IOD PARC

Learning to make space for qualitative indicators

by //
Joseph Barnes

October 2010

Confounded by qualitative indicators?

A recent study undertaken by Masters students from London School of Economics and Political Science investigated the potential of The Full Frame Initiative framework to be used as a predictor of organisational impact (Kalashnikova et al 2010). The study considered three case studies of international development agencies of various sizes. This was a major extension of the Full Frame concept, which consists of a number of organisational principles, such as 'life is messy'. These principles were originally derived by looking at high-performing human services voluntary organisations in the US. These principles were translated into qualitative indicators for the purpose of the research (Ibid).

Whilst the approach clearly holds great promise, the researchers nevertheless found themselves confounded by one of the major challenges that face the real world use of qualitative indicators. Namely, with the amount of information realistically available to them, it was enormously challenging to differentiate organisations meaningfully on the range of scoring that researchers were able to use. The most noticeable trend was that even very different organisations ended up clustering around the 'safe' middle scores rather than the 'extremely bad' or 'extremely good' ends of the rating scale.

As with many qualitative indicators, the approach taken by this particular research was to 'quantify' a qualitative observation by using a rating scale. Rating scales create a range of scores that aim to map a set of possible scores that can be aggregated and compared. A simple example is the 1-5 scores that can be used to rank how much people like a song or photograph. This can be improved by creating a range that has no natural 'middle' score, e.g. 1-10, thereby forcing reviewers away from the middle ground. There are many other adaptations to this approach, for example the Qualitative Impact Assessment methodology used in the water sector in India employs a scale from 1-100 that defines each quintile (0, 20, 40, 60, 80, 100) but leaves room for reviewers to select intermediate scores where reality falls between the descriptions given.

All of these approaches, however, suffer from similar issues in their implementation. Well known challenges include problems with reliability and validity that come from unrepresentative samples of reviewers who each apply their own unique understanding of what a particular score represents. A lesser-discussed issue is that a limited 'linear' range of possible scores (i.e. with the 'difference' between a score of 1 and 2 representing a similar sized improvement to the 'difference' between a score of 2 and 3) mismatches with the 'bell curve' distribution of scores that might be expected from any analysis. As a result, the bulk of ratings will be given around the middle scores, making it harder to disaggregate between two 'normal' examples than between two 'extreme' examples at either end of the scale.

Both of these problems can be resolved to some extent by using large sample sizes. For example, Amazon and Apple use large numbers of ratings from their users to offset inconsistencies in the way in which each reviewer interprets what 'five stars' represents. Similarly, IOD PARC runs a qualitative rating system on behalf of UNICEF to assess the quality of its evaluation reports. The hundreds of reports that have passed through this system have arguably allowed successive reviewers to make full use of the full range of 1-5 scores and to thus disaggregate the range of quality in the reports in a more nuanced way. So even though each score is carefully defined by UNICEF, the process of applying these to hundreds of real life reports has enabled reviewers to adapt these in ways that are more relevant and useful.

Challenges in the real world

Unfortunately, large numbers of reviewers and/or samples are rarely available in the case of performance management monitoring. This makes the accurate use of qualitative indicators more challenging, but by no means impossible. For example, during the recent round of quality controlling UNICEF's global evaluation reports, IOD PARC employed multiple checks and balances to ensure that the small team of reviewers were 'all on the same page'. These included sampled peer reviews, working group discussions and investigation of 'outlier' scores. The reliance of qualitative indicators on these non-standard processes means that much more information needs to be included in a monitoring system for a qualitative indicator than for its quantitative counterpart.

A simple example of this can be illustrated by comparing a common quantitative education indicator (gross enrolment), with a common qualitative education indicator (parents' perception of school performance). Gross enrolment includes simple cardinal baselines and targets (sex disaggregated). We might refer to the Ministry of Education management information system, or school register as the source for this data. But we do not need to specify any more detail because the process for obtaining gross enrolment is well known. This is not the case, however, with our qualitative example. In terms of perception, if we wish to be 'robust', then

we need to specify who will be asked, how, when, who will ask, who will interpret the findings, how this will be analysed, how the findings will be presented, what 'satisfactory' means, and so on.

Not only can this required level of detail appear to be daunting at the project design phase, the logframe format (as used by many donors at the commissioning stage) is simply not designed to contain so much information. As a consequence, qualitative indicators can appear to be vague and rather forlorn in the logframe: leading some users to shy away from taking the qualitative approach even where they may wish to do so. This experience, which we see repeated over and over again, would suggest that two strategies within an organisation's project design phase could ease the pathway for qualitative indicators; and open up the valuable extra dimension to monitoring that they offer.

Two strategies for opening up performance management to qualitative indicators

The first of these strategies is more early-on investment in performance management system design, such as the support IOD PARC has provided to [AusAID Indonesia](). The detailed design work on performance management systems only tends to take place once a project (including logframe) has been signed off. This is for obvious reasons, such as availability of money. However, as a consequence, it is often only after things have already started that it becomes clear whether the qualitative indicators in the logframe are feasible or whether they can be integrated into other management tasks. At worst, it means that activities get implemented and difficult-to-answer qualitative indicators get forgotten about until it is too late to make them useful.

Resolving this issue through up-front investment in performance management system design opens up the possibility of more effective use of qualitative indicators. It also enables greater confidence in the use of qualitative indicators in the logframe, by allowing drafters to refer to sources of data that are already well understood and well documented by both the donor and the project team. At first this may require additional funding and expertise being made available to projects at the design phase (perhaps as an advance). However, as more familiarity with a greater resource-bank of qualitative indicators grows, we are likely to be able to reduce the level of such support.

The second strategy is to institutionalise the fundamental message of the Managing for Development Results (MfDR) agenda: that results-based performance management system design should move away from emphasising a project's ability to 'predict' a future state and towards an emphasis on successfully adapting to changing conditions (McAllister, forthcoming). Qualitative indicators are powerful tools in adaptive management strategies: because the meaning and interpretation of indicators can be tuned into the context. This means that we move away from holding project designers to account for their ability to foresee numbers and towards holding them to account for keeping their project relevant to the environment they serve.

In practical terms, embracing the adaptive management vein of MfDR thinking opens the door to creative performance management systems that reflect the real world rather than a predicted world. These will support good managers to achieve real-life outcomes. A prime example of this can be illustrated by the use of 'coding' in qualitative indicators.

Coding is a method of processing qualitative data in order to make it possible to apply quantitative tools. As a simple example, feedback forms from a workshop can be processed to determine how many positive and negative statements they contain. The 'codes' are the words (or combinations of words) that are chosen to indicate a positive or negative statement. In this scenario, the index of 'codes' chosen is crucial to the final score that a feedback form receives.

There are two main approaches to applying these code indexes for use in qualitative assessment. The first is to use as set of predetermined codes based upon research and experience. This is likely to be the norm in a traditional performance management system because it means that at the beginning of the project we already know how we will assess the evaluation data generated at the end. As a result, we can make sure that our baseline analysis and final evaluation are consistent even though they are many years apart. In many cases this will be useful for management and open up areas of information that had not previously been incorporated into decision making. In practice, such coding is also often improved over time, through both experience and specific research.

There is, however, another approach to coding: emergent coding. Emergent, or inductive, coding looks for themes and words that stand out in a sample of qualitative texts. Emergent code indexes therefore adapt as the contents of the source material evolves over time. The benefit of emergent approaches to coding means that management teams get a real insight in the whole picture of the world in which they operate.

One way of implementing such an approach in a development project could be to transcribe interviews with key stakeholders about the performance of a particular organisation or function. These transcripts could then be combined and run through a free tool such as [Wordle] to create a word cloud. The word cloud can be used to pick out which words related to the indicator appear most often. Over time this process can be repeated and the change in the words used by stakeholders (and their frequency) can be tracked to give an indication of progress over time.

Fig. 1: Word cloud based on this article

Unlike 'traditional' indicators, such an emergent approach does not allow us to predict a numerically-comparable baseline and target. But, in 'difficult to measure' areas, such as influencing value systems, these creative qualitative indicators have the potential to provide deep insight. And, rather than holding project management to account for their ability to predict numbers we can assess whether, at the end of the project, people are expressing the type of values that it was the intention of the project to bring about.

These two strategies of investing in performance systems early on and embracing emergent indicators would both require significant political will and space from donors. However, they are fundamentally achievable within the boundaries of existing programme cycles and tools. And, perhaps, the prize of making qualitative indicators more accessible, trusted and useful means that it would be a worthwhile investment to make.