

To rubrics or not to rubrics?

An experience using rubrics for monitoring, evaluating and learning in a complex project

Samantha Stone-Jovicich, CSIRO

August 2015

In this Practice Note I share our experience using an evaluation and monitoring approach called 'rubrics' to assess a complex and dynamic project's progress towards achieving its objectives. Rubrics are a method for aggregating qualitative performance data for reporting and learning purposes. In M&E toolkits and reports, rubrics looks very appealing. It appears capable of meeting accountability needs (i.e. collating evidence that agreed-upon activities, milestones, and outcomes have been achieved) whilst also contributing to enhanced understanding of what worked, what was less successful, and why. Rubrics also seems to be able to communicate all of this in the form of comprehensive, yet succinct tables. Our experience using the rubrics method, however, showed that it is far more difficult to apply in practice. Nonetheless, its value-add for supporting challenging projects – where goal-posts are often shifting and unforeseen opportunities and challenges continuously emerging – is also understated. In this Practice Note, I share the process of shaping this method into something that seems to be a right fit for the project (at the time of producing this note the project is ongoing and insights still emerging).

Project context

The experiences shared in this Practice Note come from the Food Systems Innovation (FSI) initiative. FSI is a partnership between the Department of Foreign Affairs and Trade (DFAT), the Commonwealth Science and Industrial Research Organisation (CSIRO), and the Australian Centre for International Agricultural Research (ACIAR). Launched in 2012, FSI's aim is to bring together Australian and international partners and expertise to help improve the impact of Australian-supported aid investments in agriculture and food systems in the Indo-Pacific region.

FSI is ambitious and complex. It brings together three Australian government agencies into a new mode of collaboration (i.e., partnership versus donor-client relationship) whilst also working closely with partners overseas and with global networks of food systems experts and other domain specialists. It aims to balance the planning and management of a detailed, collectively agreed-upon set of activities and associated outputs with flexibility to take on new opportunities and partner needs as they emerge. These, along with other aspects that make FSI a challenging project, have made it quite difficult for the FSI team to identify and develop effective monitoring, evaluation and learning (MEL) approaches and tools. When we came across the rubrics approach, it seemed to tick many of the boxes we needed: the ability to

monitor activities and outputs delivered (primarily quantitative information) and assess progress towards outcomes (largely qualitative evidence). An additional value-add was that it looked promising as a platform for fostering reflection and learning among the FSI team and partners. We have yet to test the latter but in this Practice Note we share our experiences-to-date using rubrics as a monitoring and evaluation tool.

So what are rubrics?

Rubrics – often called 'evaluative rubrics' – is a qualitative assessment tool. Rubrics involves articulating and clarifying 'the things that matter' in a project or initiative, which can encompass aspects related to the performance, quality, usefulness, and effectiveness of project activities, services or products. These are in essence the things that are considered by those involved in the project as important to pay attention to. These are evaluated using a qualitative rating system (e.g. excellent, good, adequate, poor). Rubrics typically look like a table or a matrix.

There are many readily-available rubrics guidelines and reports online. Some of the resources we used and found most useful are listed on the last page of this Practice Note.



Our experience with rubrics in FSI

We were first introduced to rubrics by an evaluation practitioner who was helping us reflect on why the approaches and tools we had developed were simply not working for us (an indicator-based framework and a home-grown narrative approach – see *Practice Note 'Journey to Fit for Purpose'* for more details). Rubrics was proposed as a middle-ground between the rigid, quantitative, accountability-oriented indicator M&E system, on the one hand and the fluid, qualitative-heavy, process-focused, reflection-oriented Learning Trajectory approach, on the other hand. Below we describe our journey with rubrics.

Sussing rubrics out

A small team of us (evaluation practitioner included) spent half a day brainstorming whether the rubrics approach was worth pursuing (see Figure 1 below). This was a critical decision as we had already invested a significant amount of time in the development of previous MEL tools only to find them failing. At this stage we were not only low in energy but also quite cognisant, and nervous, that FSI did not have a functioning, running MEL system to support the management and implementation teams.

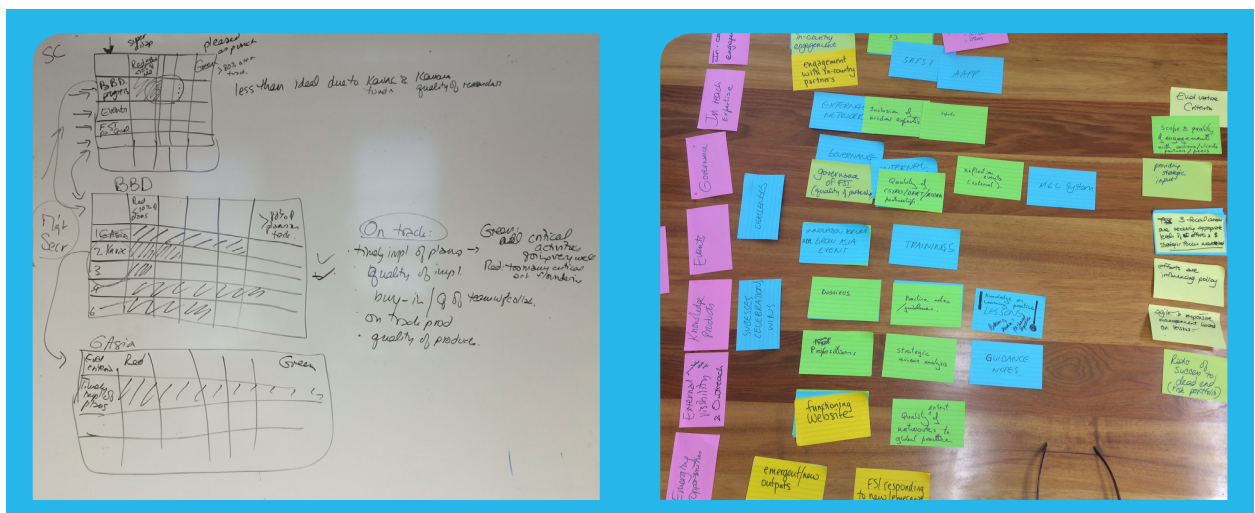
So how did we go about making the decision? In our discussions, three questions dominated:

- Would rubrics allow us to collate, in a non-cumbersome and rapid way, the evidence needed to demonstrate that we were delivering on what we promised? – *we thought yes*
- Equally critical, would it enable us to report on our outcomes which, up to that point, we were struggling to do? – *rubrics' focus on outcomes confirmed this for us*
- And, would the rubrics outputs (tables) be an effective way to communicate our achievements (as well as challenges and deviations from plans) to FSI's Steering Committee? – *again we thought yes*

The rough road to a first draft of rubrics

Identifying the evaluation criteria. We began by looking at various guidelines available online on how to develop rubrics. The first thing these told us was that we had to revisit FSI's programme logic, particularly the outcome statements which are at the heart of rubrics' evaluation process. For each outcome, rubrics requires that you clearly define the key qualities or changes that are deemed critical for your project's progression towards achieving its goals.

FIGURE 1 Brainstorming the potential value of using rubrics is a critical step. For us, this entailed scrawling on a blackboard a rough illustration of what it would look like and how it would support different groups within FSI's governance structure and revisiting the project's programme logic using coloured cards.



This sounds easy, but it's not. We found pinpointing which specific aspects of outcomes we wanted our evaluation to focus on an incredibly challenging task. It became quickly apparent to us that these would differ depending on the 'eye of the beholder' (i.e. each member of the FSI team and the Steering Committee would most likely define these differently). This is not a new insight – the guidelines on rubrics emphasize the importance of doing this using participatory, consensus-building processes. Our challenge was that we simply did not have the time to do this; we had to pursue the less optimal (but, in projects, typically more common) path of having to rely on a small set of team members – those responsible for FSI's MEL system – to define these. While far from ideal we did have more than a year's worth of working closely as a team and liaising frequently with Steering Committee

members. We did the best we could to ensure that the reworked outcome statements reflected (to the best of our knowledge) the aggregate perspectives of FSI members and partners. After many, many iterations we came up with a series of carefully worded outcome statements (re-labelled 'implementation and output performance statements' and 'primary outcome'; see Figure 2 below). These formed the basis for the next step – the evaluation criteria used in the rubrics tables.

We then created rubrics tables for each of the outcome statements (an example for 'learning events' is provided in Figure 3 on the next page). We thought the hardest part was behind us. Not so. We found that developing the ratings that would be used to assess the evaluation criteria was equally difficult.

FIGURE 2 The FSI initiative's outcome statements were rewritten in such a way that the specific qualities and changes being targeted become explicit (words/phrases underlined). These subsequently formed the evaluation criteria listed in the rubrics table (see further below).

ACTIVITIES AND OUTPUTS	IMPLEMENTATION AND OUTPUT PERFORMANCE STATEMENTS	PRIMARY OUTCOME
Learning events	<u>Relevant, timely, appropriately-designed</u> learning events that involve <u>Australian and/or in-country and partners and their networks</u> , and are perceived as <u>worthwhile</u>	Enhanced <u>knowledge-exchange; learning; and networking</u> among FSI partners and other stakeholders in Australia and overseas, thereby strengthening capacity to progress food systems innovation
Knowledge products	Knowledge products are <u>relevant, practice-based and practice-oriented, and aligned with current and emerging (ag, food and nutrition) international development thinking, practices and needs</u> within Australia and overseas; <u>collaboratively-created, reader-friendly and audience-appropriate</u> , and <u>produced and delivered in a timely manner</u>	
Expertise and practice networks	An <u>expanded range of relevant experts and development practitioners</u> who <u>actively contribute</u> to FSI Australian and in-country partners' and their networks' international development discussions, designs, and practices in ways that are perceived as <u>collaborative, salient, credible and useful</u>	
In-country engagement	FSI creates <u>opportunities</u> for in-country partners to <u>participate in reflection and learning in food systems innovation</u>	
Outreach and external visibility	Activities and a web-based platform that bring together analyses, opinions, lessons, <u>experiences and capacity building resources</u> derived from FSI and its Australian and regional partners; are <u>easily accessed and easy to navigate</u> ; and <u>audience-appropriate and reader-friendly</u> ; and <u>updated on a regular and in a timely fashion</u>	
FSI governance	FSI key partners (DFAT, ACIAR, CSIRO) <u>work together collaboratively, respecting the agreed-upon partnership principles, to collectively learn and manage FSI in a responsive and agile manner</u>	

FIGURE 3 Rubrics for FSI's Learning Events

Implementation and output performance statement: Learning events are relevant, timely, appropriately designed; involve Australian and/or in-country partners and their networks; are perceived as worthwhile; and are effective in progressing FSI's primary outcome (enhanced knowledge-exchange; learning; and networking)

EVALUATION CRITERIA	RATINGS					EVIDENCE
	Poor	Adequate	Good	Excellent	Insufficient evidence	
Relevant						
Timely						
Appropriately designed						
Involve Australian and in-country partners and networks						
Worthwhile						
Enhance knowledge-exchange						
Enhance learning						
Enhance networking						

Defining the ratings categories. Guidelines on rubrics provide very clear definitions of each of the rating categories (i.e., 'poor', 'adequate', 'good', 'excellent', 'insufficient evidence'). These are purposely generic which meant that we had to reflect on 'so what do these really mean for us in the context of FSI?' We came up with contextualised descriptions of the categories (see Figure 4 below). These were never formally included in any of the rubrics reports; rather they were used informally to help us pin-down what we thought we needed to do next: provide a description for *each* of the rating categories, for *every* evaluation criteria, across *all* outcome statements (an example of what this meant for us is provided below). This is where we made a costly mistake.

We spent weeks writing up the descriptions for ratings categories. Soon the rubrics became bigger than Ben Hur, taking up pages and pages of tables filled with minuscule

writing (an example is provided in Figure 5). For us, the proverbial 'straw that broke the camel's back' came with the feedback from team members: no-one could agree on what constituted a 'poor', 'adequate', 'good', and 'excellent' learning event in terms of its relevance, for example. This was the case for the majority of the other evaluation criteria across different outcome statements. We decided to do something we should have done at the very start: consult an evaluation practitioner who had extensive experience with rubrics. The first thing she told us (contrary to some of the online guidelines we had consulted!) was that we did not need to define the ratings categories. "Let the people who are attending the learning events (or involved in other FSI activities) tell you why they ticked 'poor', 'adequate', 'good' or 'excellent'" she said. She added that, like us, she had learned this lesson the hard way but that years of trialling different approaches to rubrics had demonstrated that the simpler method worked best.

FIGURE 4 The ratings categories used in rubrics, generic descriptions, and our FSI-contextualised descriptions

CATEGORIES	GENERIC DESCRIPTIONS	TRANSLATION (a.k.a. what this really meant for us within the context of FSI)
Poor	<ul style="list-style-type: none"> Performance is weak. Does not meet minimum expectations/requirements. 	<ul style="list-style-type: none"> What we don't want.
Adequate	<ul style="list-style-type: none"> Performance is reasonable. Still major gaps or weaknesses. 	<ul style="list-style-type: none"> Still not where we want to be; supporting status quo and unlikely to contribute to new ways of thinking and practice that will improve ODA impacts overseas.
Good	<ul style="list-style-type: none"> Performance is generally strong. Might have a few slight weaknesses but nothing serious and are being managed effectively. 	<ul style="list-style-type: none"> Exposing partners including in-country partners and their networks to new ideas and practices which hopefully we see being picked up in their policies, programs, projects.
Excellent	<ul style="list-style-type: none"> Performance is clearly very strong or exemplary. Any gaps or weaknesses are not significant and are managed effectively. 	<ul style="list-style-type: none"> The ideal; all conditions needed for maximising space for innovation (individual to organisational learning, etc. to systemic changes) are being enhanced.
Insufficient evidence	<ul style="list-style-type: none"> Evidence unavailable or of insufficient quality to determine performance. 	<ul style="list-style-type: none"> Flags areas we need to be collecting more evidence on and/or simply can't because of time-lags or other factors.

FIGURE 5 An example of our (erroneous) attempt to describe each ratings categories for each evaluation criteria

RATINGS					
EVALUATION CRITERIA	Poor Majority of learning events	Adequate Majority of learning events	Good Majority of learning events	Excellent Majority of learning events	Insufficient evidence
RELEVANT	<ul style="list-style-type: none"> Are on topics or issues that are irrelevant or misaligned with key partners' and learning event participants' interests or needs, and with global/regional/country-specific development priorities and strategies/programs/projects. Are relevant for Australia (Canberra-based) partners and not for in-country partners. 	<ul style="list-style-type: none"> Meet partners' and participants' mandated or technical needs (e.g. training in a particular M&E system). 	<ul style="list-style-type: none"> Are not focused only on technical issues but also on current or emerging areas of development thinking and practice that are aligned with global/regional/country-specific development priorities and strategies (e.g. Aid4Trade, private-public partnerships). 	<ul style="list-style-type: none"> Are on current and emerging areas of development thinking and practice that are aligned with global/regional/country-specific development priorities and strategies, and Make an explicit link to the specific development priorities and strategies/programs/projects/practices of participants' organisations and networks. 	<ul style="list-style-type: none"> Not enough evidence available to assess this objective.
TIMELY	<ul style="list-style-type: none"> Are out of sync (not timely) with partners' and participants' information and skills needs for enhancing their capacity to design and delivery effective strategies/programs/projects. 	<ul style="list-style-type: none"> Are 'hit-and-miss', i.e. some events are offered in a timely manner, whereby partners and participants can make use of the information; whereas others are out of sync. 	<ul style="list-style-type: none"> For the most part are offered at the right time, i.e. partners/participants are able to incorporate at least some of the information and learnings gained from the event to inform their strategies/programs/projects/practices. 	<ul style="list-style-type: none"> Are offered at the right/appropriate time, i.e. most of the information and learnings can be used to inform partners' and participants' strategies/programs/projects/practices. Include some development thinking/practices that are emerging or are on the cutting-edge (i.e. are not mainstreamed likely to be) that offers partners and participants a 'head start'/'competitive edge' in global development arena. 	<ul style="list-style-type: none"> Not enough evidence available to assess this objective.

Towards a more manageable and practical rubrics (where we are at now)

With all of the ratings category descriptions deleted, the rubrics tables became so much more user-friendly and useful. We were finally able to use them.

Gathering the evidence. We drew up a rubrics table for each learning event we held – i.e., for each workshop, training events, and reflection session (we also produced rubrics tables for the other types of activities and products FSI was engaging in/producing). Rather than show the rubrics tables to participants of the events, we put together a very simple 1-page 'survey' that we invited participants to fill out at the end of each event. Other sources of evidence that enabled us to complete the rubrics

tables included: reflections from FSI team members who had led or participated in the event; unsolicited e-mails from participants about the event; and follow-on activities and 'wins' that we could trace back to having been triggered or influenced by the event.

Reporting the rubrics. The report that emerged from the rubrics tables was comprised of three different layers of information (see Figures 6 and 7 on next page): (1) a series of individual tables for each activity and output (e.g. each learning event held), (2) a summary table that aggregates the data from the individual tables (e.g. one table that summarises all of the learning events held), and (3) a one page rubrics table with supporting narrative that distils the massive amount of information collated in the individual and aggregate tables.

FIGURE 6 FSI Rubrics approach: An example for Learning Events (fabricated data)

Implementation and output performance statement: Learning events are relevant, timely, appropriately designed; involve Australian and/or in-country partners and their networks; are perceived as worthwhile; and are effective in progressing FSI's primary outcome (enhanced knowledge-exchange; learning; and networking)

1. INDIVIDUAL RUBRICS FOR EACH EVENT

EVALUATION CRITERIA	RATINGS	EVIDENCE
Relevant	Excellent	<p>Feedback from participants who filled post-event evaluation form (N=13):</p> <ul style="list-style-type: none"> Aligned with my current work requirements or needs (85%) Knowledge I gained can be used to improve my work (77%) Structured in a way that supported my learning style (100%) Benefits of attending the seminar outweighed the time away from the office (85%) Will share the information I learnt at the seminar with colleagues (77%) I met people who have the potential to be valuable in my work (100%) <p>Selection of additional comments from participants:</p> <ul style="list-style-type: none"> "I learn some new things in this presentation. I will use some of the things learned to my work." "The information presented was very useful for me as a practitioner and researcher. The example gave me a clear idea on how important it is to consider these issues." <p>Other evidence of success:</p> <ul style="list-style-type: none"> Since the training, participant X and Y have re-designed their program to incorporate concepts and practices from the event Participant Z wrote the following, unsolicited e-mail: "I learned a lot from the event and was wondering if FSI will be offering a follow-on training course." <p>Sources of evidence:</p> <ul style="list-style-type: none"> Participants' responses to evaluation feedback form Reflections from FSI team member leading event Feedback from collaborators and FSI steering committee
Timely	Good	
Appropriately designed	Excellent	
Involve Australian and in-country partners and networks	Excellent	
Worthwhile	Excellent	
Enhance knowledge-exchange	Excellent	
Enhance learning	Excellent	
Enhance networking	Good	



2. SUMMARY RUBRICS OF ALL EVENTS

LEARNING EVENTS	EVALUATION CRITERIA & RATINGS							
	Relevant	Timely	Appropriately designed	Involve partners	Worthwhile	Enhance knowledge	Enhance learning	Enhance networking
Workshop 1	Good	Good	Excellent	Good	Excellent	Excellent	Excellent	Good
Seminar	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Excellent	Good
Lunch seminar	Excellent	Excellent	Good	Good	Excellent	Good	Insufficient information	Insufficient information
Reflection event	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Excellent	Good
Workshop 2	Excellent	Excellent	Good	Excellent	Excellent	Excellent	Excellent	Good
Training event 1	Excellent	Good	Excellent	Good	Good	Insufficient information	Insufficient information	Excellent
Training event 2	Good	Good	Good	Good	Good	Excellent	Good	Adequate
Presentation	Adequate	Good	Good	Adequate	Good	Insufficient information	Insufficient information	Insufficient information
OVERALL	Excellent	Good	Good	Good	Excellent	Excellent	Good	Good

- Excellent
- Good
- Adequate
- Insufficient information



3. SUMMARY REPORT

OVERALL RATING – GOOD

A series of training events on private-public partnerships and Theories of Change have been delivered and well received. There is a growing demand from in-country programs for similar events and 3 are being planned in the next quarter.

We found getting from the summary rubrics tables (#2 in Figure 6) to the 'summary report' (#3) a challenge. The practitioner and online guidelines we consulted recommended that a consultative, participatory process be used (e.g., team members and partners spend a day together looking at all the individual and summary rubrics tables and collectively write up the summary report). Again, for the same reasons previously mentioned (i.e., time constraints, logistical challenges in getting everyone together) we had to come up with a different process. In our case it was having the project leader look at the rubrics tables across all FSI activities, outputs, and outcomes and summarising them into a single overarching table that was accompanied by a succinct narrative statement; and it had to fit in 1 page (see Figure 7 on the next page).

FIGURE 7 Overarching summary report based on rubrics tables (fabricated data)

ACTIVITIES/ OUTPUTS	RATING	COMMENTARY, EVIDENCE AND PROSPECTS
Learning events	Good	A series of training events on private-public partnerships and Theories of Change have been delivered and well received. It was a slow start but there is a growing demand from in-country programs for similar events and 3 are being planned in the next quarter.
Expertise and practice networks	Adequate	External expertise has been used in the areas of M&E, partnership development and impact investing. Engaging more readily accessible regional expertise has been a challenge but tasks have been defined to strengthen these ties, with an immediate focus linking to the inclusive agribusiness innovation round table.
In-country engagement	Excellent	The project has been building strong alliances and collaborations with partner projects in East Africa, South Asia, Indonesia and Timor Leste. This is demonstrated by positive feedback and increasing demand for further activities from in-country partners. Two major events in South Asia are being planned that will further strengthen these relationships and enhance knowledge-exchange, learning and capacity-building.
Outreach and external visibility	Poor	The interactive web based platform developed for the project is not being used. In the coming quarter, the site will be substantially redesigned to reflect its greater outreach orientation and encourage a wider set of stakeholders to engage with and contribute material.
FSI Governance	Good	FSI has been effectively responding to numerous and frequent requests from FSI key partners and their overseas partners to collaborate. Partnership principles, as articulated in the Partnership Agreement, are being used to guide FSI governance processes. There is further work to be done to foster joint learning and ownership by all partners.

This one-page summary report is what we communicated to our Steering Committee. All of the other more detailed tables were included but (and this is important – we discuss why in the next section), they were filed away as appendices.

So what do we like about rubrics?

1. ITS MULTIPLE-FUNCTIONALITY

The main purpose of the rubrics is to enable evaluation of progress towards outcomes. It does this really well, but it also is quite effective as a monitoring tool for assessing delivery of activities and outputs. For example, the summary tables (table #2 in Figure 6) provide an easy way to list all of the activities and outputs that have been achieved. It thus works well for accountability purposes, which is a desired, demanded and important aspect of M&E systems. Furthermore, the process of rating the evaluation criteria essentially captures qualitative information. Supplementing the ‘tick-the-box’ (or, in our case, ‘colour the box’) ratings categories with written and verbal feedback in the ‘evidence’ column provides another layer of information: what was it about those particular

qualities or changes that were important? And why (or why not)? For us, these complimentary narrative-based sources of evidence enable the M&E system to have performance management capability, which opens the possibility of exploring the less tangible aspects and dynamics (e.g. knowledge brokering, trust building, co-ownership) that were facilitating (and, reversely, hindering) FSI to progress towards its goals.

2. NOT INDICATORS, BUT NOT TOO FAR OFF

Using indicators for measuring success is standard practice in many organisations and projects. So much so that it is arguably imperative that whatever MEL system you come up with has to have some capacity to ‘hold conversations’ with existing assessment structures formally being used by the organisations or groups you are working with. What we liked about the rubrics was that the evaluation criteria are essentially indicators but the process of rating these brings in additional information on the ‘how’, ‘why’, and ‘so what’? Even the most die-hard advocates of quantitative indicators acknowledge the importance of being able to capture this more qualitative, process-level information.

3. YOUR PICK OF PROCESSES TO CONSTRUCT RUBRICS

The rubrics approach is not wed to any singular approach to developing them. One can choose to take a bottom-up, participatory route or a more top-down one, or somewhere in the middle. This makes it quite a flexible approach that can be developed in a context-appropriate way, that is sensitive to time- and funding-constraints, and other issues (for example, conflicts among project/program participants which may make it undesirable or unrealistic to pursue a participatory approach to building rubrics). In our case, time-constraints and urgency dictated a less participatory approach to developing the rubrics.

4. INDUCTIVE, EMERGENT RATINGS PROCESS

The process of defining the outcome statements involves, to various degrees, bringing together the different perspectives and interests of those involved in a project or initiative. But perhaps one of the greatest (and understated) value-adds of the rubrics is not that it just integrates diverse viewpoints but that it can do so in an inductive, emergent way (i.e. as one gets more and more feedback, one starts to get a better picture of what 'excellent', 'good', and 'poor' means through the eyes of participants). This is particularly true with regards to the ratings. Rather than needing to define 'poor' upfront (which then throws you in the conundrum we faced), it becomes defined over time as one accumulates feedback from a range of people and perspectives (e.g. different team members, external partners, participants in project sponsored activities, etc.). This is not insignificant. Let me share an example. Had we used the predefined ratings categories we had first come up with and used those to assess the learning events, a good portion of them would have more or less bombed on the selected evaluation criteria. That is because our (FSI team's) assumptions about what comprised a "successful" learning event was heavily informed by our broader thinking underpinning the project: that to achieve innovation in food systems thinking and practices, our learning events had to go beyond the standard-of-the-mill stuff, i.e. training courses, seminars (in other words, transfer-of-knowledge stuff). We have since found, however, that participants who attended conventional learning events overall tended to rate them more positively than we would have. More importantly, the written comments that accompanied their ratings reveal a more nuanced, and complex, picture of the role of these types of learning events in contributing towards food systems innovation.

Having said this, it is not always easy to get participants or users to provide comments and thus help give meaning to the criteria evaluation. For example, we have not been able to set up a mechanism for capturing feedback – in a succinct, regular, and non-cumbersome manner – from readers of our reports, Practice Notes, and the dozens of other knowledge products we have produced. While this does not mean that we cannot use rubrics for assessing knowledge products, it does mean that our assessment is limited because it is being done through the eyes of project team members only.

5. SUCCINCT YET RICH – BRINGS SIMPLICITY TO THE COMPLEXITY OF PROJECTS WITHOUT BEING SIMPLISTIC

Finally, we liked the nested tables and what they provide in terms of information. They enable moving from the nitty gritty to broad level assessments. The one-page summary report that draws upon the detailed tables are perfect for the quick read on short flights between Brisbane and Canberra, or during commuting-to-work-time in Jakarta. For time-poor people with multiple competing demands (as most members serving project management and steering committees are) this is essential. The more detailed tables can be included as appendices, allowing those interested in more details to have access to the evidence base.

What have we found challenging?

1. THOSE *%! RATING CATEGORIES

We have already said enough on this issue. We highlight it again as our blunder of developing detailed descriptions for the ratings categories is apparently a very common mistake. And a very time consuming one.

2. THE WHOLE IS MORE THAN SIMPLY THE SUM OF ITS PARTS

Rubrics, because it forces you to break down specific attributes and changes you want to assess, it focuses your attention on different bits and pieces of a project. Don't get me wrong – this is very valuable (no one can argue that it is not important to know that a workshop you invested lots of money in and lost considerable sleep over, was great at helping people build networks and enhance knowledge). However, while rubrics are great for assessing individual or particular activities and outputs in isolation they are less effective in capturing the cumulative outcomes or impacts of multiple dimensions of a project, as it does not capture how they interact and what emerges as a result. Also the more components you have the more daunting becomes the task of pulling the data together into tables.

3. IT IS A LONG, TIME-CONSUMING PROCESS

It took us 10 months to get from initially assessing whether the rubrics approach was worth pursuing to developing a report. There are many reasons why it took so long. We were novices in this approach and this meant a steep learning curve and time-consuming mistakes along the way. We also aimed to cover over a year's worth of FSI activities and outputs; it was not a matter of developing a handful of rubrics tables but, rather, several dozens. Moreover, as highlighted in this Practice Note, rubrics is not just a table but rather involves a whole series of steps and processes. These began with refinement of the outcomes statements, to the development of different tools, to gathering and collating feedback

and other sources of evidence. Each of these bits and pieces that lead to the actual rubrics table takes considerable forethought and time. And finally (and perhaps most significant) was the fact that developing rubrics was only one of many, many other project activities that the team had to juggle at the same time. It was thus by necessity a piecemeal process that had to fit and be woven in and amongst other project activities, the majority of which were of greater priority. That said, now that we have the basic structure of the rubrics table and report along with the methods for collating supporting evidence, we anticipate that using the rubrics approach for assessing future activities and outputs can be done relatively quickly.

In summary: if we had to do it all over again, what would we have done differently?

RUBRIC STEPS AND PROCESSES	WHAT WE DID	WHAT WE WOULD HAVE DONE DIFFERENTLY (IN AN IDEAL WORLD)	WHY
<i>Refining the project outcome statements and defining the evaluation criteria</i>	Two team members worked together to reword FSI's outcome statements as defined in the programme logic, and in the process define the evaluation criteria.	A face-to-face participatory process whereby the whole FSI team, Steering Committee members and other key stakeholders jointly rework the outcome statements and define the evaluation criteria.	Other practitioners' experience using rubrics have found that having everyone jointly articulate and agree upon, in an explicit and transparent manner, the evaluation criteria (i.e. what matters, what is important to focus on) increases interest in and ownership of the rubrics process and ultimately, the assessments made.
<i>Developing the ratings</i>	We first defined, for each evaluation criterion, what each of the ratings category were (i.e. detailed descriptions of what would constitute an 'excellent', 'good', 'adequate', and 'poor' learning event along each of the criteria identified).	Leave the ratings undefined (what we ended up doing, once we realised our mistake in attempting to fill them out); allow them to become defined over time; feedback from participants/users accumulated.	<p>Defining the ratings is incredibly time-consuming and produces lengthy, unreadable (and ultimately unusable) rubrics tables.</p> <p>It is impossible to develop a set of ratings descriptions that will be agreed-upon by everyone (unless several days are spent, face-to-face, to jointly develop them).</p> <p>They are rigid: having a set of predetermined ratings descriptions does not allow for changes in how 'excellent', for example, may be perceived and understood by participants/users over the course of a project.</p> <p>Allowing participants/users to assign their own meaning provides a mechanism for different perspectives of what matters and why to emerge (arguably leading to more accurate and valid assessments).</p> <p>Other practitioners' experiences using undefined ratings has shown that it works just as well, if not better.</p>

RUBRIC STEPS AND PROCESSES	WHAT WE DID	WHAT WE WOULD HAVE DONE DIFFERENTLY (IN AN IDEAL WORLD)	WHY
<i>Gathering evidence</i>	The evidence for the majority of FSI's activities and products is FSI team members' personal assessments and judgement calls. The exception is the Learning Events for which we developed a concise survey to gather feedback from participants.	Develop other non-cumbersome, systematic methods for incorporating a broader set of opinions and perspectives for the other FSI activities.	Currently, the FSI assessments provided through the rubrics are very internally-focused. Different viewpoints from a broader range of people are required to get more diverse lines of evidence. This is needed for achieving a more transparent, shared, and valid assessment of how well a project is tracking.
<i>Communicating the rubrics</i>	First drafts of the rubrics-based report included in the main body of the report all three levels of rubrics information (individual tables, summary tables, and overarching summary report). We subsequently reduced the report to one-page, filing all the detailed tables as appendices.	Focus on developing the one-pager and co-develop it with the Steering Committee.	The look and feel of the report depends on your audience. In our case, our main audience has been the Steering Committee. Previous experiences with MEL reporting made it clear to us that whatever we produced had to be 'short and sweet'. The one-pager we produced (which is still in the works) was designed to meet the needs, and fit in with the busy schedules of the Steering Committee.

So, to rubrics or not to rubrics?

While we are still in experimenting with rubrics, we can share some emerging insights and words of wisdom.

1. GIVE YOURSELF PLENTY OF TIME TO PUT TOGETHER THE RUBRICS

As we learned the hard way, putting together the rubrics is not quick nor easy. It is a multi-step process, whose steps and time-frames vary depending on how well-developed the project programme logic is, the degree of consensus among project members and other key stakeholders about the particularities of the outcomes most critically in need of being tracked, and competing demands from other project activities and priorities.

2. TAKE AN EMERGENT APPROACH TO DEFINING THE RATINGS USED TO ASSESS THE EVALUATION CRITERIA

Let those providing feedback – be it event participants or team members – define what 'poor', 'adequate', 'good', and 'excellent' is. As a critical mass of assessments come back, look for patterns or common themes among people's responses.

3. EXPERIMENT WITH IT UNTIL IT WORKS FOR YOU

The generic approach to developing rubrics outlined in most guidelines may or may not work for your project. You need to allow some time to experiment with different versions.

4. DON'T GO AT IT BLINDLY – BUILD ON THE EXPERIENCES OF OTHERS USING RUBRICS

There are some fantastic resources out there, including people with extensive experience applying rubrics in different formats and varying contexts, and modifying them along the way. It is worth investing in connecting with others who have actually applied rubrics and building on their lessons learned.

5. BE CAUTIOUS ABOUT TURNING RUBRICS INTO A JACK-OF-ALL-TRADES MEL TOOL

The rubrics can seem very attractive to use as THE monitoring and evaluation tool. In all likelihood, it needs to be complemented with other approaches. For example, almost fortuitously, we found that a monthly e-mail listing detailing all of the activities and outputs that had been completed and were underway (with a few bullet point details) is, by far a, more effective monitoring tool. We have yet to trial the rubrics' value for learning-purposes, so the jury is still out on this one. But we think it has potential if it is *complemented with other tools* that have been developed explicitly for triggering reflection and learning. ■

Some useful resources

King, J., McKegg, K., Oakden, J. and Wehipeihana, N. (2013) *Rubrics: A Method for Surfacing Values and Improving the Credibility of Evaluation*. Journal of Multidisciplinary Evaluation 9 (21): 11. http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/374

Oakden, J. and Weenink, M. (2015) *What's on the rubric horizon: Taking stock of current practice and thinking about what is next*. Presented at ANZEA Conference, Auckland, 6 – 8 July 2015. <http://pragmatica.nz/wp-content/uploads/pragmatica-nz/sites/326/150718-Oakden-Weenink-ANZEA-vxx.pdf>

Oakden, J. (2013a) *Evaluative rubrics – passing fad or an important contribution to evaluation practice?* Presented at ANZEA Conference, Auckland, 6 – 8 July 2015. <http://www.anzea.org.nz/wp-content/uploads/2013/10/Judy-Oakden-2013-Evaluative-rubrics-fad-or-here-to-stay.pdf>

Oakden, J. (2013b) *Evaluation rubrics: how to ensure transparent and clear assessment that respects diverse lines of evidence*. BetterEvaluation, Melbourne, Victoria. <http://betterevaluation.org/sites/default/files/Evaluation%20rubrics.pdf>

Acknowledgements

The author would like to thank Irene Guijt and Judy Oakden for their constructive advice and invaluable discussions regarding monitoring, evaluation and learning approaches and the FSI initiative.

About the author

Dr Samantha Stone-Jovicich is an anthropologist with the Commonwealth Scientific and Industrial Research Organisation (CSIRO). Her research focuses on understanding why innovative and creative approaches (such as social learning approaches) that have emerged to address global social-environmental challenges are so difficult to implement effectively. Her primary interest is strengthening science's contribution to on-the-ground impacts by rethinking scientists' roles, practices, and the communication of scientific knowledge. Prior to joining CSIRO in 2006, she was an applied researcher working with small scale farmers and loggers in Brazil and other Latin American countries.

This series of Practice Notes facilitate shared learning and innovation to improve practice amongst research and development practitioners. The series is a deliverable of The Food Systems Innovation (FSI) Initiative, a partnership between the Australian Government Department of Foreign Affairs and Trade (DFAT), the Commonwealth Scientific and Industrial Research Organisation (CSIRO), the Australian Centre for International Agricultural Research (ACIAR) and the Australian International Food Security Research Centre (AIFSRC).

LIMESTONE AVENUE, CAMPBELL ACT 2601
PO BOX 225, DICKSON ACT 2606 AUSTRALIA
TEL. +61 2 6276 6621 • ABN 41 687 119 230

