Rubrics Methodology in Detail: Helping Save The Children Turn Children's Experiences of Discrimination and Exclusion into Rich, Trackable Outcomes

E. Jane Davidson
Real Evaluation

Thomaz K. Chianca
Real Evaluation

Nicole Dulieu
Save the Children International

Anil Sigdel
Save the Children International

Background: Save the Children's child sponsorship program (referred to as Sponsorship) supports children, their families and communities, and institutions in 20 countries to create social and behavior change together so that all children survive, learn, are protected, and live free from discrimination.

Purpose: To present a new rubrics-enhanced monitoring, evaluation, accountability, and learning (MEAL) system being developed with Save the Children's Sponsorship Program. The system helps the organization track progress on key social change outcomes, especially less tangible outcomes, such as children's experiences of inequality and discrimination.

Setting: Sponsorship's work spans 20 countries around the world. It works through a 10-year program cycle, building community capacity to create social change that addresses the root causes of inequality and discrimination for a group of children living in an impact area. The evaluation methodology showcased here was piloted in Zambia, the Philippines, and El Salvador.

Intervention: Not applicable.

Research Design: This paper presents some of the nuts and bolts of using rubrics to turn stories of change into rich, trackable outcomes, using one of the 12 rubrics developed for Sponsorship to illustrate the methodology.

Data Collection and Analysis: Focus group and interview protocols are used to gather stories and other documentation (such as health statistics and educational outcomes) from children, families, community members, educators, health care professionals, and local government officials. A detailed, step-by-step evidence interpretation guide shows how to synthesize these multiple sources of evidence and convert them into evaluative ratings using rubrics and guided evaluative reasoning.

Findings: Rubrics are a powerful tool for evaluating complex and nuanced outcomes, provided they are appropriately adapted for different contexts. For MEAL practitioners less familiar with rubrics methodology, detailed guidance and support are needed to help ensure that the framework is applied consistently across the system while also being contextually responsive. Based on Save the Children's experience, several helpful strategies are presented for introducing rubrics into an existing M&E system: starting small and building momentum; opt-in piloting and gradual rollout; smart capacity building; providing more detailed guidance to evaluators and managers; getting buy-in to a specific measurement problem and solution rather than replacing the entire MEAL system; working with expert allies; and getting excited and innovative.

For change efforts focusing on social and cultural change (e.g., social norms, behaviors, beliefs, and traditions), the most important outcomes are almost invariably the hardest to measure. Take one of Save the Children's key outcomes for its child sponsorship program (referred to as Sponsorship), for example: "Children are included and living free from discrimination." Although there are a few indicators that might be informative, the heart and soul of this outcome are children's lived experiences of inclusion, exclusion, and discrimination.

Save the Children is the world's leading independent organization for children, working in 116 countries with development and humanitarian programs (Save the Children, n.d.-a). Sponsorship is a long-term development program supporting children, their families, communities, and institutions in 20 countries to create change together so that all children survive, learn, are protected, and live free from discrimination (Save the Children, n.d.-b). A priority and challenge for the program is centering the voices of children as co-designers of the change and capturing children's own stories about the outcomes affecting them.

Stories are rich and powerful sources of evidence, especially for capturing lived experience (Salm, n.d.), but it is harder to show clearly how well change is progressing over time. Herein lay the conundrum for the designers of Sponsorship's evaluation system—how could they capture and center children's stories in a way that shows trackable change, is child-friendly, and makes sense to children, parents, communities, program staff, and other stakeholders?

Rubrics methodology provides a breakthrough for challenges like this. A rubric describes what the situation looks like at different levels of performance on an outcome (Davidson, n.d.). When the main evidence includes children's experiences, a rubric can describe what kinds of stories would indicate a situation that is, for example, beneficial (or harmful) to children's wellbeing. This makes it possible to use those stories evaluatively to draw conclusions about how problematic or beneficial the situation is at baseline, as change unfolds, at program close-out, and later, as the change is or is not sustained (Chianca & Davidson, 2020).

Evaluative rubrics have been used for a long time in education, mostly for student assessment. Their application to evaluating programs, policies, and other evaluands can be traced back several decades, at least as far as the late 1990s and early 2000s. Since that time, two of us have been leaders, along with several other colleagues around the world, in developing a nuanced and sophisticated body of knowledge and know-how around using rubrics methodology to guide sound evaluative reasoning (e.g., Davidson, 2004, 2014b, 2025; Gargani & King, 2024; King et al., 2013; McKegg, 2011; Oakden, 2013; Wehipeihana, 2011). For those interested in applied examples, there are now numerous evaluation frameworks and reports that illustrate the use of rubrics methodology in diverse sectors and regions of the world (e.g., Chianca & Davidson, 2021; Chianca et al., 2009; Davidson, 2014a; Davidson & Chianca, 2020; King, et al., 2020; Laudes Foundation, n.d.; New Zealand Qualifications Authority, 2021; Nunns & Roorda, 2010; Wehipeihana et al., 2015).

This paper describes a new rubrics-enhanced monitoring, evaluation, accountability, and learning (MEAL) system developed with one of the world's largest international non-governmental organizations. The system helps Save the Children track progress on key outcomes in the impact areas where Sponsorship programming is implemented, especially on less tangible outcomes, such as children's experiences of inequality and discrimination, equitable and gender-transformative family environments, and empowered communities creating change for and in collaboration with children.

Rubrics, applied correctly, can help shift power to children, families, and communities for design and evaluation because they *describe* lived reality (as a progression from the current to the aspirational state) rather than representing it with abstract indicators (Davidson, 2024, 2025). Descriptions of change tend to be more understandable than indicators for children and their communities, making it easier to equip them to drive and monitor the social change they seek to influence.

In this paper, we start with a very brief explanation of the work Sponsorship does, Save the Children's push for child-led development and evaluation, and why Sponsorship opted for a rubrics-enhanced measurement and learning system to support these goals.

Next, we present some of the nuts and bolts of using rubrics to turn children's (and adults') stories of change into rich, trackable outcomes. We highlight how rubrics-enhanced evaluation lends itself particularly well to helping convert nuanced, story-based and other evidence into easily visualizable ways of showing progress, using one of the rubrics developed for Sponsorship—"Children are included and living free from discrimination."—as the main example throughout the paper.

We also provide advice for practitioners wishing to try out the rubric presented here in their

own work. Options are available either for applying the full rubric in its entirety, or for applying one or more of its subcriteria if not all of them are relevant for a particular project.

Finally, we close with reflections about how to get buy-in for a switch to this approach within an organization that is heavily invested in logframes and indicators. We also discuss breakthroughs, challenges, and hard-won lessons, as well as what lies ahead for this effort.

Why Rubrics Were Seen as a Worthwhile Enhancement to Save the Children's Sponsorship MEAL System

A priority and challenge in the way Sponsorship develops programming to address inequality and discrimination is to integrate children and their communities at all stages of the program cycle. As a child rights organization, Save the Children upholds the United Nations Convention on the Rights of the Child's general principles (Article 12), which state that children and young people should be considered when people make decisions about things that affect them, and that their views should be given "due weight" (United Nations, 1989, para. 12.1). Save the Children aims to implement this through their programming, accountability, and advocacy mechanisms. That means having children's voices not only in program design and implementation, but also and especially in the two other stages of the program cycle, measuring what happened and acting on findings. Often in those processes children and communities are only involved as informants or as audiences receiving information, not as co-leaders.

Sponsorship works through a 10-year program cycle, building community capacity to create social change that addresses the root causes of inequality and discrimination for a group of children living in an impact area. The theory of social transformation, in a nutshell, is that (1) identifying the root causes of inequality and discrimination (along with the agents of change who can help change those negative causes) and (2) adding approaches like child- and community-led action, social and behavioral change, and systems strengthening into a typical health, education, or protection program make it possible to (3) address those root causes at the child, family, community, institutional, and policy levels (Figure 1).

Figure 1. Save the Children's Social Transformation Conceptual Framework for Sponsorship
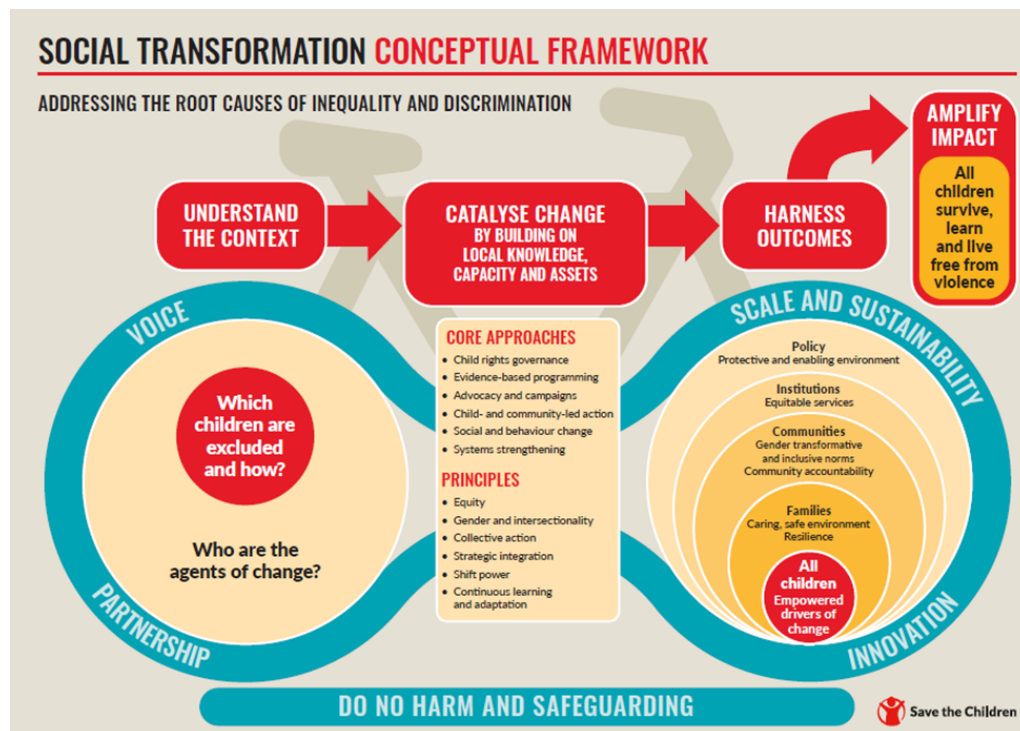
Since 2022, this program design has been piloted in El Salvador, the Philippines, and Zambia. There, children and their communities were actively involved in identifying problems. Program staff created and tested child-friendly and age-appropriate methodologies not only to listen to children's views but to prioritize with them which problems should be addressed with the program. Importantly, the program was co-designed with children and continues to engage children in decision-making via children's advisory committees. This process contributes to building on community acceptance of the program and helps Sponsorship offices to make better-informed decisions, together with children and their communities.

Working with rubrics as an evaluative methodology was a natural progression of this highly collaborative child-centered approach. Sponsorship wanted a tool that would facilitate understanding complex narratives of social change and measuring these changes over time, and that could go beyond simply gathering qualitative evidence from children.

Using Rubrics to Turn Children's Stories of Change Into Rich, Trackable Outcomes

Some of the most important social change outcomes that Sponsorship seeks to influence are also the hardest to measure; for example:

- Children are included and living free from discrimination.
- Equitable and gender-transformative family environments.
- Empowered communities creating change for and in collaboration with children.

Outcomes like these do not lend themselves well to a purely indicator-based approach to measurement. As already mentioned, Sponsorship's MEAL system like many other international non-governmental organizations' has historically been based mainly on logframes and indicators. Table 1 shows a sample of the kinds of indicators that Save the Children and their evaluation contractors typically used to measure their impact.

Table 1. Sample of Indicators Typically Used by Save the Children to Measure Outcomes

Logframe	Indicators
Overall Impact Result	
Out-of-school boys and girls at risk for early pregnancy and marriage, substance abuse, child labor and abuse are now protected, learning and thriving.	# out-of-school girls and boys # child marriages & teenage pregnancies # child labor
Outcome 1: Individual Level	
1.1 Girls and boys are increasingly aware of and exercise their rights / Out-of-school girls and boys exercise all their rights / Child-led engagement and dialogue with parents on children's rights.	% of supported girls and boys, and children with non-binary gender, that feel empowered to create change on behalf of children as a result of the engagement in Child Rights Reporting
1.2 Girls and boys have a safe and protective environment (including feeling safe in schools).	% of children who report they feel safe when they are with their family # schools passing the safety checklist with only 'none of the above' checked in all categories of the 'School Building Safety checklist'

Indicators can be relevant and should be used as one of the information sources. But, if considered in isolation, they don't lend themselves well to telling a clear story about the complex social change that Sponsorship is seeking to influence.

One of the most important ways of understanding how well social change is happening is to listen to people's stories and see how their lived

experience is changing (Falk, 2021). For a child-focused social change effort, what kinds of stories do evaluators need to ask about, and from whom, and how can they draw conclusions about social change from those stories? Rubrics are the main tool to use here.

As part of this project, the Real Evaluation team developed a set of 12 rubrics at the child, family, and community levels of analysis, in collaboration with Save the Children experts. In this paper, we focus on Rubric 11, which measures changes in social exclusion and discrimination at the child level. This is the one rubric for which we have developed detailed evidence-capture tools and a synthesis guide. We hope that the other rubrics will also be fleshed out in detail with instruments and evidence interpretation guides so that Sponsorship's MEAL practitioners around the world can use them consistently and with confidence.

What is a Rubric, and What Does This Example Look Like?

A rubric is a description of what performance looks like at different levels of an outcome (Davidson, n.d.; Davidson, 2024, 2025). For example, in Figure 2, we show the five levels created for Rubric 11 based on the outcome, "Children are included and living free from discrimination." These levels illustrate the progression from typical baseline levels (the bottom two light and dark orange levels) through to the eventual ideal situation, where children are included, valued, accepted, and treated kindly and fairly (the top two light and dark blue levels). In the full version of the rubric, these brief headings are fleshed out with rich descriptions of what the stories and other evidence should look like at each of these levels (Chianca et al., 2024).

Figure 2. Five Levels for Rubric 11: Children are Included and Living Free from Discrimination

To ensure good construct validity and contextual relevance, rubrics should be developed through collaborative processes involving the main groups affected by and interested in the initiative being evaluated. If rubrics are being created for use in just one setting, it is important to include community members in their development. For a multi-site program spanning multiple different contexts, it makes more sense to involve practitioners who have worked across many of these contexts.

In Sponsorship's case, we engaged highly experienced staff and other content experts from within Save the Children in a process facilitated by two of us as the rubrics methodology experts. Through a series of virtual meetings and several rounds of collective text revisions, this evaluation rubrics working group developed criteria and detailed descriptions to define the progression from a typical baseline situation to the desired end state for all 12 outcomes. As part of that process, the working group determined that a five-level scale used for a previous Sponsorship evaluation (*excellent*, *good*, *almost OK*, *problematic*, and *dire*) would be appropriate to establish baselines as well as to track progress over time for all 12 rubrics (Chianca & Davidson, 2020).

Based on feedback from initial pilot testing in Zambia, the Philippines, and El Salvador (see next section), we saw the need to simplify the original rubrics, starting with Rubric 11. We broke its content into six subcriteria, which cover the key

aspects in determining how well a community is doing in terms of child inclusion and nondiscrimination:

- 11a. Teasing, bullying, and abuse.
- 11b. Acceptance by families, friends, and the wider community.
- 11c. Social inclusion, exclusion, and discrimination.
- 11d. Children's understanding of the norms and barriers that perpetuate discrimination.

- 11e. Children's understanding of their rights.
- 11f. Children's collective strength to speak up.

Table 2 shows the full section of the rubric developed for subcriterion 11a ("Teasing, bullying, and abuse."). Similar descriptions were developed for the other five subcriteria (11b–11f) comprising Rubric 11. The full version of Rubric 11, including all six subcriteria, may be found on Save the Children's website (Chianca et al., 2024).

Table 2. Rubric 11, Subcriterion 11a: Teasing, Bullying, and Abuse

Levels / Criterion	Excellent	Good	Almost ok	Problematic	Dire
11a. Teasing, bullying, and abuse	Teasing and bullying are rare, mild, and always addressed quickly. Only by children (never adults).	Teasing and bullying sometimes occur and are mild, but not always addressed quickly. Only by children (never adults).	Teasing and bullying happen a lot and are mild to moderate. Adults may sometimes allow or encourage teasing/ bullying, but don't join in or abuse.	Teasing and bullying happen a lot and are moderately severe. Adults may sometimes join in, but no incidents of adult-inflicted abuse.	Teasing and bullying happen a lot and are severe, harmful, and considered normal. AND/OR: There are incidents of physical, sexual and/or emotional abuse inflicted by adults.

Pilot Testing the Rubrics

Before starting to use the rubrics more widely, it was essential to test them and make improvements as needed. The original Sponsorship rubrics were piloted by three country offices (El Salvador, Zambia, and the Philippines) in 2022, to establish a baseline for the forthcoming 10 years of work supporting the communities in the impact areas covered by their initiatives.

During the pilots, country office teams worked to adapt several of the 12 available rubrics for the three local cultures and contexts; gather story-based and other evidence to assess the nature, prevalence, and severity of children's experiences; and use this evidence to arrive at a baseline assessment on key social change outcomes.

The contextualization process involved working with children to identify which groups were most impacted by inequality and discrimination in their communities, and then listening to those impacted groups' stories to understand what kinds of things they were experiencing. That helped the local evaluation teams formulate the right questions to elicit locally meaningful stories of social exclusion, discrimination, abuse, and abandonment—as well as stories of inclusion and support.

Country office teams were then coached in sensemaking processes, in which diverse sources of evidence were synthesized alongside the rubrics to provide a clear assessment of how harmful or conducive the situation was to children's wellbeing. More details on how to do this are provided in the next section.

The pilot testing revealed some helpful insights not only for sharpening the rubrics and their application but also for better understanding the advantages and limitations of rubrics-enhanced evaluation processes in different cultures and contexts. The following are some of the key insights:

- Community members said that their participation in the rubrics-guided process was "empowering and transformative." It was empowering because it provided an opportunity for them to deeply reflect on key issues and express their perspectives and opinions throughout the process. It was

transformative because they got a chance to hear diverse perspectives on issues affecting children, shedding light on a range of family and community key issues that they may not have been fully aware of.

- Insights from children themselves revealed nuances not always acknowledged by adults. For example, in one community, children expressed concerns about gender-based discrimination, stating that "although some [village] officials and parents claim no discrimination against children … parents tend to impose more responsibilities on their daughters than their sons." In another community, children explained that "differently from what adults say, youth still experience discrimination in the [village], particularly those who become pregnant early."

- Working with Save the Children's child rights, gender, and disability experts when developing the rubrics provided rich and nuanced descriptions of what kinds of child discrimination and exclusion stories are typical. However, the rubrics had initially been designed for use by the evaluation teams, not directly in participatory processes involving children, parents, community leaders, and other key actors—processes that were now an important part of Sponsorship's approach. This made the process of applying the original rubrics with those different community groups during the piloting phase challenging. The original language used was not always child/community friendly, and various facilitators had to adapt it in the field, making it difficult to ensure consistency. Nevertheless, the use of rubrics as a participatory method to rate changes in impact areas emerged as a promising avenue to support locally led decision-making.

- Rubrics have potential to shift power by engaging children and other community members in measuring changes in inequality and discrimination. By using Save the Children's child- and community-friendly tools (simplified versions of the rubrics), children and community members can monitor the social change they are trying to achieve and advocate more effectively for themselves with decision-makers.

- Local evaluators recognized the potential of rubrics for (a) bringing together different types of evidence (qualitative and quantitative information from primary and secondary sources), (b) synthesizing the evidence using a clear framework, and (c) eliciting informed and thoughtful conversations to reach collective ratings for each community about the situation for children. However, while MEAL professionals are still gaining familiarity with rubrics methodology, much more specific tools and guidance are required.

## Determining a Rating on Each Subcriterion

The application of a rubric would typically involve the evaluation team determining for themselves what evidence would be required to make a rating on each subcriterion, and then using a structured deliberation process to consider that evidence together and arrive at a rating. However, as the pilot testing revealed, this approach requires a relatively high level of familiarity with translating rubric descriptions into evidence-capture instruments, as well as with the evaluative sensemaking and reasoning process needed to interpret a mix of qualitative and quantitative evidence from multiple sources and arrive at valid ratings.

Although some of the country offices happened to have access to a particularly experienced practitioner available to support this kind of learning, we couldn't assume that this was the case everywhere. Save the Children needed detailed instructions to ensure consistency and validity across sites and to ensure that all MEAL practitioners gained a sound understanding of this methodology, which was new to them, and could use it with confidence.

To address this issue, we worked with Save the Children content experts to develop supplementary tools to help the country offices get the assessment right and more thoroughly learn the methodology using step-by-step guidance. These tools included a set of evidence-capture tools—focus group discussion (FGD) and key informant interview (KII) protocols—as well as instructions showing exactly which evidence to apply to which sections of the rubric and how to interpret the evidence.

The protocol developed for this is outlined as a flowchart in Figure 3. On the left, under Evidence Sources, we listed the specific questions from each of the FGD and KII protocols that should be used for the analysis. In the diamonds are the deliberation questions for the evaluation team, with arrows showing which rating should be given, depending on the evidence.

Figure 3. Flowchart to Generate a Rating for Subcriterion 11a (Teasing, Bullying, and Abuse)
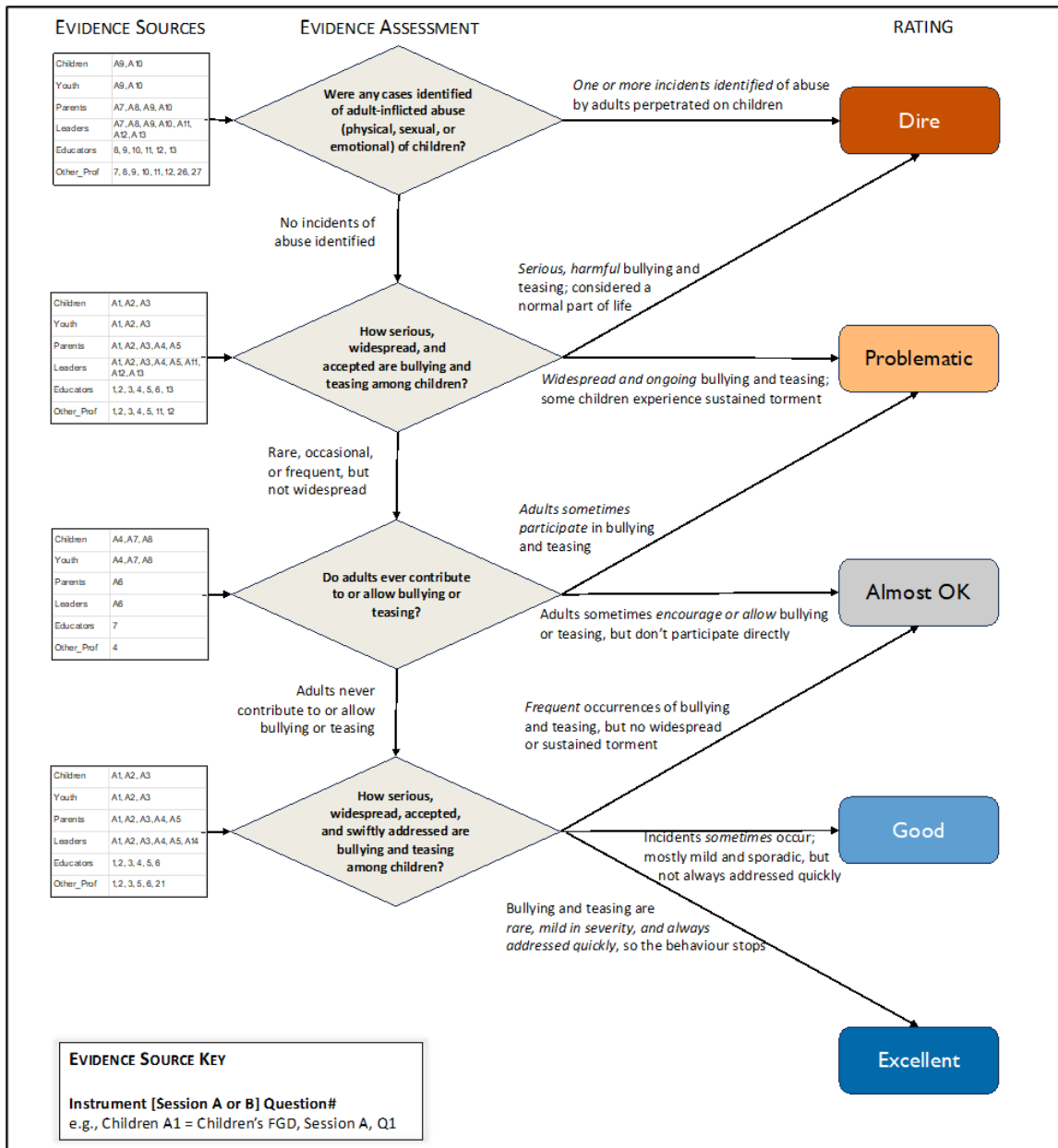
Let's walk through how this works. In the first section of the flowchart, the Evidence Sources box lists all FGD and KII questions pertaining to abuse perpetrated by adults on children. The team conducting the analysis examines the answers to these questions, along with any other relevant evidence they may have (e.g., incident reports from social workers, health care professionals, or police).

If any of the evidence identifies one or more instances of child abuse, the team is instructed to rate subcriterion 11a as *dire*. If no such instances are found, then the team proceeds to the next Evidence Assessment question, which asks about the prevalence and severity of bullying and teasing. Again, the arrows indicate whether a *dire* or *problematic* rating should be ascribed or whether

the situation is mild enough to consider the next collection of evidence sources.

The contents of the flowchart lay out the evaluative reasoning steps required to use the rubric, making explicit how the evidence should be weighed and considered to arrive at a rating on subcriterion 11a. This makes it clearer for the team conducting the analysis exactly which pieces of evidence should be considered, in which order, and what kinds of things they should be looking for. The evaluative reasoning process is not simple or algorithmic though. It requires nuanced and contextually valid judgment calls, such as what constitutes "serious and harmful" or "widespread" bullying. These are judgments best made by those who are closest to the affected children and the context in which they live, not pre-defined by those outside that context.

The flowcharts are designed to mirror the reasoning process of an experienced rubrics practitioner, providing a practical way for MEAL teams to build that know-how for themselves. Each of the six subcriteria have similar flowcharts to guide the step-by-step consideration of evidence by evaluation teams in each impact area. The intent here is to ensure enough rating consistency across the system while still allowing local deliberation and interpretation of the evidence in ways that make sense in the local context. For example, every country has somewhat different cultural norms related to how people interact regarding personal space and physical contact. So it is conceivable that an interaction that would be considered largely innocuous in one culture (kissing and hugging when greeting someone) could be considered quite inappropriate or intrusive in another (Toll, n.d.).

Allowing locally knowledgeable teams to deliberate on the evidence together can help to ensure culturally and contextually appropriate values and understandings are being applied. This is an important reason why, when designing rubrics, it is important to deliberately underspecify what the evidence should look like so as not to impose specific ways of looking at things on widely varying cultural contexts. Even in this example, where the same evidence-capture instruments are used in different contexts, the local evaluative sensemaking process allows for nuanced, culturally relevant interpretations rather than a one-size-fits-all approach to interpreting the evidence.

As with all rubrics-enhanced evaluative reasoning, there is always room for some disagreement on which rating is appropriate. However, in our decades of experience using this methodology, disagreements have seldom spanned more than two adjacent levels and have usually been resolved via skillfully facilitated discussion.

Moreover, when grappling with complex social change, conversations are often far more important and valuable than final ratings, because they push stakeholders to unpack nuances that may be helpful when trying to address traditional norms and beliefs that are harmful to children.

Synthesis of Subcriterion Ratings Into an Overall Rubric Rating

Rubrics dealing with multifaceted outcomes sometimes encompass several subcriteria rated separately, as is the case here. One common challenge is the movement from different ratings for subcriteria to an overall rating for the whole rubric.

Many evaluators would default to weighting the subcriteria and summing them up, a synthesis methodology known as Numerical Weight and Sum (NWS; Scriven, 1991). However, that methodology implies that good ratings on some subcriteria can make up for poor ones on another, which is seldom a valid assumption. The synthesis methodology for this set of subcriteria uses a more nuanced and valid approach to evaluative reasoning, which considers the more important and the less important aspects a little differently.

Not all of the subcriteria for Rubric 11 are equally important. Subcriteria 11a, 11b, and 11c are the top-priority criteria. They are the most critical for this outcome, because low ratings in these areas are the most problematic for children. When children are seriously bullied by other children or abused by adults, when they are abandoned by their families or made to feel like they are a burden, and when they are treated unfairly or not allowed to participate in important aspects of family, school, and community life—these are the things that Sponsorship's experts pointed out as most strongly affecting children's well-being with respect to this outcome.

The rest of the rubric, subcriteria 11d, 11e and 11f, are the other related aspects of social change that Sponsorship is seeking to influence. These subcriteria are important and are part of the picture, but they don't impact children's well-being quite so dramatically. When children understand the norms and barriers that perpetuate discrimination, when they can explain their rights as children, and when they have the collective strength to speak up and call out instances where children are being treated unfairly—these are important aspects of achieving real social change. However, they don't affect children's well-being

quite as dramatically as the other three, hence their slightly lower importance.

Even though that makes it a little easier to do the rating on each of the subcriteria separately, it does require a synthesis step to get to an overall rating for the whole rubric.

To guide the synthesis of subcriterion ratings, the Real Evaluation team developed a user-friendly evaluative reasoning flowchart (Figure 4). The step-by-step decision process to determine which rating to choose is guided by answers to a series of evidence assessment questions (inside the light gray diamonds).

Figure 4. Evaluative Reasoning Flowchart to Generate an Overall Assessment of Focus Children's Experiences of Social Inclusion and Discrimination.
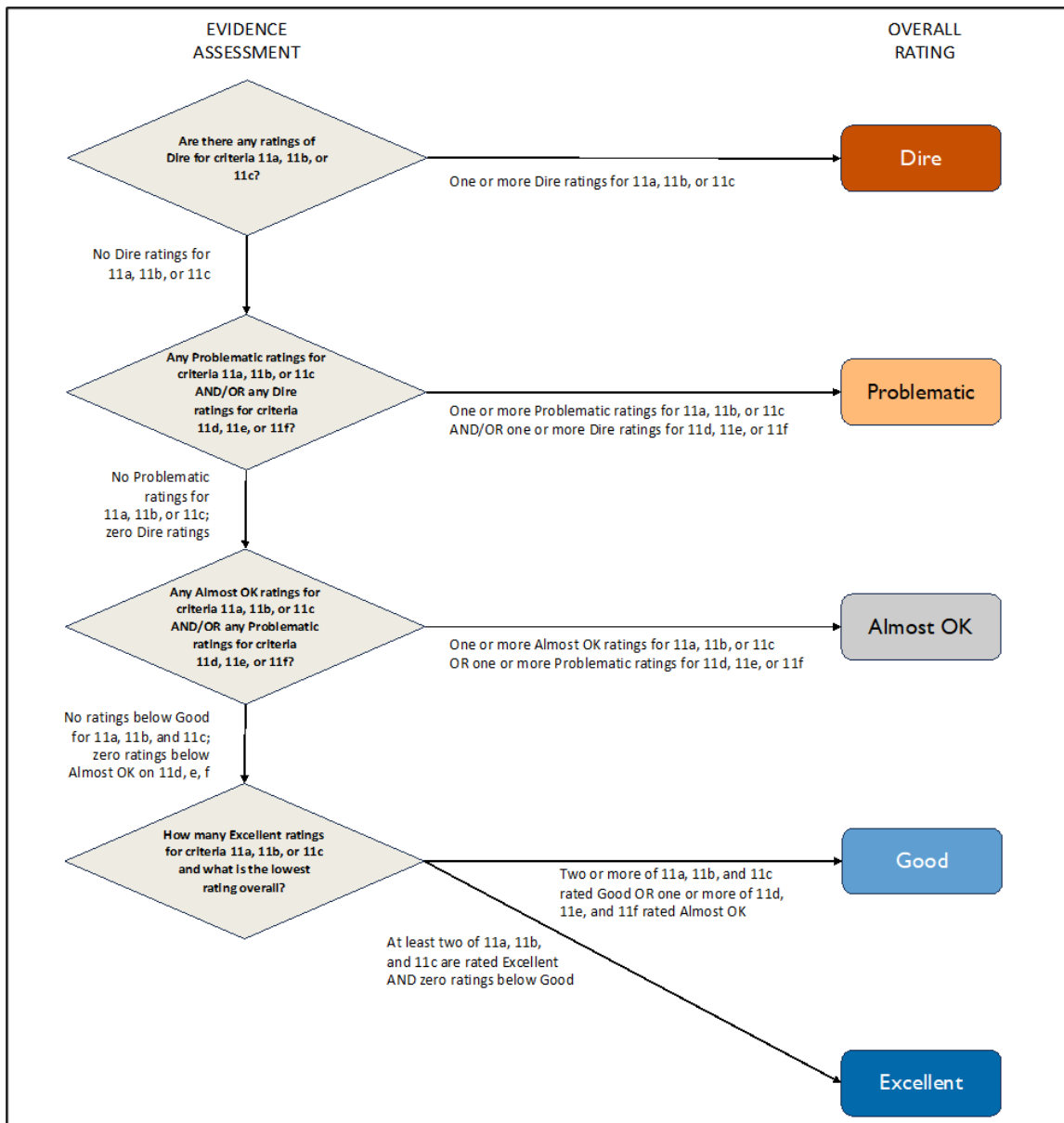
Looking at the first assessment question, if there are *any dire* ratings on the three most important subcriteria—in other words, there were instances of child abuse (from subcriterion 11a), child abandonment (11b), and/or overt exclusion (11c)—that will automatically yield an overall rating of *dire*, *regardless* of whether good things might also be happening, because these things are extremely harmful to children.

If those very serious things aren't happening, but instead evidence shows that children have no idea that they have rights (subcriterion 11d), or they

know nothing about how discrimination is perpetuated (11e), or they don't have the strength to speak up when they see something bad happening (11f), a *dire* rating may be applied in any of these slightly less important subcriteria, but the broader situation will be considered *problematic* overall rather than *dire*.

The next part of the same evidence assessment question (in the second diamond) shows the other evidence that could indicate a *problematic* rating: If one or more of the important criteria (11a, b, or c) is rated *problematic*—meaning, if mild to moderate bullying happens a lot (11a), or if some children feel only partially valued and accepted (11b), or if they tend to be left out of *most aspects* of school, family, and community life (11c)—then the overall rating will be *problematic*.

The logic is similar for the next level too. If one or more of the *less important* criteria (11d, 11e and 11f) is rated as *problematic* and/or if one or more of the *more important* criteria (11a, 11b, and 11c) is rated as *almost OK*, the overall rating will be *almost OK*.

Now, if there are no critical criteria (11 a, b, or c) rated below *good* and no less critical criteria (11 d, e, or f) rated below *almost OK*, we will be in the *good* or *excellent* rating territory. If at least two of the more important criteria (11 a, b, and c) are rated as *excellent* and there are no ratings below *good*, then *excellent* will be the final rating. Anything in between will be rated *good*.

A frequently asked question is how to come up with a flowchart like this. Isn't it all just subjective? It was certainly constructed by humans, but the Real Evaluation team did use careful evaluative reasoning to ensure robust validity. In short, it was designed to mirror the evaluative reasoning process that experienced rubrics practitioners would use with the available evidence.

As with any evaluative reasoning, simply documenting it clearly like this isn't enough to establish validity. The most important thing at this stage is to make the reasoning absolutely clear and transparent so that stakeholders and evaluation teams can clearly follow it and critique it when needed. Additional validation will come when evaluators pressure test this methodology in the field as it is rolled out by Save the Children's country offices. Evaluation practitioners applying the methodology in the field will provide feedback on whether the evaluative reasoning flowchart and its resulting conclusions make sense to them and to the communities they are working with. Important feedback will also address whether the methodology is sufficiently user-friendly and fit-for-purpose for communities to use themselves in child-led change efforts.

Rubric Ratings Can Show Trackable Progress Over Time

One of the great things about rubrics is that they are a way to turn rich stories and other mixed-method evidence into trackable outcomes that can show progress without boiling things down to a single, narrow indicator.

Figure 5 shows an adapted version of a diagram the Real Evaluation team has used in a different evaluation to show progress on a rubric-rated outcome at multiple time points (Chianca & Davidson, 2020).

Figure 5. Sample Diagram Showing Progress Over Time on an Outcome Rubric
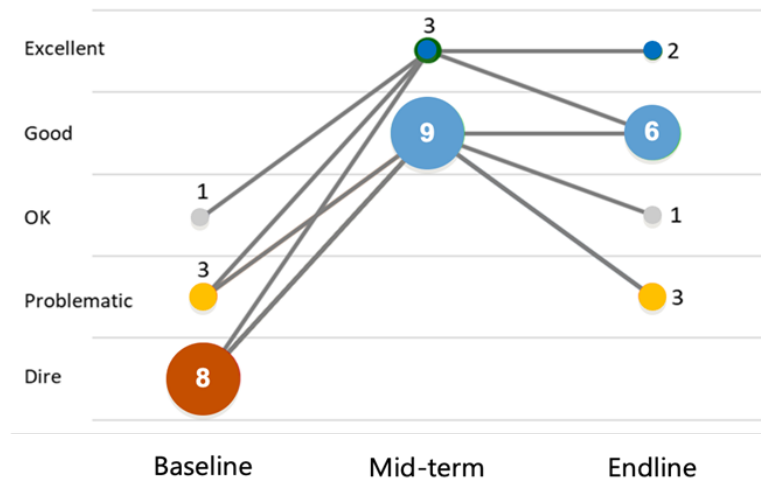
Under each of these data points lies not just one indicator but a rich mix of stories and other evidence representing a complex social reality in the community. Quotes and summaries of those stories, along with any available supporting statistics, can be used to illustrate further what the graph means.

There are many different ways to use such diagrams; for example, (i) showing change for the whole impact area (one rating at each time point); (ii) showing change for each village/settlement within the impact area (in Figure 5) 12 settlements are rated at three time points); and (iii) showing how things are changing for each demographic of children who are the focus of change in the impact area (e.g., separate or combined graphs for children with disabilities, pregnant and parenting teens, etc.).

Meeting the Challenge of Building a Toolkit for Nuanced yet User-Friendly Evaluation

In our first attempt to pilot the rubrics, we applied the usual approach of coaching the teams on how to contextualize the rubrics for their specific country/region/impact area realities and left for them the decision of how to gather the needed evidence in the most locally appropriate way. However, country office teams found this process to be quite complicated given their lack of familiarity with rubrics methodology. We then pivoted our approach to meet those needs.

Based on the piloting experience, Save the Children made the decision to create much more detailed guidelines and specific evidence-gathering instruments—focus group and interview protocols that could be used by country office teams to gather the needed evidence. These would form part of a toolkit for facilitating nuanced yet user-friendly baseline and outcome assessments that could be consistently and validly used by country office staff and the local external evaluators contracted to support their work.

To be aligned with Save the Children's principles, the toolkit is designed to center the perspectives of children and involve them as much as possible in the whole process. This was achieved through close collaboration with some of Save the Children's top child participation specialists. With their support, the instruments and guidelines have been simplified as much as possible and include interactive exercises to make the focus group discussions more fun and engaging for the children. Although focus group discussions are not the most child-friendly of methods, the team sought to create tools using the methodologies most familiar to country office teams.

The child-friendly evidence gathering protocols were then adapted for use with youth, parents and caregivers, community leaders, educators, health care professionals, and other professionals who worked closely with the impact area communities (e.g., health care professionals, social workers, police, etc.). The full version of the toolkit is freely

available for public use on Save the Children's website (Chianca et al., 2024).

One of the challenges from Save the Children took the Real Evaluation team into some very new territory with rubrics methodology. With their emphasis on child-led development and evaluation, they were keen that the children themselves be asked to rate the situation in their communities using the rubrics. As one of the Save the Children's child participation specialists working with us pointed out, this would make the process not just an evidence-gathering exercise, but also a pedagogical and empowering experience for the children and the other community members and professionals living or working in the impact areas.

To create child-friendly rubric snippets that could be used to elicit ratings from children, we had to go through each subcriterion, identify each one's main components, and come up with simple questions that the children could answer as a group, after completing a collective exercise, discussion, or game about the issue at hand. These simplified

rubric snippets were used to elicit children's (and others') assessments of the situation on each subcriterion. Separate tools were developed for different age groups, to ensure that age-appropriate questions and methods were used.

One of the simplified rubric snippets is shown in Table 3. These are the questions children should answer to assess the situation related to bullying, teasing, and abuse (i.e., subcriterion 11a) in their communities. First, we asked them how often bullying and teasing happen in their schools and communities. Next, we asked how serious that bullying and teasing is. Finally, we asked them what adults do that either helps or hurts children. The response options for the third question are used to assess how responsive adults are to issues of bullying and teasing; whether they allow, encourage, or join in; and whether there are any instances of abuse being perpetrated by adults on children.

Table 3. Child-Friendly Rubric Snippet to Assess the Extent and Severity of Bullying, Teasing, and Abuse

Questions:	Excellent	Good	Almost OK	Problematic	Dire
1. How often do bullying and teasing happen at your school or in the community?	Bullying and teasing <u>almost never</u> happen	Bullying and teasing happen <u>sometimes</u>, but not a lot	Bullying and teasing happen <u>a lot</u>		
2. How serious is this bullying and teasing?	It is <u>never</u> serious		It is sometimes <u>a little</u> serious	It is sometimes serious	It is sometimes <u>very</u> serious
3. What do adults do that either helps or hurts children?	When bullying and teasing happen, adults <u>always</u> help stop it quickly	When bullying and teasing happen, adults <u>don't always</u> help stop it quickly enough	When bullying and teasing happen, adults sometimes <u>allow or encourage</u> it	When bullying and teasing happen, adults <u>allow or encourage</u> it and <u>may sometimes join in</u>	Some adults do or say cruel things to some children again and again

To make it easier for the children to rate each question, the Real Evaluation team was careful not to include too many levels. For example, when asking how often bullying happens, there were only three options: *almost never*, *sometimes*, and *a lot*. The response options in each line of the simplified rubric snippet mirror the evaluative reasoning in the original rubric, without building in a level of precision that didn't exist in the real world (Davidson, 2024, 2025).

In addition to the simplified rubric snippet, the Real Evaluation team also created a simplified rubric synthesis guide (Table 4) to help the children determine which overall rating best represented the situation of their community for that specific criterion. The guide starts from the *dire* situation. If any of the children's answers to Questions 2 or 3 are *dire*, then *dire* is the overall rating. If nothing is *dire* and answers to Questions 2 or 3 are rated *problematic, then problematic* is the final rating. The other levels are addressed similarly.

Table 4. Simplified Rubric Synthesis

Similar rubric synthesis guides have been created as part of the evidence-gathering protocols for other key informant groups—youth, parents/caregivers, community leaders, educators, and other professionals (health, social work, police, etc.). These are available on Save the Children's website (Chianca et al., 2024).

We are optimistic that the current version of the child-friendly toolkit will work better than the previous version that was piloted in the field. However, after ensuring that we have clearance from Save the Children's ethical review boards, we will be paying careful attention to how user-friendly the processes are from the perspective of the country offices and their contracted evaluation teams, as well as children and community members themselves. Some adaptations of the toolkit contents and formatting will certainly be required along the way. Rubrics and their accompanying toolkits should always be considered living documents and refreshed periodically to capture new understandings and priorities.

How to Start Using These Rubrics in M&E Work

How would an evaluation professional working with an initiative seeking to influence similar outcomes for children and youth go about trying out this rubric? We have some tips and resources that may be useful:

If you haven't yet had the opportunity to work extensively with rubrics under the guidance of someone highly experienced in the methodology, take some time first to understand how rubrics work. A number of references are provided in this paper; perhaps the easiest one to start with is the short description of rubrics methodology on the Real Evaluation website (Davidson, n.d.) and then moving to more detailed explanations (e.g., Davidson, 2004, 2025; McKegg, 2011; King et al., 2013; Oakden, 2013; Wehipeihana, 2011).

Next, carefully review the full version of Rubric 11, which is available for download from Save the Children's website (Chianca et al., 2024). Compare the content of the rubric with the description of the relevant outcome you need to evaluate for your current project. Are all six subcriteria (11a through 11f) relevant? If so, you can use the full toolkit in its entirety. If not, you may elect to use one or more subcriteria.

Whether all six subcriteria are to be used or not, start by identifying the subgroups of children and youth who are most affected by discrimination and exclusion in the communities where the initiative might influence change. Instructions for this are included in the set-up sections of each FGD and KII protocol.

Depending on the nature of the initiative being evaluated and what is likely to be most useful for those leading and guiding the change, you may choose to make rubric ratings separately for each subgroup of children, separately for each village/settlement/school/etc., or overall, for the entire population of children covered by the initiative. This way, the rubric ratings may be used to track progress over time at whatever level of analysis will be most informative.

For some projects, you may find that some of the subcriteria are relevant but not others. For example, if you are evaluating a program that is

trying to help children understand their rights, then subcriterion 11e ("Children's understanding of their rights.") may be useful, but perhaps not the others. In this case, treat subcriterion 11e as the entire rubric.

Go through the relevant section of the "Evidence Interpretation Guide" (Chianca et al., 2024)—for subcriterion 11e, see page 11 of the guide—and note exactly which questions from which FGD and KII protocols are relevant to the subcriterion (or subcriteria) you plan to use. Copy these questions into your own FGD and KII protocols. For example, for subcriterion 11e, find questions A11 to A19 in the "Facilitator's Guide for Focus Group Discussions: Children's Sessions A & B" (Chianca et al., 2024) and drop these questions (and the associated instructions, including those at the front end of the packet) into your own FGD protocol for children. Do the same for each of the evidence capture instruments (FGD and KII protocols) until you have extracted all questions relevant to subcriterion 11e for all stakeholder groups you plan to speak with.

Conduct the FGDs and KIIs with the relevant individuals and groups. Once the evidence is gathered, return to the "Evidence Interpretation Guide" and follow the instructions. On page 4 of the guide, summarize the simplified ratings on subcriterion 11e made by the individuals and groups during the FGDs and KIIs. Keep each individual's and group's reasoning for each rating handy so you can refer to them later.

Using the flowchart on page 11 of the "Evidence Interpretation Guide," review the answers to the relevant questions (listed in the boxes on the left) and assess them using the evidence assessment questions (in the diamonds down the middle), choosing at each juncture which response best fits the evidence, and following the flowchart accordingly until you arrive at a rating.

Be sure to work through this process as a group, ideally including not just evaluation professionals but also program staff working directly with communities, and any community members working closely with you on the evaluation. Make a note of any outliers or anomalies in the evidence; this will be useful later as you review your findings and respond to any questions about them.

At the end of this collaborative sensemaking process, you should have a rating for subcriterion 11e: "Children's understanding of their rights."

If all six subcriteria in Rubric 11 are relevant for your project, use the complete FGD and KII protocols to guide your evidence gathering for all of the individuals and groups you are able to speak with. Using the Table on p. 4 of the "Evidence Interpretation Guide," summarize the simplified ratings made by the individuals and groups during the FGDs and KIIs.

You may do this separately for each subcriterion, or for the overall rating, or for both (we recommend both). Keep the individual's and group's reasoning for each rating handy so you can refer to them later. Finally, work through the guide to make ratings on each subcriterion (11a to 11f), and then to synthesize all six ratings into an overall rating on Rubric 11, using the level of analysis that you have determined will be most useful to inform and understand change.

Tips for Bringing Rubrics Into Your Own Organization's MEAL System

Like many large organizations working in international development, Save the Children prioritizes delivering funding directly to programs impacting children. Each country invests 5% of their budget in MEAL to demonstrate results. This leaves very little for central analysis and consolidation of results, which is why rubrics are the ideal tool to consolidate information. Teams face additional barriers that others also experience: low resources available, competing priorities, and overburdened country offices with limited bandwidth.

The sponsorship program also faces some barriers and considerations that are a little more unique:

- The work takes place in complex humanitarian environments.
- Sponsorship is a long-term program, with 10-plus-year programming cycles.
- MEAL staff may not be specialized in multiple evaluative techniques.
- Everything needs to be informed by children's voices and stories.

Sponsorship's M&E tools and processes have historically struggled to capture impact at all of the different levels the program aims to help change (child, family, community, institutions, and policy). A particularly difficult challenge is that some of the most complex social change efforts are aimed at norms, beliefs, and traditions that are deeply embedded in local culture but several of which are harmful to children. So it is a delicate balance between being appropriately culturally responsive while also prioritizing children's fundamental human rights, with all of the measurement and evaluation challenges that come with that.

To help overcome these barriers, Sponsorship Program colleagues at Save the Children adopted several helpful strategies:

Starting Small and Building Momentum

Sponsorship's MEAL teams decided to ask the Real Evaluation team to start with the "transformational change" outcomes in the global sponsorship theory of change (Figure 1), because this was where they were having the greatest difficulty trying to meaningfully measure social change. The initial set of 12 rubrics were at the child, family, and community levels; the institutional- and policy-level outcomes were parked for later development.

Opt-in Piloting and Gradual Rollout

Second, Sponsorship chose to pilot a selection of these early rubrics in just three countries whose evaluation teams expressed an interest. This was to establish proof of concept and to guide improvements to the system before attempting to roll them out globally. Sponsorship's ongoing program of field testing and refinement is not mandated anywhere, but instead offers opportunities to try out this new methodology when countries are conducting their evaluations. Based on the learning from the initial piloting phase, Sponsorship developed the system in a strategic and targeted way using budgetary and other resources as judiciously as possible while also maximizing buy-in.

Smart Capacity Building

The Real Evaluation team worked deeply with the three pilot countries to coach and mentor teams of people (not just MEAL practitioners, but program staff as well) through the rubrics application process over the course of 6 months.

Providing More Detailed Guidance

Working collaboratively with Save the Children's internal experts, the Real Evaluation team developed detailed FGD and KII facilitation guides and toolkits to guide and create a common way for the various country offices to collect and make sense of evidence while still allowing for appropriate localization. These tools and guides have been designed to give the level of detail that the pilot offices said were needed in order to use the rubrics system with confidence while assuring good reliability and validity. Save the Children plans to pilot these new tools and guides in the near future.

Getting Buy-in to a Specific Problem and Solution Rather Than Replacing the Entire System

Save the Children's long-established M&E systems primarily use logframes and indicators. Rather than trying to convince internal leaders to change the entire system to rubrics, Sponsorship staff focused on establishing buy-in to a specific measurement problem. Rubrics are offered as a solution to the issue of the organization not having standard measures to evaluate changes in discrimination. In other words, Sponsorship is filling a problematic gap in the current system rather than trying to replace it entirely.

Working with Expert Allies

We chose to work with "warmer" allies across the Save the Children movement to develop the definitions of the rubric levels. The process of developing rubrics required significant input from in-house experts to ensure that the content reflected the best and deepest knowledge accumulated internally. By finding allies across the organization who had the right kind of expertise and were interested in the approach, the project not only benefited from their input but also ensured that key experts across Save the Children International would have strong familiarity with rubrics methodology and could help support the work, both as resource people and as advocates.

Getting Excited and Innovative

Enthusiasm and interesting ideas draw people in. By positioning not just the MEAL function but the rest of the Sponsorship portfolio as an innovative team, Sponsorship has been able to generate excitement about the work. The program team is working on failure festivals to celebrate learning from failures, investing in innovation capacity, trialing new program design methods, and finding creative new ways to measure important outcomes. Rubrics are seen as one of many exciting new approaches that Save the Children International is pioneering, and that in itself is a win.

Save the Children is certainly convinced that rubrics are the right path to follow and a powerful complement and enrichment to their existing indicator-based tracking systems.

References

Chianca, T. K., & Davidson, E. J. (2020). *Retrospective impact evaluation: Save the Children's Sponsorship programming in Woliso impact area, Ethiopia (2002–2010).* Save the Children Child Rights Resource Center. https://resourcecentre.savethechildren.net/document/retrospective-impact-evaluation-save-childrens-sponsorship-programming-woliso-impact-area/

Chianca, T. K., & Davidson, E. J. (2021). *Fashion for good: Independent evaluation.* COMEA Relevant Evaluations, Rio de Janeiro, Brazil. https://www.laudesfoundation.org/media/y4ydafsd/2021-02-19ffg-report-final.pdf

Chianca, T. K., Davidson, E. J., Dulieu, N., & Sigdel, A. (2024). *Rubrics 11: A toolkit to address child discrimination and exclusion.* Save the Children Child Rights Resource Center. https://resourcecentre.savethechildren.net/document/rubrics-11-a-toolkit-to-address-child-discrimination-and-exclusion/

Chianca, T. K., Scriven, M., & Kern, R. (2009). *A qualitative and quantitative approach to impact evaluations of international development interventions: The case of Heifer International in 20 countries, 2005–2009* [Presentation]. American Evaluation Association conference. https://comm.eval.org/viewdocument/a-qualitative-and-qu

Davidson, E. J. (n.d.). What are rubrics? *Real Evaluation LLC.* https://realevaluation.com/what-are-rubrics/

Davidson, E. J. (2004). *Evaluation methodology basics: The nuts and bolts of sound evaluation.* Sage.

Davidson, E. J. (2014a). *Evaluation of the Hutt Valley's first teen road safety expo, June 2014.* Real Evaluation, Auckland, New Zealand. https://www.safecommunities.org.nz/application/files/7614/4963/3978/Appendixes-_Safe_Hutt_Valley_reaccreditation.pdf

Davidson, E. J. (2014b). *Evaluative reasoning* (Methodological Briefs: Impact Evaluation 4). UNICEF Office of Research.

Davidson, E. J. (2025, forthcoming). *Evaluation methodology basics: The nuts and bolts of sound evaluation* (2nd ed.). Sage.

Davidson, E. J., & Chianca, T. K. (2020). *Retrospective impact evaluation: Save the Children's sponsorship programming in Woliso Impact Area, Ethiopia (2002–2010).* Real Evaluation, Seattle, WA.

https://resourcecentre.savethechildren.net/document/retrospective-impact-evaluation-save-childrens-sponsorship-programming-woliso-impact-area/

Falk, E. (2021, June 27). Op-Ed: Why storytelling is an important tool for social change. *Los Angeles Times.* https://www.latimes.com/opinion/story/2021-06-27/stories-brain-science-memory-social-change

Gargani, J., & King, J. (2024). Principles and methods to advance value for money. *Evaluation, 30*(1), 50–68.

King, J., McKegg, K., Arau, A., Schiff, A., & Garcia Aisa, M. (2020). *Social and economic impact assessment of mutation breeding in crops of the RCA Programme in Asia and the Pacific.* International Atomic Energy Agency. https://www.iaea.org/sites/default/files/20/11/social-and-economic-impact-assessment-of-mutation-breeding-in-crops-of-the-rca-programme-in-asia-and-the-pacific.pdf

King, J., McKegg, K., Oakden, J., & Wehipeihana, N. (2013). Rubrics: A method for surfacing values and improving the credibility of evaluation. *Journal of Multidisciplinary Evaluation, 9*(21), 11–20. https://doi.org/10.56645/jmde.v9i21.374

Laudes Foundation (n.d.). *Measuring and learning with rubrics.* https://www.laudesfoundation.org/how-we-work/measuring-with-rubrics/

McKegg, K. (2011). Knowing about rubrics is one thing … tips, tricks and traps. In E. J. Davidson, N. Wehipeihana, K. McKegg, & P. Rogers (Eds.), *The rubric revolution* (pp 24–44) [Presentation at the Australasian Evaluation Society conference, Sydney, Australia]. https://www.betterevaluation.org/tools-resources/rubric-revolution

New Zealand Qualifications Authority (2021). *Guide to external evaluation and review – for tertiary education organisations.* NZQA, Wellington, New Zealand. https://www2.nzqa.govt.nz/assets/Tertiary/External-evaluation-and-review/Guide-to-EER/eer-guidelines.pdf

Nunns, H., & Roorda, M. (2010). *Final evaluation report of the Recognised Seasonal Employer Policy (2007–2009).* New Zealand Department of Labour, Wellington, New Zealand. https://www.betterevaluation.org/tools-resources/merit-determination-rubrics

Oakden, J. (2013) *Evaluation rubrics: How to ensure transparent and clear assessment that respects diverse lines of evidence.* BetterEvaluation, Melbourne, Australia.

https://www.betterevaluation.org/tools-resources/evaluation-rubrics-how-ensure-transparent-clear-assessment-respects-diverse-lines-evidence

Salm, M. (n.d.). Personal stories. *Better Evaluation.* https://www.betterevaluation.org/methods-approaches/methods/personal-stories

Save the Children (n.d.-a). *Where we work.* https://www.savethechildren.net/where-we-work#

Save the Children (n.d.-b). *Sponsorship programs.* Save the Children Child Rights Resource Center. https://resourcecentre.savethechildren.net/collection/sponsorship-programs/

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Sage.

Toll, E. (n.d.). *Greetings and customs around the world.* DiversityResources. https://www.diversityresources.com/greeting-customs-around-the-world/

United Nations Convention on the Rights of the Child, November 20, 1989, https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-rights-child

Wehipeihana, N. (2011). Using rubrics to give voice to Indigenous values. In E. J. Davidson, N. Wehipeihana, K. McKegg, & P. Rogers (Eds.), *The rubric revolution* (pp. 15–23) [Presentation at the Australasian Evaluation Society conference, Sydney, Australia]. https://www.betterevaluation.org/tools-resources/rubric-revolution

Wehipeihana, N., McKegg, K., Thompson, V., & Pipi, K. (2015). Cultural responsiveness through developmental evaluation: Indigenous innovations in sport and traditional Māori recreation. In Patton, M. Q., McKegg, K., & Wehipeihana, N. (Eds.), *Developmental evaluation exemplars: Principles in practice* (pp. 25–44). Guilford Press. Project summary, Te Whetu Rēhua rubric, and link to a video explaining the rubric available online: https://www.kinnect.co.nz/he-oranga-poutama/