

Using the Delphi Technique for Assessing Evaluation Capability Building Needs

Julie McGeary, Victorian Department of Primary Industries

Introduction

This paper describes the use of the Delphi Technique for a component of a Needs Assessment to inform the development of an Evaluation Capability Building Plan for the Agriculture and Fisheries Group of the Victorian Department of Primary Industries.

The Victorian Department of Primary Industries (DPI) is a large public sector organisation promoting the sustainable development of primary and energy industries. Responsibility for Victoria's agriculture and fisheries sectors rests with the Agriculture and Fisheries Group (AFG) within DPI.

To support its Investment Strategy, AFG recently produced an Evaluation Framework which articulates the organisation's evaluation requirements and expectations. To ensure AFG's evaluation capability is sufficient to meet the requirements of the Evaluation Framework, an Evaluation Capability Building (ECB) Plan was requested. As an Evaluation Specialist within AFG's Evaluation Unit, with a specific responsibility for ECB activities, I was allocated the task of developing an ECB Plan.

Evaluation Capability Building

ECB has been defined as *"the intentional work to continuously create and sustain overall organisational processes that make quality evaluation and its uses routine"* (Stockdill et al 2002 p. 14).

While ECB is currently undertaken within AFG through mechanisms including a formal evaluation training program, mentoring arrangements, provision of technical advice and resources, and the operation of Evaluation Communities of Practice, these activities are not codified in an ECB Plan, the development of which is an important component of any ECB effort. King (2007) stresses the value of a purposeful, explicit, written capacity-building plan. Similarly, Mackay (2002) lists the preparation of a realistic ECB action plan as a key issue, while the development and implementation of *"a purposeful long-term ECB plan for the organisation"* is one of the checklist items provided by Volkov and King (2007). The lack of such an ECB Plan within AFG is a key driver for undertaking an ECB Needs Assessment.

Needs Assessment

Needs assessment has been defined as:

a systematic set of procedures undertaken for the purpose of setting priorities and making decisions about program or organisational improvements and allocation of resources. The priorities are based on identified needs (Witkin & Altschuld 1995 p. 4).

Within the Needs Assessment (NA) literature, need is usually defined as the discrepancy between a target state and an actual state (Roth 1990, Witkin & Altschuld 1995, Lee et al 2007).

Based on this definition, the five main components of a NA are: “(1) *determining the What Should Be (target) status*; (2) *ascertaining the What Is (actual) status*; (3) *quantifying discrepancies between What Should Be and What Is*; (4) *analysing the causes of discrepancies*; and (5) *establishing priorities*” (Lee et al 2007, Owen 2006, Witkin & Altschuld 1995).

As well as considering different types of need, Witkin and Altschuld (1995) identify three different levels of need which they relate to different NA target groups. The first level refers to the needs of service recipients. Witkin and Altschuld (1995 p. 12) stress that valid NAs should be focused on Level 1 as these are the needs of the people for whom the system ultimately exists. Level 2 refers to the needs of service providers, and the third level refers to the organisational needs or resources.

Data to determine the ideal state of evaluation capability within AFG was collected using the Delphi Technique with participants from Level 1 and Level 2 target groups, ie. both recipients and providers of evaluation services in DPI. This component of the NA is the subject of the next section of this paper.

Establishing the ‘What should be’ state using the Delphi Technique

Choice of Delphi

The Delphi Technique is recognised as an appropriate method for eliciting information for a NA (Owen 2006, Witkin & Altschuld 1995, Garavalia & Gredler 2004, Critcher & Gladstone 1998, Jones & Hunter, 1996). It has been described as “*a consensus method providing a means of harnessing the insights of appropriate experts to enable decisions to be made*” (Jones & Hunter 1996 p. 46), and “*used not only to determine consensus but also to enhance consensus building*” (Witkin & Altschuld 1995 p. 194).

The Delphi Technique is an iterative process involving a series of survey rounds with the same panel of ‘experts’. Each round is informed from responses from preceding rounds. While the rounds can continue until consensus is approached or achieved, most Delphi processes involve three or four rounds (Jones & Hunter 1996, Witkin & Altschuld 1995, Critcher & Gladstone 1998, Garavalia & Gredler 2004).

The technique was selected for this component of the ECB NA because of its advantages in enabling a large group of geographically-dispersed, time-poor participants to be contacted cheaply by email, avoiding the organisational constraints and expense of bringing them together physically. Administration by email allowed rapid turn around, while providing participants with anonymity, equal status and equal opportunity to participate. It also avoided some of the disadvantages associated with face-to-face meetings such as personality influences or individual dominance (Critcher & Gladstone 1998, Garavalia & Gredler 2004, Witkin & Altschuld 1995, Jones & Hunter 1996).

Method

Selection of participants

Potential participants were selected on the basis that they had expertise in evaluation or had a stake in its use within AFG. Participation was invited from those who receive evaluation services (Level 1 target group), and from those who provide evaluation services (Level 2 target group), resulting in a potential participant pool of 80 (47 Level 1 and 33 Level 2).

Round 1

Potential participants were invited by email to provide opinions based on their knowledge or experience, in response to the specific question: “*What would strong evaluation capability within DPI’s Agriculture and Fisheries Group look like?*” Following an analysis procedure suggested by Witkins and Altschuld (1995 p. 196), the 107 opinions received were clustered and themed independently by the coordinator and three other colleagues, then formed into a set of ten statements (refer to Table 2).

Round 2

Collated statements from Round 1 were distributed to all potential participants, who were invited to allocate ten points across the ten statements and provide comments justifying their decisions if desired. Responses were analysed to calculate rating frequencies and all 59 comments received were collated.

Round 3

Group responses from Round 2 were redistributed to Round 2 participants, together with their own points allocation for comparison, and the same rating and commenting process was repeated. This provided participants with an opportunity to rethink their original ratings in light of the data on the views of the group.

Resolution

Consensus was approached regarding which three statements were rated highest, and which three were rated lowest. This supported the decision that further Delphi rounds were unnecessary. All potential participants were then provided with an analysis and summary of the group data.

Findings

Participation

Table 1 shows a breakdown of participation in each round by target group Level. Overall, 24 per cent of the initial sample completed Round 3. It is interesting to note that Level 2 participants, those who provide evaluation services within the organisation, participated at almost double the rate of those from Level 1, who receive evaluation services.

Table 1: Participation by Round by Level

	Round 1	Round 2		Round 3	Overall Response rate
Level 1 (N=47)	8	11		8	17%
Level 2 (N=33)	12	14		11	33%
Totals (N=80)	20	25	Total (N=25)	19	24%
Response rate by Round	25%	31%		76%	24%

Scoring

Table 2 shows percentage points allocated for each statement from Round 2 and Round 3. The table shows that the 'voting' for the top two statements firmed up in the third round, providing confidence that there is general agreement amongst participants that

they consider these two issues the most important. Statement number 1 was rated third highest in both rounds, while statements 8, 9 and 2 were rated the lowest in both rounds.

Table 2: Results from Rounds 2 and 3

Number	Statement	R2 % points	R3 % points
5	AFG should have clear guidelines detailing organisational expectations for all aspects of evaluation.	13.2	18.4
6	AFG should have streams of timely evaluation information continuously being used by decision makers at all levels of the organisation.	12.4	16.8
1	AFG should be able to show it is delivering social, economic and environmental benefits for the community of interest.	11.6	11.1
3	Evaluation capability should be distributed across AFG Divisions with leadership provided by a core group of evaluation specialists.	10.4	10.5
7	AFG should have ongoing evaluation capability building in a variety of forms including technical advice, support materials, training, communities of practice, mentoring and supportive organisational structures.	11.2	8.4
4	AFG should have an accessible store of evaluation information, including previously collected data and examples of tendering documents, evaluation plans and reports.	9.6	7.9
10	AFG should have a core of evaluation specialists to advocate for evaluation, build capability and provide support for all staff across the organisation.	10.4	7.9
8	AFG should be using a variety of forms of evaluation at different levels of the organisation to suit different purposes.	8.8	7.4
9	Evaluation should be valued at all levels of AFG.	8.0	7.4
2	Each Project Assessment Group should be responsible for resourcing evaluations.	4.4	4.2

Agreement

Figure 1 shows a breakdown of Round 3 points allocation by target group Level. It reveals a lack of agreement between the two Levels regarding their views of the importance of different statements. For example, statements 4, 7 and 10 received much greater support from Level 2 participants than from Level 1 participants. Conversely, statement 8 was supported more by Level 1 participants than by Level 2 participants. These differences are masked if only the total points allocated to each statement are considered.

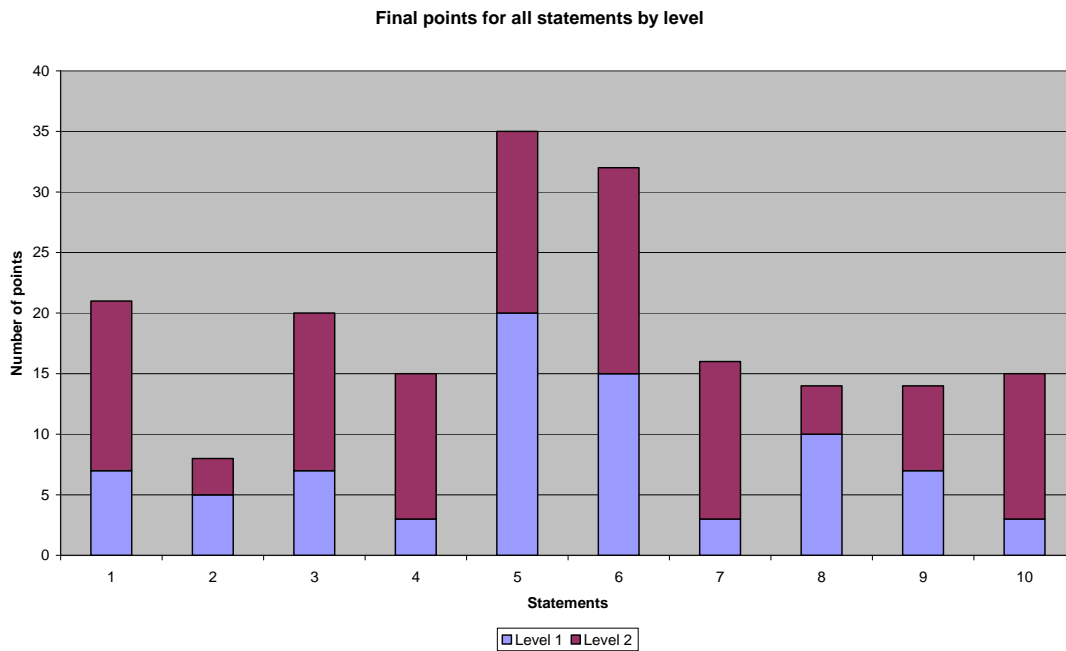


Figure 1: Round 3 points allocation by Levels.

Discussion

The purpose of this Delphi process was to gain agreement from an informed panel on the ideal state of evaluation capability within AFG. Jones and Hunter (1996 p. 47) distinguish between two forms of ‘agreement’. One form is the extent to which each respondent agrees with the issue under consideration, and the other is the extent to which respondents agree with each other. Overall consensus was approached for the three most important issues, and for the three least important issues, showing a reasonably high level of agreement by respondents with overall importance of the issues. Differences which emerged between Level 1 and Level 2 respondents, however, indicate lower levels of agreement between respondents. It will be important when developing an ECB Plan aiming to address identified evaluation capability discrepancies that the different perceptions of service providers and service receivers are taken into consideration, while at the same time remaining focused on the needs of those for whom the service ultimately exists.

The following three statements were rated by respondents to be the most important:

- Item 5. AFG should have clear guidelines detailing organisational expectations for all aspects of evaluation.
- Item 6. AFG should have streams of timely evaluation information continuously being used by decision makers at all levels of the organisation.
- Item 1. AFG should be able to show it is delivering social, economic and environmental benefits for the community of interest.

However, these preferences cannot be understood fully without reference to the comments provided by respondents to justify their points allocation. For example, the highest rated statement called for clear evaluation guidelines. This statement received 25 per cent of Level 1 points, and 13.6 per cent of Level 2 points. Two comments for this statement were provided by Level 1 respondents, and they both referred to AFG’s recently developed Evaluation Guidelines, with one claiming they were comprehensive,

and the other questioning the clarity of expectations they contain. Furthermore, three of the four comments provided by Level 2 respondents recognised the benefits of guidelines, but cautioned against them being too rigid or too prescriptive. This qualitative information will be an important aid to interpreting the quantitative 'voting' results, and care must be taken to ensure this further level of understanding is incorporated into the design of the proposed ECB Plan.

Meanwhile, the following three statements were rated by respondents as least important:

- Item 8. AFG should be using a variety of forms of evaluation at different levels of the organisation to suit different purposes.
- Item 9. Evaluation should be valued at all levels of AFG.
- Item 2. Each Project Assessment Group should be responsible for resourcing evaluations.

Comments provided for these statements also assist in understanding their apparent importance ratings. For example, statements 8 and 9 each received 7.4 per cent of all points allocated, indicating a low opinion of the importance of these issues. However, many comments indicated that respondents considered these issues to be '*self evident*', '*stating the obvious*', '*goes without saying*' and '*a bit of a motherhood statement*', suggesting the issues are still considered inherently important.

It should also be noted that although statement 8 rated low overall, it did receive the third highest rating from Level 1 respondents. With their needs requiring more consideration than those of Level 2 respondents in a demand-driven model, this issue should remain a focus for attention.

Methodological issues

Development of the initial question

Witkin and Altschuld (1995 p. 195) point out that developing the initial question for the first Delphi round can be a difficult exercise because it needs to capture the imagination of participants, generating intensive thought and interest. The 'visioning' form of the opening question; '*What would strong evaluation capability within DPI's Agriculture and Fisheries Group look like?*' was deliberately chosen to generate considered responses. The wording of the question avoided any reference to 'need' in an attempt to minimise confusion between needs and wants, a common problem with NA that has been identified in the literature (Scriven 1990, Owen 2006, Lee et al 2007, Witkin & Altschuld 1995, Altschuld & Witkin 2000, Gaber 2000, Titcomb 2000, Reviere et al 1996).

Another reason for not using the term 'need' was to minimise the likelihood of respondents thinking immediately of solutions rather than concentrating on the ideal state as requested. This confusion commonly occurs in NAs when 'need', used as a noun to describe the gap between actual and target states, becomes confused with 'need' used as a verb to indicate what is required to fill the gap (Witkin & Altschuld 1995).

While some of the opinions generated in response to the opening question suggested solutions, most described an ideal state.

Analysis of data from Round 1

Round 1 generated 107 opinions which were clustered into ten themes from which ten statements were developed. Because of the potential arbitrary nature of this process, efforts were made to test and validate results with both evaluation practitioners and non-

practitioners. The resultant ten statements were, however, fairly broad, expressing ideal goals with high desirability. The nature of the statements greatly influenced the design of Round 2, which in turn influenced the final number of statements.

Design of Round 2

Witkin and Altschuld (1995 p. 135) warn that when category scales are used to rate importance of socially desirable items which have high face validity, clustering at the high end of the scale often occurs. This makes establishing priorities from the result very difficult. To avoid such clustering, they recommend methods that force choices from respondents rather than those which invite ratings of each individual item. Because of the nature of the statements which emerged from Round 1, a variation of the 'budget allocation method' (Witkin & Altschuld 1995 p. 136) was used. This involved forcing respondents to choose through their allocation of ten points in response to the question: *"Which of the following 10 statements do you believe are most important for a strong evaluation capability in DPI's Agriculture and Fisheries Group?"*

Decision to confine the Delphi to determine the 'ideal' state

It was originally my intention to use the Delphi Technique to determine both 'ideal' and 'actual' states for the ECB NA. Discrepancies between the two states would then be quantified, analysed for causes, and then prioritised for treatment in an ECB Plan. Initial question designs were investigated, and a preferred design employing category scales addressing importance and current performance on items generated was proposed. Similar methods applied in NA are presented by Lee et al (2007) and by Rodski (2005).

However, upon closer reading of Witkin and Altschuld (1995), I was dissuaded from taking this course of action. In relation to interpreting data from discrepancy questionnaires, Witkin and Altschuld (1995 p. 60) warn: *"Too often, however, comparisons are made on noncomparable items, such as importance of an objective versus perceived achievement."* Discussing the disadvantages of the NA written survey, Witkin and Altschuld (1995 p. 129) caution that *"It is usually not an appropriate vehicle for directly determining discrepancies, that is, by seeking responses simultaneously to 'what is' and 'what should be' questions."* This warning is repeated later in the same book: *"We do not usually recommend a two-response format for surveys (judging 'what is' and 'what should be' on the same instrument)."* (Witkin & Altschuld 1995 p. 142).

Heeding the warnings given above, it was decided to confine the purpose of the ECB Delphi to determining the 'what should be' condition of evaluation capability within AFG. The actual state and subsequent discrepancies will be assessed using different methods. It was only after this decision was taken that I discovered the following statement by those same two authors in a later book: *"Well-constructed NA surveys contain (at a minimum) double-scaled items that ask for ratings about current and desired status, in accord with the definition of need (the measurable discrepancy between...). With two scores for each item, it is possible to calculate a numerical discrepancy or an index of need."* (Altschuld & Witkin 2000 p. 53) Whether or not the correct decision was taken remains to be seen.

Next steps

Using the Delphi Technique, a broad range of opinions has been generated from an 'expert' panel to determine the ideal or 'what should be' state of evaluation capability within AFG, and preliminary priorities have been established. The next task will be to test this information against recommended best practice ECB from the literature to ensure all

'essential' ECB elements are included in the priority list, and to incorporate recommendations from relevant previous NA studies.

This 'testing' will be undertaken in an attempt to overcome identified weaknesses of the Delphi Technique. Jones and Hunter (1996 p. 52) warn of the danger of the Delphi Technique deriving collective ignorance rather than wisdom. A similar warning is offered by Critcher and Gladstone (1998 p. 443) who caution that ". . . *what appears to be a high level of consensus might also be interpreted as the lowest common denominator of opinion*". These same authors remind the practitioner that ". . . *the Delphi method is an aid to decision making and not a substitute for it*" (Critcher & Gladstone 1998 p. 443).

Once 'reality' tested, the information generated from the Delphi process will be used to focus data gathering activities for establishing the actual or 'what is' state of evaluation capability within AFG. Existence of, and reasons for, discrepancies will then be identified. Methods other than the Delphi Technique will be used for these steps. This is consistent with recommendations from the NA literature that more than one type of data collection method should be used for NA to ensure that a variety of viewpoints are considered, to capture a broad understanding of identified discrepancies, and to distinguish between a need and a want better (Owen 2006, Witkin & Altschuld 1995, Rouda & Kusy 1995, Gaber 2000, Waterman 1990).

Final steps required for the completion of the ECB NA include deciding which identified needs should be given priority for action, followed by the development of an ECB Plan to address the needs. The selection and design of treatment strategies will need to be undertaken in close consultation with the evaluation audiences identified earlier. This step is described by Altschuld & Witkin (2000) as a major challenge but a necessary component of NA.

References

- Altschuld, J. W. and Witkin, B. R. (2000). *From Needs Assessment to Action: Transforming needs into solution strategies*. Thousand Oaks, Ca: Sage.
- Critcher, C., and Gladstone, B. (1998). Utilising the Delphi Technique in Policy Discussion: A case study of a privatised utility in Britain. *Public Administration*, 76 (Autumn), 431-449.
- Gaber, J. (2000). Meta-Needs Assessment. *Evaluation and Program Planning*, 23, 139-147.
- Garavalia, L. and Gredler, M. (2004). Teaching Evaluation through Modeling: Using the Delphi Technique to assess problems in academic programs. *American Journal of Evaluation*, 25(3), 375-380.
- Jones, J. and Hunter, D. (1996). Consensus Methods for Medical and Health Services Research. In N. Mays and C. Pope (Eds.), *Qualitative Research in Health Care* (Ch 6 pp. 46-58), London, BMJ.
- King, J. (2007). Developing Evaluation Capacity through Process Use. *New Directions for Evaluation*, 116, 45-59.
- Lee, Y., Altschuld, J. W., and White, J. L. (2007). Problems in Needs Assessment Data: Discrepancy analysis. *Evaluation and Program Planning*, 30, 258-266.

- Mackay, K. (2002). The World Bank's ECB Experience. *New Directions for Evaluation*, 93, 81-99.
- Owen, J.M. (2006). *Program Evaluation: Forms and Approaches*. (3rd ed.) Crows Nest, NSW: Allen & Unwin.
- Reviere, R., Berkowitz, S., Carter, C. C., & Ferguson, C. G. (Eds) (1996). *Needs Assessment: A creative and practical guide for social scientists*. Washington DC: Taylor & Francis.
- Rodski Survey Research. (2005). *Department of Primary Industries: Employee Opinion Survey Report*. Department of Primary Industries, Victoria, unpublished.
- Roth, J. (1990). Needs and the Needs Assessment Process. *Evaluation Practice*, 11(2), 141-143.
- Rouda, R. H., & Kusy, M. E. (1995). Needs Assessment: The first step. http://alumnus.caltech.edu/~rouda/T2_NA.html, Access date 1 April 2008.
- Scriven, M. (1990). Special feature: Needs Assessment. *Evaluation Practice*, 11(2), 135-140.
- Stockdill, S., Baizerman, M. and Compton, D. (2002). Toward a Definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7-26.
- Titcomb, A. L. (2000) *Needs Analysis*. <http://calscf.calsnet.arizona.edu/icyf/docs/needs.pdf>, Access date 1 April 2008.
- Volkov, B. and King, J. (2007). *A Checklist for Building Organizational Evaluation Capacity*. *Evaluation Checklists Project*. Available from www.wmich.edu/evalctr/checklists. Access date 13 August 2008.
- Waterman, S. (1990). Comment on Needs Assessment Article in EN #2. *Evaluation Practice*, 11(2), 140-141.
- Witkin, B. R. and Altschuld, J. W. (1995). *Planning and Conducting Needs Assessments: A practical guide*. Thousand Oaks, CA: Sage.