evaluation practices. For example, co-design with stakeholders is a core component of SC2.0, and the evaluation team's common practice is to use collaborative or participatory approaches to evaluation (Fetterman et al., 2014). So Standard U2 (Attention to Stakeholders) has been included explicitly in several places in the evaluation toolkit since its initial drafts:

- The introduction to the evaluation guide includes advice on forming a broad-based evaluation advisory committee to make decisions on all aspects of planning and executing the evaluation. In addition, in the evaluation methodology section of the guide, collaborative and participatory approaches are suggested among the recommended evaluation approaches.
- The SC2.0 ToC was co-constructed with stakeholders, and advice is provided in the ToC narrative document to tailor it to specific implementations of SC2.0 using a participatory approach.
- The evaluability assessment checklist includes questions to assess an organization's or community's readiness to conduct a participatory or collaborative evaluation.

Moreover, SC2.0's implementation guide includes a stakeholder framework that includes different groups: (Group A: leadership/administration, Group B: clinical staff, Group C: clients/families/peer support, Group D: internal stakeholders, and Group E: community stakeholders, including informal mental health supports [e.g., outside of mental health professionals]) (Carey et al., 2021). This SC2.0 stakeholder framework was used to create a stakeholder engagement grid that can be used to conduct a stakeholder analysis while developing an evaluation plan. An example of a completed stakeholder grid is included in the evaluation guide (see Table 3), and the evaluation plan template includes a blank version of the grid that can be used in an evaluation plan.

Furthermore, as the SC2.0 model has an explicit focus on the meaningful inclusion of people with lived and living experience of marginalization, considerable attention is paid throughout the evaluation toolkit to the meaningful inclusion of "the full range of individuals and groups invested in the program and affected by its evaluation" (Yarbrough et al., 2011, p. 3).

Table 2. Sample Stakeholder Engagement Grid from the Evaluation Guide

Group	Why are they important?	Individuals	How they can contribute to the evaluation	What they need from the evaluation	Actions to take	When	Who is responsible for taking action?
Leadership/ administration	▪ Have the ability to affect changes in organizational policies and processes	Person A's name	Participation in evaluation advisory committee (EAC)	Evaluation findings to answer their evaluation questions	Recruit to EAC	Before evaluation kickoff	Director
		Person B's name	Provide organizational approval for data release	Assurance that data privacy and security policies will be followed	Submit request for data	After data collection plan is complete	Evaluator
		Person C's name	Provide approval of policy and process changes	Recommendations for improvements based on evaluation findings	Prepare briefing note	After evaluation report is complete	Evaluator
Clinical staff	▪ Have direct contact with clients ▪ Are part of the delivery of SC 2.0	Person D's name	Participation in evaluation advisory committee	Evaluation findings to answer their evaluation questions	Recruit to EAC	Before evaluation kickoff	Director
		List of type of roles to be surveyed/ interviewed	Participation in surveys or interviews	An opportunity to share their experience and contribution to service improvements	Recruit for surveys or interviews	After data collection plan is complete	Evaluator
Clients, families, and peer support	▪ Have lived experience ▪ Have valuable expertise to contribute ▪ Are the ones who SC 2.0 is meant to serve	Person E's name Person F's name	Participation in evaluation advisory committee	Evaluation findings to answer their evaluation questions	Recruit to EAC	Before evaluation kickoff	Director
		List of type of roles to be surveyed/ interviewed	Participation in surveys or interviews	An opportunity to share their experience and contribution to service improvements	Recruit for surveys or interviews	After data collection plan is complete	Evaluator

Group	Why are they important?	Individuals	How they can contribute to the evaluation	What they need from the evaluation	Actions to take	When	Who is responsible for taking action?
Internal stakeholders	▪ Are part of the delivery of SC 2.0 ▪ Have close knowledge of services and process	Person G's name	Participation in evaluation advisory committee	Evaluation findings to answer their evaluation questions	Recruit to EAC	Before evaluation kickoff	Director
		List of type of roles to be surveyed/ interviewed	Participation in surveys or interviews	An opportunity to share their experience and contribution to service improvements	Recruit for surveys or interviews	After data collection plan is complete	Evaluator
Community stakeholders	▪ Are part of the delivery of SC 2.0	Person H's name Person I's name	Participation in evaluation advisory committee	Evaluation findings to answer their evaluation questions	Recruit to EAC	Before evaluation kickoff	Director
		List of type of roles to be surveyed/ interviewed	Participation in surveys or interviews	An opportunity to share their experience and contribution to service improvements	Recruit for surveys or interviews	After data collection plan is complete	Evaluator

Standards P1 (Responsive and Inclusive Orientation) and P3 (Human Rights and Respect) were also included in the initial draft of the evaluation toolkit (in several sections of the evaluation planning guide, as well as in the evaluability checklist). These were included in alignment with SC2.0's principle, "Social justice drives effective care system transformation and is an intervention in itself" (Carey et al., 2021, p. 7) and the model's person-centric approach. The Utility standards U5 (Relevant Information), U6 (Meaningful Processes and Products), and U7 (Timely and Appropriate Communicating and Reporting) all align with SCS's Core Component 4 ("Continuous service improvement is achieved through ongoing monitoring and improvement cycles" (Carey et al., 2021, p. 13)) and were all included in the initial draft of the evaluation planning guide.

Standards A2 (Valid Information), A3 (Reliable Information), A4 (Explicit Program and Context Descriptions), A5 (Information Management), and A6 (Sound Designs and Analyses) were all also included in the initial draft of the evaluation toolkit. As Gullickson et al. (2019) note, evaluators tend to focus a great deal on methods, which is in alignment with our having included these PES in the evaluation tools even before reviewing the PES. Similarly, F1 (Project Management) was built into the evaluation toolkit, in both the evaluability assessment and the evaluation planning guide. Since these tools are focused on determining the feasibility of evaluating a given evaluand and on creating a plan to conduct an evaluation, respectively, it is reasonable that they would naturally include effective project management strategies.

While many of the PES were explicitly included in the initial drafts of the tools, others had not been given sufficient attention. For example, even though the team knew that the tools were being designed for individuals and groups who had varying levels of evaluation expertise and who would be undertaking different evaluation activities, Standard U1 (Evaluator Credibility) was not initially addressed in any of the tools. Substantial efforts were made to create a robust toolkit with ample guidance on all aspects of evaluation, but discussion about what knowledge, skills, and dispositions are required to conduct high-quality evaluation work had not been included. Prompted by the review of Standard U1, a discussion of both *Competencies for Canadian Evaluation Practice* (Canadian Evaluation Society, 2018) and *AEA Evaluator Competencies* (American Evaluation Association, 2018) was added to the introductory section of the evaluation planning guide. In keeping with P1 (Responsive and Inclusive Orientation), it was also noted that one need not be professionally identified as an "evaluator" in order to conduct high-quality evaluation work, but having or building the knowledge and skills to do so—including understanding and being able to apply the PES—is important.

Standard U4 (Explicit Values) was another PES that, despite being critical to evaluation work, was not one of the primary areas of focus in the initial drafts of the tools, beyond recommendations for those engaged in evaluation work to reflect on their own values and how those values come into play in evaluation. Gullickson and Hannum (2019) indicate that this lack of explicit attention to values is common in evaluation and suggest some ways of making values explicit throughout the evaluation process. Thus, the following content related to the to Standard U4 (Explicit Values) was added to the Evaluation Guide:
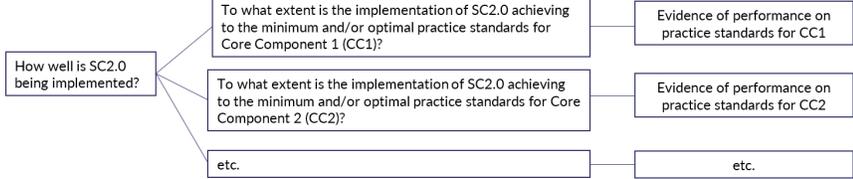
Excerpt 1.

Values are a critical part of the evaluation process, but they are often not discussed and their role in evaluation is often not made explicit (Gullickson & Hannum, 2019). Making decisions about what to evaluate, what types of evidence are deemed legitimate, what criteria will be used to evaluate, and what is deemed "good" all require value judgments. Reflecting on values—our own values, what others value, organizational values, the principles of SC 2.0—allows us to be more explicit and transparent in how decisions about the evaluation are made and will allow us to better articulate the evaluation reasoning we use to draw conclusions from the data we collect.

Similarly, A7 (Explicit Evaluation Reasoning) was not included in the original draft, but reviewing the PES prompted the evaluation team to add the following section on this topic to the Evaluation Guide: (see Table 5).

Excerpt 2.

Once you have your key evaluation question(s), it is useful to think about how you are going to go about answering them in an explicitly evaluative way. Evaluation questions "must ask not only 'What were the results?' (a descriptive question) but also 'How good were the results?' This cannot be done without using evaluative reasoning to evaluate the evidence relative to the definitions of quality and value" (Davidson, 2014, p. 4). Evaluative reasoning involves thinking through how you will decide what criteria you will use to draw conclusions and what the standards are to decide if the performance on that standard is "good." [This table] shows a process by which you can make your evaluative reasoning explicit.

Evaluation Type	Evaluative Reasoning
Process Evaluation	In a process evaluation you could ask the question "how well is SC2.0 being implemented?" and the definition of "quality" of implementation could be the practice standards. Thus, an implementation can be compared to the SC2.0 practice standards to determine the answer to the question "how well is it being implemented?"
Outcome Evaluation	In an outcome evaluation, you need to determine what your definition of "quality" or "good" is so that when you collect evidence of performance on an activity or outcome, you will be able to determine if it is "good." The diagram below can be used to guide evaluation reasoning for an outcome evaluation.
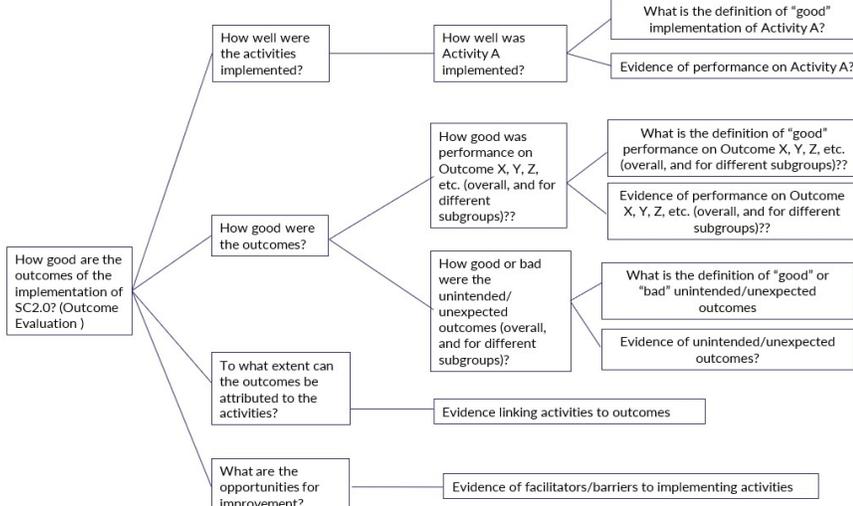
P4 (Clarity and Fairness), like P1 (Responsive and Inclusive Orientation) and P3 (Human Rights and Respect), is in alignment with SC2.0's principle related to social justice and with the model's person-centric approach (Carey et al., 2021), but the evaluation team had not made it explicit within the tools. Reflecting on ways of making values explicit throughout evaluation planning and execution (U4, Explicit Values, and P5, Transparency and Disclosure), as well as the

"Implementing P4" section of *The Program Evaluation Standards*, helped to identify areas where the PES could be integrated into the toolkit

Excerpt 3.

For example of the following text related to Standard P4 (Clarity and Fairness) was added to the Evaluation Guide:

Worldviews are important for evaluation because they shape how we conceptualize problems, devise solutions, and evaluate things.

> "Everyone has [a worldview], and it constitutes a set of beliefs that guides their judgment making and action taking in all spheres of activity. Different people have different worldviews, and individual worldviews change and develop on an ongoing basis." (Rousseau & Billingham, 2018)

Our worldview guides how we do things, regardless of whether or not we are aware of it. Often evaluations default to a Euro-centric worldview,[1] which is not universal. For example, assuming that things that can be measured "objectively" are more legitimate forms of evidence than lived experience, or assuming that a randomized control trial is the "best" design for any evaluation, are beliefs that can come from a Euro-centric worldview.

Taking time to explore one's own worldview and the worldviews of others can help us to uncover our own biases, create more welcoming spaces for evaluation, be more explicit about how we design our evaluations, and create more equitable and culturally appropriate evaluations.

Reflective practice, which is a competency for evaluation practice and is promoted in SC2.0 Core Competency 5, is a way in which those participating in the evaluation can discuss their and each other's worldviews and the implications of their worldviews for the evaluation.

It is also important to recognize that evaluation and research have been used in harmful ways in historically oppressed communities. This includes extracting knowledge from communities and using culturally inappropriate methods to conduct evaluations that serve to maintain the status quo rather than in service of social justice. It can take time and effort to earn the trust of people and communities. To avoid perpetuating harm, it is important to be intentional about embracing Indigenous and alternative ways of knowing when implementing and evaluating SC2.0.

——

[1] The Encyclopedia of Case Study Research (Mills et al., 2010) defines eurocentrism as "*a body of knowledge that has been used to interpret the histories and cultures of non-European societies according to the European experience. Embedded in present-day thought, values, and education, Eurocentrism is naturalized as common sense that underscores non-Western values and ways of knowing.*"

Finally, a metaevaluation section was not originally included in the evaluation plan template and evaluation guide, but reviewing the PES prompted the team to add this. While aspects of E1 (Evaluation Documentation) were already included in the evaluability assessment and evaluation plan template and guide, after reviewing the E1 chapter of *The Program Evaluation Standards*, the team incorporated Standards E2 and E3 into a dedicated section on metaevaluation in the evaluation plan template and guide. This section includes reference to a metaevaluation checklist based on the program evaluation standards (Stufflebeam, 1999), as well as reflecting on ways in which the SC2.0 evaluation toolkit could be improved.

Table 3 shows a summary of where each PES is now included in the SC2.0 evaluation toolkit. Notably, every PES appears in the evaluation plan

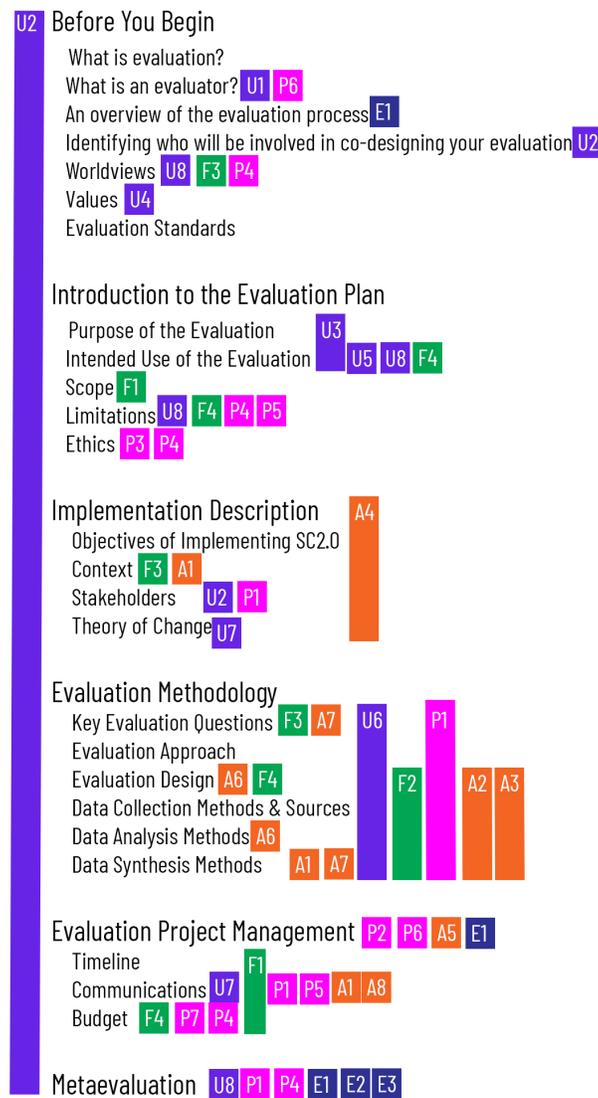template and in the guide, which was developed to facilitate the development of a comprehensive evaluation plan and thus covers all aspects of an evaluation. Figure 1 shows where each PES is included in the evaluation plan template and guide.

Table 3. Summary of Where Each PES Is Included in the SC2.0 Evaluation Toolkit

Program evaluation standard	Component of SC2.0 evaluation toolkit			
	Evaluability assessment checklist	Theory of change	Evaluation plan template & guide	Evaluation matrix
U1 Evaluator Credibility	■		■	
U2 Attention to Stakeholders	■		■	
U3 Negotiated Purposes	■		■	
U4 Explicit Values	■		■	
U5 Relevant Information	■		■	■
U6 Meaningful Processes and Products			■	
U7 Timely and Appropriate Communicating and Reporting		■	■	
U8 Concern for Consequences and Influence	■		■	
F1 Project Management	■		■	
F2 Practical Procedures	■	■	■	■
F3 Contextual Viability	■		■	
F4 Resource Use	■	■	■	■
P1 Responsive and Inclusive Orientation	■		■	
P2 Formal Agreements			■	
P3 Human Rights and Respect	■		■	
P4 Clarity and Fairness			■	
P5 Transparency and Disclosure	■		■	
P6 Conflicts of Interests			■	
P7 Fiscal Responsibility	■		■	
A1 Justified Conclusions and Decisions			■	
A2 Valid Information	■		■	■
A3 Reliable Information	■		■	■
A4 Explicit Program and Context Descriptions	■	■	■	
A5 Information Management	■		■	
A6 Sound Designs and Analyses		■	■	
A7 Explicit Evaluation Reasoning			■	

Program evaluation standard	Component of SC2.0 evaluation toolkit			
	Evaluability assessment checklist	Theory of change	Evaluation plan template & guide	Evaluation matrix
A8 Communication and Reporting			▮ (orange)	
E1 Evaluation Documentation	▮ (navy)		▮ (navy)	
E2 Internal Metaevaluation			▮ (navy)	
E3 External Metaevaluation			▮ (navy)	

Figure 1. Map of Evaluation Standards Included in the Evaluation Planning Guide

[U2] **Before You Begin**
  What is evaluation?
  What is an evaluator? [U1] [P6]
  An overview of the evaluation process [E1]
  Identifying who will be involved in co-designing your evaluation [U2]
  Worldviews [U8] [F3] [P4]
  Values [U4]
  Evaluation Standards

Introduction to the Evaluation Plan
  Purpose of the Evaluation [U3]
  Intended Use of the Evaluation [U5] [U8] [F4]
  Scope [F1]
  Limitations [U8] [F4] [P4] [P5]
  Ethics [P3] [P4]

Implementation Description [A4]
  Objectives of Implementing SC2.0
  Context [F3] [A1]
  Stakeholders [U2] [P1]
  Theory of Change [U7]

Evaluation Methodology
  Key Evaluation Questions [F3] [A7] [U6] [P1]
  Evaluation Approach
  Evaluation Design [A6] [F4] [F2] [A2] [A3]
  Data Collection Methods & Sources
  Data Analysis Methods [A6]
  Data Synthesis Methods [A1] [A7]

Evaluation Project Management [P2] [P6] [A5] [E1]
  Timeline [F1]
  Communications [U7] [P1] [P5] [A1] [A8]
  Budget [F4] [P7] [P4]

Metaevaluation [U8] [P1] [P4] [E1] [E2] [E3]

Legend: Purple = Utility standards; green = Feasibility standards; pink = Propriety standards; orange = Accuracy standards; blue = Evaluation Accountability standards.

The second drafts of the tools were provided to SCS and were further refined based on SCS's feedback. The final drafts will be provided to organizations and communities working with SCS to guide their evaluation work, with support from the SCS consultants. The toolkit is currently being used for an internal SCS project; this and future projects will provide opportunities to gather feedback for continuous improvement of the toolkit.

Limitations and Areas for Future Research

One of the challenges of developing this evaluation toolkit was writing for a wide range of potential users, from those who have little experience with evaluation to evaluation experts. Using the PES as the evaluation team drafted the toolkit helped ensure key aspects of evaluation were not overlooked.

Another challenge was to keep the evaluation guide to a manageable size yet still provide guidance on conducting a high-quality evaluation from start to finish. The evaluation team kept the size of the toolkit manageable by writing concisely and providing links to high-quality external resources.

Evaluating the value of the toolkit, and the utility of the PES in particular, to end users is an area for future study. Stepped Care Solutions is currently using the evaluation toolkit with one of their projects and we, the evaluation team, have embedded a metaevaluation component to allow for an exploration of the value of the toolkit, including how well it works for different types of implementations of SC2.0, and areas where the toolkit can be further improved.

Conclusion

The quality of the evaluation toolkit was strengthened by intentionally integrating the PES. The toolkit will support actionable evaluations of the SC2.0 model and promote continuous improvement to optimize implementation. As experienced evaluators, it was useful for the evaluation team to reflect on the PES to create tools for evaluators (some of whom might be new to evaluation, or who might be researchers branching into evaluation) to plan and execute high-quality evaluations.

While the clear discussion and concrete examples *in The Program Evaluation Standards* (Yarbrough et al., 2011) helped identify ways that the PES can be applied in evaluation, integrating

the PES into the toolkit was not without its challenges. As noted in that guide:

> standards require adaptive, responsive, and mindful use. They should not be applied literally and superficially following a simple recipe [...] Much attention has gone into the wording of the standard statements, but they are only compact advance organizers for the concepts and issues they represent. Those who apply the standard statements while ignoring clarifications, rationales, suggestions, and applications are missing out on the richness that makes the standards useful. (p. xxxii)

The team included the above statement in the toolkit's evaluation guide, along with the suggestion for evaluators using the toolkit to explore the PES in full. It is hoped that the toolkit will make novice evaluators aware of the PES and their potential to enhance evaluation quality, and remind more seasoned evaluators to revisit them.

References

American Evaluation Association. (2018). *The 2018 AEA evaluator competencies.* https://www.eval.org/Portals/0/Docs/AEA%20Evaluator%20Competencies.pdf

Canadian Evaluation Society. (2018). *Competencies for Canadian evaluation practice.* https://evaluationcanada.ca/txt/2_competencies_cdn_evaluation_practice_2018.pdf

Carey, S., Jaouich, A., Churchill, A., Cornish, P., Impey, D., & Kim, M. (2021). *Stepped Care 2.0 revised implementation guide.* Mental Health Commission of Canada.

Cornish, P. (2020). *Stepped Care 2.0: A paradigm shift in mental health.* Springer.

Davidson, E. J. (2014). *Evaluative reasoning (Methodological Briefs Impact Evaluation No. 4).* UNICEF Office of Research–Innocenti. https://www.betterevaluation.org/en/resources/guide/evaluative_reasoning

Davies, R. (2013). *Planning evaluability assessments: A synthesis of the literature with recommendations.* (Working Paper 40). UK Department for International Development. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf

Dhillon, L., & Vaca, S. (2018). Refining theories of change. *Journal of MultiDisciplinary Evaluation, 14*(30), 64–87.

Fetterman, D., Rodríguez-Campos, L., Wandersman, A., & O'Sullivan, R. G. (2014). Collaborative, participatory, and

empowerment evaluation: Building a strong conceptual foundation for stakeholder involvement approaches to evaluation (A response to Cousins, Whitmore, and Shulha, 2013). *American Journal of Evaluation*, *35*(1), 144–148. https://doi.org/10.1177/1098214013509875

Gullickson, A. M., & Hannum, K. M. (2019). Making values explicit in evaluation practice. *Evaluation Journal of Australasia*, *19*(4), 162–178. https://doi.org/10.1177/1035719X19893892

Gullickson, A. M., King, J. A., LaVelle, J. M., & Clinton, J. M. (2019). The current state of evaluator education: A situation analysis and call to action. *Evaluation and Program Planning*, *75*, 20–30. https://doi.org/10.1016/j.evalprogplan.2019.02.012

Leviton, L. C., Collins, C. B., Laird, B. L., & Kratt, P. P. (1998). Teaching Evaluation Using Evaluability Assessment. *Evaluation*, *4*(4), 389–409. https://doi.org/10.1177/13563899822208699

Mason, P., & Barnes, M. (2007). Constructing theories of change: Methods and sources. *Evaluation*, *13*(2), 151–170. https://doi.org/10.1177/1356389007075221

Mills, A. J., Durepos, G., & Wiebe, E. (2010). Eurocentrism. In *Encyclopedia of Case Study Research*. Sage Publications. https://methods.sagepub.com/reference/encyc-of-case-study-research/n132.xml

Rousseau, D., & Billingham, J. (2018). A systematic framework for exploring worldviews and its generalization as a multi-purpose inquiry framework. *Systems*, *6*(3), 27. https://doi.org/10.3390/systems6030027

Stufflebeam, D. L. (1999). *Program Evaluations Metaevaluation Checklist (Based on* The Program Evalaution Standards*)*. https://wmich.edu/sites/default/files/attachments/u350/2014/program_metaeval_short.pdf

Yarbrough, D., Shulha, L., Hopson, R., & Caruthers, F. (2011). *The program evaluation standards: A guide for evaluators and evaluation users*. Sage.