
What's on the rubric horizon

*Taking stock of our current practice
and thinking about what is next*

Judy Oakden, Pragmatica Limited – a member of the Kinnect Group

Melissa Weenink, Ministry of Education

Presented at ANZEA Conference

Auckland

6 – 8 July 2015

Contents

01 How we got here

02 What is a rubric?

03 Reaching a broad description

04 Managing the tension – broad brush or detail?

05 Synthesising data

06 Summary

How we
got here

Setting the scene

Introduction

Overview

Evaluators in Aotearoa New Zealand are increasingly using rubrics in their evaluative practice. We now have a working knowledge about using rubrics and have some sense of what makes them more or less effective. While rubrics have shifted our evaluation practice, it has not been without challenges.

A rubric is supposed to express an evaluand's dimensions of value with sufficient precision to allow for clear evaluative judgements. Getting this clarity is hard. Some of the recent challenges we have seen are that:

- reaching a broad description that is also nuanced enough to encompass variations in implementation is a difficult balancing act (for instance, capturing key elements of a policy overall)

- managing the tension between representing something complex in an elegant and succinct way while also ensuring people feel that their view or values are represented is tricky
- synthesising a range of values or different perspectives into a rubric without it becoming unwieldy is hard.

In this e-book we will explore some of the challenges we have encountered using rubrics in our practice. We also include feedback from a discussion during a practice-based session at the ANZEA Conference in Auckland, New Zealand in 2015 where we explored difficulties we and others face with rubrics.

This document aims to provide more input as together we work out how we might steer our rubric boat to better navigate the choppy waters we sometimes encounter.

Introduction

Benefits of rubrics

Rubrics can be really useful right throughout the evaluation process.

At the ANZEA Conference in Hamilton, New Zealand in 2013, one of the authors of this e-book, Judy Oakden, gave a presentation entitled: *Rubrics passing fad or an important contribution to evaluation practice?*

At the time, she had discussed using rubrics with a number of evaluators – both those working as independent consultants and those working as internal evaluators within organisations. In general people were interested to engage in discussion about the use of rubrics. One of the internal evaluators particularly interested in the discussion was Melissa Weenink who is co-author of this document.

From those early discussions there was a general consensus that rubrics can be useful as they:

- help us frame the evaluation
- enable a flexible approach to evaluation in teams
- aid a shared understanding of what “good” looks like
- aid in mapping the data collection
- aid in the data synthesis process by providing a framework for making judgements
- can provide a useful reporting framework.

Introduction

Challenges of rubrics

However, we also observed there were a number of places where things could go wrong.

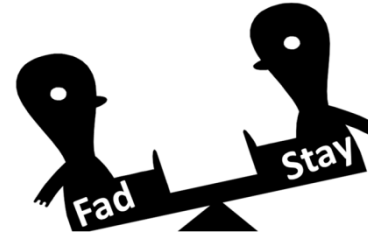
By 2013 both independent consultants and those evaluators working within agencies had observed that:

- rubrics were more difficult to use than we had initially expected
- they required a time investment from clients at the start
- rubrics could blow out, becoming big and unwieldy
- the synthesis was still difficult at times
- within organisations, rubrics had the potential to take on a life of their own beyond the original intended use.



Introduction

Rubrics - passing fad or here to stay?



During the presentation at the ANZEA Conference in Hamilton, New Zealand in 2013 Judy Oakden said:

“I’d put my money on rubrics being here to stay, but I think we need to understand the challenges of using them and mitigate against the risks.”

Judy considered that the benefits of using rubrics outweighed the challenges and she planned to continue to use rubrics in her own work.

She then went and shared that presentation with a number of government agencies in Wellington, New Zealand. Through these sessions she had the opportunity to discuss the challenges and opportunities that internal evaluators in a number of agencies were experiencing.

She also had ongoing discussions with Melissa Weenink who reflected on her experiences working with rubrics within New Zealand’s Ministry of Education. This document is a joint update on the progress we are both seeing two years later.

So two years on, what have we learned – where are the new directions in the use of rubrics?

What is a rubric?

Recap

What is a rubric?

Definition

Firstly, let's be specific as to what we are referring to when we talk about rubrics.

Evaluative rubrics offer a transparent process for making explicit the judgements in an evaluation and are used to judge the quality, the value or the importance of the programme or service being evaluated.

Evaluative rubrics include:

- **WHAT:** the aspects of performance the evaluation will focus on – that's what we are evaluating
- **HOW GOOD:** descriptions of what performance looks like at different levels
- **IS IT GOOD ENOUGH?** so we can determine how good the level of performance is and whether it is good enough.

With evaluative rubrics, the aim is to build layers of evidence till there is sufficient to make a judgement. They can be used for evaluating whole systems, for evaluating policy or programmes.

What is a rubric?

Example of a rubric

So that readers are clear of what we are talking about when we talk of evaluative rubrics, here is an example of ratings for a fictional party.

You can see the **performance ratings** range from “insufficient evidence” to “excellent” on a number of different **aspects of performance**:

- intended guests invited and attend
- there is a suitable venue
- there is appropriate food and drink
- people get the chance to socialise
- the guest of honour feels special.

So how do we make the judgements?

Aspects of performance	Performance ratings					
	Insufficient	Poor	Adequate	Good	Very good	Excellent
<i>Rating of the quality of the party – extent to which...</i>						
Overall rating				●		
The intended guests are invited and attend					●	
There is a suitable venue <i>(this includes parking, dance floor, kitchen, bathrooms, cloak room, near transport etc)</i>				●		
There is food and drink <i>(age and occasion appropriate)</i>		●				
People get the chance to socialise			●			
Guest of honour feels special			●			

What is a rubric?

Example: Rating a party

At this fictitious party, let's assume we had a great turnout – lots of people attended. It was held at a lovely venue and it was beautifully decorated, although there was not much parking available. So while it was very good, it was not excellent.

Most of the guests got to socialise and made the guest of honour feel special celebrating her 40th birthday.

Unfortunately three people came down with food poisoning during the event. One person became really sick and was taken by an ambulance to hospital later in the night. That person remained hospitalised for a week. Even though the party went well for many, this is what we'd call a “deal breaker”.

So while overall it might have been an “excellent” party for many of the party goers, we would only rate it “good” overall because of the people that were compromised.

Aspects of performance	Performance ratings					
	Insufficient	Poor	Adequate	Good	Very good	Excellent
<i>Rating of the quality of the party – extent to which...</i>						
Overall rating				●		
The intended guests are invited and attend					●	
There is a suitable venue (<i>this includes parking, dance floor, kitchen, bathrooms, cloak room, near transport etc</i>)				●		
There is food and drink (<i>age and occasion appropriate</i>)		●				
People get the chance to socialise			●			
Guest of honour feels special			●			

Reaching a broad description

Our learnings

Reaching a broad description

What is the challenge?

We are now going to talk about the changes we've seen in the last couple of years in the use of rubrics, including our own journeys.

But we'll keep using the party as the example for simplicity.

What are the key challenges we have noticed? All the ones we noticed earlier are continuing to occur. As a result, we have observed that people seem more cautious in the use of rubrics.

Over the last two years some new challenges have emerged as well.

Firstly, reaching a broad description that is also nuanced enough to encompass variations in implementation can be a difficult balancing act (for instance, capturing key elements of a policy overall) –but it is possible.

One thing we have noticed is that at times the rubrics were pitched at too low a level. By low level we mean that they may have described programme implementation in a very specific context without identifying the the general features which would allow you to make judgements across different implementation contexts. Then, they have the potential to be very repetitive. One solution we have found is to “lift them up higher” to a more principles-based level.

Reaching a broad description

How might we understand this?

We have developed an exercise called “the party” to help people understand that reaching a broad description which is also nuanced enough to encompass variations in implementation is possible and useful.

To demonstrate this, imagine we gather four groups of people in a room and get them to each design a party for a different age group or circumstance – child’s birthday party, retiree leaving party, a fortieth birthday for a girlfriend, and a wedding.

We have found that when the groups all think about their party it is possible to get quite different, rich descriptions of what good parties look like.



However, overall we have found that the critical and core dimensions for assessing parties are very similar, regardless of the age of the party-goers or the reason for the party.

Reaching a broad description

What have we learned?

People quickly get to see that at a broader level, all these parties have very similar criteria for success.

So, looking across the different variants of parties, they see the common features may be, for example:

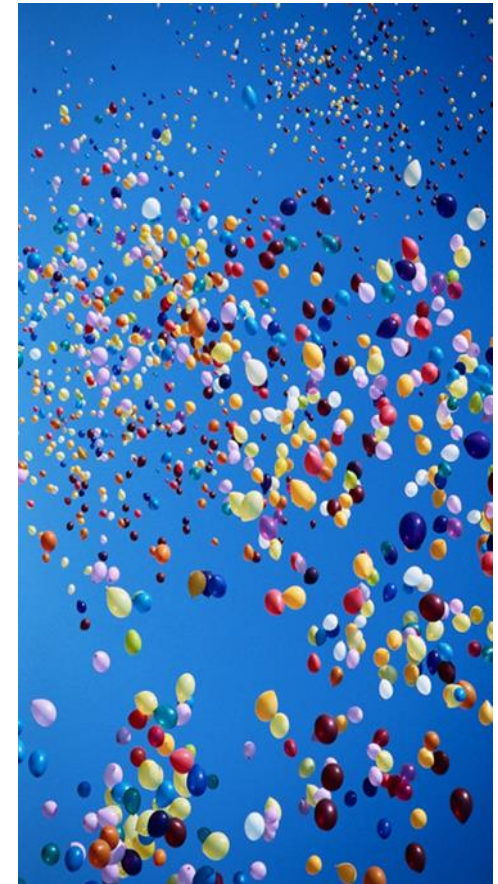
- the intended guests are invited and attend
- there is a suitable venue (this includes parking, dance floor, kitchen, bathrooms, cloak room, near transport, etc)
- there is food and drink (appropriate to age and occasion)
- people get the chance to socialise
- people make those at the centre of the occasion feel special.

We can see that if we were to be too detailed in our criteria

for a party, they might be fit for the purpose of a child's birthday party, for instance, but not the wedding – or vice versa. This would allow us less flexibility in the way we used the rubrics.

Therefore we are advocating that rubrics need to be specific enough that people associated with a programme can see themselves in them, but not so detailed that they will no longer be useful if the programme changes, or as the programme plays out in different contexts.

If the descriptions are too narrow and specific then its difficult to apply the criteria more broadly in different instances.



Managing the tension— broad brush or detail

**Making the rubric fit for
the purpose**

Managing the tension – broad brush or detail

What is the challenge?

In the last example we decided on the dimensions we were going to assess. But it still didn't give us measures of how good the performance was.

We have found it is challenging to manage the tension between representing something complex in an elegant and succinct way, while also ensuring people feel that their views or values are represented.

While it can be tricky to develop measures of how good the performance is, it is possible. Again if we take the birthday party example – in our last example we came up with some measures of aspects of performance for the party, but we didn't have a measure of how good the party was.



In the past there have been two approaches to measuring how good something is:

one is to try and capture a rich description of what poor to excellent looks like for all the dimensions,

and the other is to use a generic rubric for performance.

Managing the tension – broad brush or detail

Example: a rubric with descriptions of performance at different levels

Here the descriptions of performance for a party are set out for each level. A rich description is given of what performance looks like at each level: poor, good or excellent.

Using this approach we end up with a rubric that clearly illustrates levels of performance for a party on a number of criteria. This is often how people first start developing rubrics.

While this works well for a simple exercise, it can quickly become cumbersome for a large complex Government policy or programme.

Party Criteria	Poor	Good	Excellent
The intended guests are invited and attend	Only a few people turned up, and none of the important guests attended the event.	Half the guests turned up including all the most important guests, but not those from overseas.	Most of the guests turned up including all the important guests and many who have travelled from overseas.
There is food and drink (<i>age and occasion appropriate</i>)	The food for special diets did not arrive, and some of the food made party goers sick. There was generally not enough food. The drinks ran out after an hour.	There was a good selection of food and drink in plentiful quantity. Some of the food was a bit lukewarm, and looked a bit 'tired' but overall the food was tasty.	The food was delicious and the guests commented on this many times. There was plenty to drink, and party goers drank in moderation.
People get the chance to socialise	People did not mix well with each other. Two groups in particular did not get on and argued with each other.	People were standing in cliques, some met new people, but many stayed with those they already knew.	A warm atmosphere meant people introduced themselves to those they had not met before and had convivial conversations.
Guest make those at the centre of the occasion feel special	The guest of honour got 'bailed up' by someone they did not really want to spend a lot of time with.	The guest of honour moved from group to group and was generally made to feel special.	The guest of honour was made to feel special and was delighted to see guests talking to people they had not met before.

Managing the tension – broad brush or detail

Benefits of the detailed approach

Detailed rubrics help us capture a rich description of what “poor” to “excellent” looks like for all the dimensions of interest in the particular situation.

Detailed rubrics also tease out what good performance really looks like, and this can be valuable when evaluating a policy or programme.



Managing the tension – broad brush or detail

Example: a generic rubric

Another way to use rubrics is to apply a generic rubric: the evidence is taken and an assessment of performance is made based on a generic rubric that could apply to range of situations.

So for the party example, instead of deciding beforehand what “poor”, ‘good’ and “excellent” would look like, we might take the evidence and then make a judgement about it.

On the next page we show you how this is done.

Generic rubric	
Rating level	Generic description
Excellent (Always)	<ul style="list-style-type: none">• Clear example of exemplary performance or best practice in this domain: no weaknesses.
Very good (Almost always)	<ul style="list-style-type: none">• Very good to excellent performance on virtually all aspects; strong overall but not exemplary; no weaknesses of any real consequence.
Good (Mostly, with some exceptions)	<ul style="list-style-type: none">• Reasonably good performance overall; might have a few slight weaknesses, but nothing serious.
Adequate: (Sometimes, with quite a few exceptions)	<ul style="list-style-type: none">• Fair performance, some serious, but non-fatal weaknesses on a few aspects.
Poor: Never (Or occasionally with clear weakness evident)	<ul style="list-style-type: none">• Clear evidence of unsatisfactory functioning; serious weaknesses across the board on crucial aspects.
Insufficient evidence	<ul style="list-style-type: none">• Evidence unavailable or of insufficient quality to determine performance.

Managing the tension – broad brush or detail

Judging venue using generic rubrics

This approach would be useful for an events company to use if they run lots of parties and want to see how they are doing overall, as well as for assessing each party on its merits.



Poor
Clear evidence of unsatisfactory functioning; serious weaknesses across the board on crucial aspects.

Party A: What happened?


- Venue felt empty (was too big for the number who attended)
- No parking available – even for VIPs
- Waiter service was patchy
- Toilets flooded
- Belongings were stolen from cloak room
- There was very little transport home available including public transport – long wait for taxis.



Good
Reasonably good performance overall; might have a few slight weaknesses, but nothing serious.

Party B: What happened?

- Venue mostly the right size – the dancing area was a bit small, and there was a cold draft
- Parking for the guest of honour only
- Waiters were generally helpful
- Bathrooms checked once during the evening
- Cloakroom was secure
- Transport home was easily obtained – at party-goer's expense.



Very good
Very good to excellent performance on virtually all aspects; strong overall but not exemplary; no weaknesses of any real consequence.

Party C: What happened?

- Venue felt comfortably full and warm
- Parking was available for all guests
- Waiting staff was highly attentive
- Bathrooms regularly refreshed
- Secure cloak room and guests were given an identity tag
- Transport home for those not driving was included in the party fee.

Managing the tension – broad brush or detail

Benefits of the generic approach

The generic approach to rubrics is useful when we need to be flexible, or when context and circumstances might change from one year to the next.

Generic rubrics are also useful in situations where there is a lot of change in how programmes are delivered but not in the actual outcomes that organisations are aiming for. This is much like the

situation soldiers have found themselves in during war – they had direction (orders) about their overall objective but latitude in working out how they achieved it.

We believe developing evaluative criteria that are sufficiently generic can allow us to work across a broader system. Interestingly, in his recent writing on developmental evaluation Michael Quinn Patton



has also started talking about “guiding principles” that provide direction but are open to interpretation and can be adapted so as to work in different contexts. This is also akin to Glenda Eoyang’s “simple rules” as a way for building cohesion.

•

Synthesising data

Our learnings

Synthesising data

What is the challenge?

The final area we want to discuss is the challenge of synthesising a range of values or different perspectives into a rubric without it becoming unwieldy.

- Do we treat all aspects the same or treat some as more important than others?
- Are some things “deal breakers”?
- Who says what’s more important?
- And how do we actually do the analysis?



Synthesising data

Some criteria **are** more important

There are usually some criteria that are more important than others.

All criteria are not of equal weight; some are more important than others. How do we capture this in our evaluation? There are a number of ways the relative importance of criteria can be arrived at – the literature, stakeholder groups, experts or user groups. Scriven recommends that a graded scale for weighting the importance of evaluative criteria is used (qualitative weight and sum), and recommends a range of five different weights be applied (1991, pp. 293-295).

To go back to the party analogy, it will not be a success if no-one turns up. Also if there are instances of food poisoning, someone gets hurt or there are other adverse events, the party won't be judged a success. In other words some things are just essential.

But there are other aspects, like lovely flowers or festive decorations, that make a party hum and give it atmosphere but aren't "deal breakers".



All criteria are not created equal – if we treat them as equal this can lead to judgements that don't give us the best results.

Synthesising data

Rubrics evolve during an evaluation

The proof of a rubric's effectiveness is made evident when it is used to make evaluative judgements at the end of the evaluation.

In the past we said, it's not until you have actually used the rubrics to make evaluative judgements that you know how well they work.

Being able to think through to how you will use evidence to make judgements is important. Keeping the line of sight between a rubric

and its eventual use to make evaluative judgements is critical.

Indeed we have found that at times we need to tweak the wording to ensure they capture important aspects appropriately and that nothing is missing.

Using our party example, it might turn out that adequate seating is important, and if we do have not place where we capture this, it will need to be added.



Rubrics may need to evolve over time. It's ok to tweak the wording or add extra criteria if the data supports this.

Summary

Key learnings

Summary

Using rubrics still of benefit



This leaves us to sum up our learnings to date about rubrics.

We still believe that using rubrics (that is evaluative criteria and some form of performance framework) is beneficial to the quality of our evaluations. We believe there is more benefit when we keep the evaluative criteria quite general – at the principles level – think of the party example.

Indeed, change is so constant that we have found evaluative criteria have to be flexible to retain their currency on projects and to be able to cope with how programmes play out in different implementation contexts.

Summary

Care in use of rubrics

Rubrics that are used as **guides to performance** are different to those used by evaluators to assess the **level of performance**.

There is a risk that rubrics used as guides to performance by service users become unwieldy and very complex, which will ultimately lead to a loss of utility. This is something we need to guard against.

While everyone wants to see themselves in the levels of performance, this can be quite a tricky aspect to navigate.



Our advice is when a rubric starts to get too detailed, it starts to lose its power to be useful longer-term to make evaluative judgements.

Feedback

ANZEA Conference 2015
Participant feedback

Feedback from participants

Ongoing challenges

These comments and observations were made by participants after our presentation, following a 10-minute discussion with other attendees.

You need to be careful how you name your performance levels – you don't just want pass and fail. You need to develop some performance measures in your rubrics.

Rubrics can be really useful for building evaluation capacity, particularly with schools. A school is developing a programme and saw the use of rubrics as quite powerful. But we also acknowledge that in education settings [rubrics] can be useful in providing clarity around expectations as well as for monitoring and evaluation.

Thinking of analysis of the data – we were in different places in terms of... the ability to do the assessment and form an opinion overall

We talked about the time it takes to do rubrics – 15 minutes to half an hour on your own and several months with a group!

We acknowledged the range of experience - that some of us haven't used rubrics, while some of us using them in multiple settings at once.

It's important to manage the interactions of the stakeholders at the table. Rubrics need to be broad enough to capture the key aspects but neither *too* broad nor so detailed you get lost in the detail

Feedback from participants

Ongoing challenges

We acknowledged that there is a space for developing rubrics *for* and *with* stakeholders. We observed how they look different and that there is a challenge this brings for some of the stakeholders we are working with as well as for ourselves.

Rubrics look like they might make work simpler but they might really not. Rubrics don't *make* things more complex but they can *be* really complex.

In this group we talked of our experience of rubrics being used to operationalise performance – as opposed to evaluating the quality or worth of something.

There was a sense: if you are not struggling with them, you are probably not using them right!

Question: Reflecting on the analogy of the party – my daughter would have a wonderful experience at her party, but for me it is a complete headache. So where does perspective fit in rubrics?

Answer: Yes, perspective does come into it – whose values you judge the party by. Often this is the tension between looking from the perspective of the user and the provider in our work. Sometimes there is a level of agreement. At times you may have a separate section in the rubric for organisers and party-goers. Or, possibly, the level of agreement between users experiences and providers perceptions of success is a possible criteria itself.

References

Further reading

References

Further reading

Publications

Davidson, E. J. (2012). *Actionable evaluation basics: getting succinct answers to the most important questions [minibook]*. Auckland: Real Evaluation Ltd

Davidson, J. (2014). *Evaluative Reasoning*. Florence: UNICEF..

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 Vol 9 No 21 pgs11–20 http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/374/373

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

Oakden, J. (2013) *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>

Scriven, M. (1991). *Evaluation thesaurus (Fourth ed.)* (4th ed.). Newbury Park: Sage.

Other rubric examples

Ministry of Education (2013) *Measuring and reporting progress*. Wellington: Ministry of Education.
<http://www.minedu.govt.nz/theMinistry/PolicyAndStrategy/KaHikitia/MeasuringandReportingProgress.aspx>

New Zealand Qualifications Authority. (2009). *Policy and guidelines for the conduct of external evaluation and review*. Wellington: New Zealand Qualifications Authority.

Thank you



Contacts

Melissa Weenink

melissaweenink@gmail.com

Judy Oakden

Pragmatica Limited – member of
the Kinnect Group

judyoakden@gmail.com

www.pragmatica.nz

www.kinnect.co.nz