

Case study: Using a modified evaluation methodology for assessing the performance of contract service providers

Greg Davies & Chris Milne, ARTD Pty Ltd

Paper presented to the Australian Evaluation Society International Conference, Darwin, September 2006

1. Introduction

Youth homelessness was a significant and high profile problem in Australia in the 1990s (Burdekin Report). Complex social problems often require complex solutions, and youth homelessness is one such example. Research at the time indicated that early intervention was most effective, either prior to the young person leaving home, or in the initial first few weeks. It showed the importance of a family focus, working with both the young person and the family to improve the potential for reconciliation. It showed that collaboration was needed with multiple agencies in the local community (such as schools, state care and protection agencies, youth services, counselling services, Centrelink) both to identify young people at risk of leaving home, and in providing support. It showed that interventions needed to be developed and refined over time to suit the local need as other services gained greater capacity to deal with the problem.

The Australian Government established the Reconnect program in 1999 to reduce youth homelessness through family- focussed early intervention, building on this research and experience of a two-year pilot program. Over the next three years 98 service providers around Australia were funded to deliver the program to communities ranging from inner city metropolitan areas to remote sites in indigenous communities in the Northern Territory and Western Australia. The service providers, ranging from small community committees to large NGOs, were selected through a competitive tendering process with an initial funding agreement for three years.

An important question became how to assess their delivery of services when there was a large amount of variation between the services in the context of the target populations, the extent of other services in the region, the maturity of the organisation, and the methods used for community work and for intervention. The approach had to be independent of the services and of FACSIA, be credible with all stakeholders and utilise the existing data and reports . An audit approach was not appropriate as it relied on relatively standard definitions of service, and Reconnect was a new program that also expected services to unfold their methods to meet the needs of their communities. An audit approach could also favour the services that were effective in documenting their work rather than delivering it. On the other hand, an external evaluation of each service that could take account of the context and different approaches was prohibitively expensive and could be a substantial burden on services.

This paper presents an overview of the approach that we developed, an independent performance assessment using evaluation methods that built on the available data and reporting from the services, interviewed stakeholders to corroborate the services claims, and made judgements about performance taking into account the service's context and stage of development.

2. The program

The program was therefore developed around five Service Requirements

- Sound management including planning, stakeholder involvement, evaluation and reporting
- Improved coordination between agencies
- Effective early intervention with a family focus

- Application of learning from the experience of the service through a structured process of Action Research
- Effective links to Centrelink in assessments for young people seeking income support

The program also included ‘Good Practice Principles’ that articulated the general approach to delivering services. These included:

- Provision of accessible services
- Client driven service delivery
- Holistic approaches to service delivery
- Working collaboratively
- Culturally and contextually appropriate service delivery

Importantly, the program did not set specific targets, but rather was outcome focussed, indicating that services should be tailored to suit local need. The program also placed a high emphasis on using a ‘tool box’ of interventions for young people, so that services were expected to provide a range of different supports to individuals.

The program was rolled out starting in 1999 in three annual rounds of funding through competitive tender – the first round in 1999 included many service providers had also been pilot sites, the second round in 2000 and the third round in 2001. The third round was tendered on the basis of ‘innovative’ approaches and interventions for youth homelessness and allowed for a broader interpretation of the program guidelines.

Services received on average \$200,000 per year in funding, generally sufficient for two staff, a part-time manager and expenses (rent, cars, phones). Some small amounts of additional funding were available for specific projects, such as Action Research projects targeting issues that were common to a number of services.

Services varied significantly in the support they provided to young people – some focused on working with young people in groups, some focused on individual casework and others combined the two. Young people were supported for shorter or longer periods of time, and the intervention may have been focused on family counselling, mediation or practical support such as linking young people to education or vocational support. The numbers of clients for each service varied widely – in 2005 (Round 2), services commenced an average of 132 clients in the two years prior to the assessment.

3. Evaluation and contract management

Services were required to report progress at a number of levels as part of the contract for service delivery. Details of each client were recorded at intake and when the client left the service. Key details on exit were length of time with the service, the nature of support provided, how the relationship ended (e.g. ‘no contact for more than 60 days’; or ‘referred and no further support’).

Services prepared annual self-evaluation reports, demonstrating performance against the guidelines. Agency staff in each state office reviewed the self-evaluation reports and provided comments to the service. Services also prepared annual workplans. As for many programs, the quality of the self-evaluation reports and workplans was variable and the capacity of the state office to provide timely feedback was also variable.

Services were funded for a three year period. During the third year of each cycle, the funding agency (Australian Government Department of Family and Community Services, now Families, Community Services and Indigenous Affairs, or FACSIA) needed to decide whether to refund the service provider or retender the contract.

ARTD performed an independent assessment of the performance of each service provider on a three yearly cycle.

At a broader level, FACSIA separately commissioned longitudinal studies of a sample of clients and a program evaluation to monitor the effectiveness of the overall program. This is important for framing our performance assessments – the program evaluation and longitudinal studies assessed whether the program was working as intended and if it was meeting the needs of the young people. Our performance assessments were to determine if the service providers were meeting their contract requirements.

4. The method for the independent assessments

Our approach relies on an assessment of the performance of a service provider against the program requirements, drawing on data from the previous two years. The categories are (Table 1):

Table 1: Performance ratings for Reconnect Service Requirements

<i>Rating</i>	<i>Basis</i>	<i>Required FACSIA response</i>
A	Fully effective - evidence is consistent with effective performance in this area. Minor concerns can be dealt with internally as part of ongoing service management.	Low risk – normal management
B	Reasonably effective - moderate concerns exist about the performance in this area, either because of shortfalls in performance or limited or inconsistent evidence. The service should be undertaking corrective actions as part of its ongoing work.	Moderate risk - the Department should expect evidence of improvements within an agreed period.
C	Significant concerns exist about the performance in this area. There is little or no evidence that a service requirement is being met or a practice principle being applied, or performance is contrary to a practice principle. Effective corrective action is not being undertaken.	Significant risk - the Department should take corrective action, eg. develop an action plan with the service to undertake improvements.

Performance was assessed against each of the five service requirements. A performance report for each service was produced using the five service requirements as the chapters, which included the Good Practice Principles within the relevant service requirement, and an opening chapter outlining the key factors which influenced the service in its context (e.g. urban or rural; small stand alone service or a service unit of a larger organisation; previous history as a youth service or family service).

We integrated reporting against the Good Practice Principles into an equivalent Service Requirement. Table 2 outlines the framework for each report.

Table 2 - The assessment framework in each Report

<i>Service Assessment Report</i>	<i>Service requirements</i>	<i>Good Practice Principles</i>
1. The service in its context		
2. Management	Use sound management practices	6. Ongoing review and evaluation
3. Improved coordination	Improve coordination of services delivered by government and the community sector;	4. working collaboratively 7. building sustainability
4. Family focussed early intervention strategies	Use family focussed early intervention strategies to reconnect young people with family, education, training, employment and the community;	1. Provision of accessible services 2. client driven service delivery 3. holistic approaches to service delivery 5. culturally & contextually appropriate service delivery

5. Working with Centrelink	Assist Centrelink by contributing to assessments for young people seeking income support.	4. working collaboratively 7. building sustainability
6. Action Research	Action Research	

Our reports consider the risk implications for future performance and suggest the level of action the Service and FACSIA should take. The ‘risk’ orientation of the reports is an important element of the design. The process does not attempt to accurately capture absolute performance – e.g. an A rating is low risk, meaning that there is evidence to indicate that the service is likely to continue to provide adequate service to its region in future years. It does not necessarily mean that a service is demonstrating best practice. Similarly, a number of ‘C’ ratings indicates poor performance but does not automatically indicate that a service should be defunded – rather it is intended to indicate to the funding body that there are risks that require attention – the funding body has discretion about how to proceed and the intervention that it is required.

5. The process

Our approach involved a sequence of steps for each service provider:

1. *Issues Paper* – we drafted an Issues Paper for each service based on the self-evaluation report, work plan, and an initial interview with the primary contact in the state office of the funding agency.
2. *Comment on the Issues Paper* - we forwarded the issues paper to the service asking the service to comment and to provide further information.
3. *Initial risk assessment* – we estimated the likely risk rating for each service based on the preliminary information. Higher risk services received a face to face visit because for these services, more effort or time was required to discover what was happening. Further, if a service ultimately received a ‘C’ rating, a face to face visit provided the service with a sufficient chance to tell their story. FACSIA also required that all indigenous services be visited, as this was the most culturally appropriate approach to the performance assessments.
4. *Interview* - We interviewed to confirm issues identified in the issues paper, fill in gaps and ask for documentation to substantiate the statements made in the self-evaluation reports. Preparing the issues paper prior to the interview had multiple advantages – it demonstrated that the self-evaluation reports were being utilised in a meaningful way, provided a clear focus for the interview, and allowed us to clarify any potential misunderstandings that may have arisen from the interpretation of the self-evaluation reports.
5. *Interview stakeholders* – in the service interview, we asked for a range of contacts in schools, Centrelink, youth refuges, child welfare agencies, supported accommodation providers and other youth programs (such as JPET – an employment program). Key people for this program are ‘first to know’ agencies who are the people who may first become aware that a young person was at risk of leaving home, such as schools. Stakeholders were asked to comment on the performance of the service in the key program areas – particularly improved coordination (referrals, accessibility) and family focused early intervention (what did the services do in their interventions with the young people).
6. *Draft report* – the draft report weighed up the evidence to make initial evaluative judgements about the performance of the service within its context and allocated draft ratings.
7. *Calibration* – In each funding round, there were up to 40 services. We included a step to calibrate the draft ratings across each round to ensure that ratings were equivalent for similar levels of performance, allowing for the different contexts.

8. *Service comment* – We forwarded the draft report to the Service to comment, primarily to confirm that we had represented the Service accurately. Importantly, this was **not** a negotiation about the ratings unless there was new evidence or the evidence had been interpreted incorrectly.
9. *Draft report to funding body* – We forwarded all the draft reports to the funding body for a second check on the calibration and accuracy. If services disagreed with the ratings in the draft report, we included a response from the Service (as a ‘Service Response’).
10. *Final report* – The final report included our response to comments from the funding body. If the evidence from the funding body led to a change in ratings, the service was notified before the final report was submitted. The funding body then negotiated with the Service about the future contract and what changes in service management may be required.

As a separate step, we provided a summary overview of the performance of the whole funding round compared to previous years, and noted thematic issues which had arisen in the course of the assessments, such as concerns about accuracy of client data or issues arising for some services in growth areas where the population (and hence the demand for the service) was increasing. These thematic issues were only observations and were not framed as judgements of the overall effectiveness of the program.

6. Why characterise this as a ‘modified evaluation method’?

The method for assessing performance for this program needs to allow for the different contexts and interventions – e.g. there is a very wide spread in the numbers of new clients for each service. There are many factors to weigh up in interpreting these numbers – how many clients were in groups; how long were the clients with the service; what was the nature of the intervention. This is a circumstance where evaluative judgements provide a good basis for determining whether the service is doing enough work, or the right kind of work for that environment.

Our method employs triangulation to provide the evidence for these judgements from a range of sources: the service reports and information from the interview; information from the funding body as the contract manager; client data; and interviews with key stakeholders.

A comprehensive evaluation of each service would attempt to collect evidence from the service users. This method does not do this. Where it was available, we used existing data collected by services (such as exit interviews or satisfaction surveys) where the methodology was sound. Otherwise collecting client data was not feasible. If a service was clearly under-performing, the data offered little added value. For other cases, it could have intruded on case work; raised problems getting an unbiased sample and taken considerable time to gain consent and then to reach the clients. The broader question of whether the program was meeting the needs of young people was addressed through the longitudinal studies and program evaluation.

Other possible approaches to assessing performance could have been:

- *performance auditing* – whilst our method is similar to performance auditing, audits are typically based on agreed standards or targets. In this case, the varied contexts and flexible program guidelines would not be well suited to an auditing approach
- *risk assessment* – a strict risk assessment approach would analyse the systems within each service (e.g. were management systems adequate; were there service delivery standards in place). Given the diversity of the services involved in this program, this would have been problematic – e.g. indigenous services in remote sites could not reasonably be expected to perform adequately against these kinds of standards. We used systems as evidence in some cases (e.g. were there systems for professional supervision of case workers?) but these needed to be supplemented with other evidence. Similarly, some services that

performed adequately in management demonstrated higher risks in service delivery, indicating that adequate management systems are not a sufficient indicator on their own of adequate service delivery.

7. Strengths and weaknesses

There are a number of lessons from our experience of running this process over five cycles:

Strengths

- **Clear and consistent perspective on service performance** – The independence of these assessments is valuable in providing a clear and consistent view of service performance. The funding body for this program has two roles as both contract manager and mentor to services, and in practice these roles tend to overlap.
- **Added credibility** – an independent assessment also adds credibility for discussion with central agencies (e.g. Finance for refunding)
- **Method allows for comparison across different contexts** – one of the key features of this program is the different approaches taken by each service and the different contexts for service delivery. This method allows for an effective comparison of similar services in very different circumstances.

Weaknesses

- The process still takes a **reasonable amount of elapsed time** – whilst it is a relatively quick and cost effective process and does not take as long as a comprehensive evaluation or review, it is still time intensive – in 2005, it was approximately four months in elapsed time from project commencement to completion of 95% of the reports.
- **Data issues** – as for many programs, collection of client data was mixed, and there were some problems with on-line data collection. The central data was not used consistently for the services' own monitoring. In some cases, our issues paper was the first data report that the service had seen for some time. Services have become aware of the value of accurate data over time, providing the data is used and reported back to them.

Issues and lessons

- **Expectations change over time** – generally higher performance is expected from services after five years than in first year, as the program matures and the service and funding body have a better sense of what is achievable within the available resources
- However, there is **high staff churn** in services and in the funding body – constant change in staff is endemic to the sector, and a common consequence of restructuring of government agencies. This is an important factor to consider in the expectations of the services.
- **Not all of the program parameters were equal** – over time, some of the service requirements have taken on more significance. Poor performance in core functions (family focused early intervention) tended to indicate higher concern than poor performance in Action Research, which may have indicated a lack of research expertise, or in some cases, simply a limited ability to adequately document activity.
- **Next steps** – all the steps in performance management (client monitoring, self-evaluation, program evaluation) have been integrated into a performance framework, with improved monitoring processes which reduce the need for an assessment process on this scale.

8. Conclusion

This is an effective method for undertaking a periodic assessment of performance of a contracted provider which gives the funding body a good understanding of the current

activities of each service against its contract requirements. The method allows for consistent assessments and findings between different contexts and intervention approaches. Importantly, this method is relatively quick whilst providing a reliable basis for the funding body to recommend refunding a contract service provider or to intervene in the management of a service.