

TU13

Developing an approach for a summative meta-evaluation

1. What is the purpose of the present paper?

Why should evaluators conduct sound evaluations? Because decision-makers are more likely to use information from evaluations they believe are sound (Uusikylä and Virtanen 2000, Owen with Rogers 1999).

How do evaluators judge whether evaluations are sound? They evaluate them! This is commonly known as ‘meta-evaluation’ (Stufflebeam 2001, Owen with Rogers 1999). There are two types of meta-evaluation: formative, summative.

Evaluators undertake a *formative* meta-evaluation while planning and/or conducting their own evaluation. The evaluation is checked against selected criteria for sound evaluation practice, such as the Joint Committee on Standards for Educational Evaluation’s (the Joint Committee) Program Evaluation Standards (1994 in Owen with Rogers 1999). The findings shape the evaluation because they guide decisions about its design and/or implementation. Evaluators use *summative* meta-evaluation to critically appraise their own or another evaluator’s completed evaluation. The evaluation is checked against selected criteria and its strengths and weaknesses are identified. Evaluators, clients¹ and audiences² use this information to judge whether the evaluation is sound (Stufflebeam 2001, Owen with Rogers 1999).

The purpose of the present paper is to develop an approach for a summative meta-evaluation. The paper begins with an analysis of meta-evaluation and evaluation literature. This analysis identifies three key issues. The first is the criteria used to judge whether an evaluation is sound. The second is whether evaluators who work inside or outside an organisation should evaluate its evaluation. The third concerns the methodology that meta-evaluators use. Then the paper discusses the effect of these issues on a summative meta-evaluation of a communication skills evaluation. The paper concludes with an approach for this summative meta-evaluation.

2. Which meta-evaluation criteria are appropriate?

2.1 Program Evaluation Standards (1994) and Guiding Principles for Evaluators (1995)

Stufflebeam (2001: 185 his emphasis) defines meta-evaluation as:

...the process of delineating, obtaining, and applying descriptive information and judgmental information – about the utility, feasibility, propriety, and accuracy of an evaluation and its systematic nature, competent conduct, integrity/honesty, respectfulness, and social responsibility – to guide the evaluation and/or report its strengths and weaknesses.

¹ ‘Clients’ are people who commission an evaluation or meta-evaluation.

² The ‘primary audience’ are the people most likely to use the meta-evaluation’s findings. The ‘secondary audience’ are people interested in the meta-evaluation.

Stufflebeam's (2001) definition is based on the Joint Committee's Program Evaluation Standards (1994) and the American Evaluation Association's (AEA) Guiding Principles for Evaluators (1995). Jim Sanders (the Joint Committee's Chair in 1994) compared these Standards and Principles. He concluded that they were consistent and compatible³ (Stufflebeam 2001, Owen with Rogers 1999).

The Joint Committee groups its Standards For Evaluations Of Educational Programs, Projects, And Materials (1981) and its Program Evaluation Standards (1994) under four characteristics of sound evaluation: utility, feasibility, propriety, accuracy (Stufflebeam 2001, Owen with Rogers 1999, Sanders 1994, Joint Committee on Standards for Educational Evaluation 1981).

Utility refers to whether an evaluation provided practical information that met its primary audience's needs. *Feasibility* concerns whether an evaluation was practical, cost-effective and politically viable (i.e. whether the evaluators identified and worked cooperatively with relevant interest groups). *Propriety* relates to whether an evaluator conducted an evaluation ethically (i.e. with due regard to the welfare of the evaluation's participants and people affected by its findings). *Accuracy* is about whether an evaluation produced information that was valid and reliable for its intended use (Stufflebeam 2001, Owen with Rogers 1999, Sanders 1994, Joint Committee on Standards for Educational Evaluation 1981).

The AEA's Guiding Principles (1995) require evaluators to:

- conduct systematic, data-based evaluations
- practice within the limits of their professional competence
- be honest and act with integrity
- respect the privacy, safety, dignity and self-worth of the people participating in and affected by an evaluation
- consider an evaluation's contribution to the 'public good' (Stufflebeam 2001, Owen with Rogers 1999).

Stufflebeam's definition, the Joint Committee's Standards (1994) and the AEA's Principles (1995) were developed for American evaluators (Stufflebeam 2001). The Joint Committee argues that evaluators should assess whether it is appropriate to apply the Standards in other countries (ibid). It seems reasonable to apply this argument to Stufflebeam's definition and the AEA's Principles, because they were also developed for American evaluators.

2.2 Guidelines for the Ethical Conduct of Evaluations (2002)

The Australasian Evaluation Society's (AES) (2002: 4) Guidelines for the Ethical Conduct of Evaluations (2002) "...are designed to suit the cultural, social and institutional contexts of evaluation in Australia and New Zealand."

³ The Program Evaluation Standards (1994) is the second edition of the Joint Committee's Standards. The Standards For Evaluations Of Educational Programs, Projects, And Materials (1981) is the first (Curran 2000). The Joint Committee (1981) states in their Standards (1981) that evaluators should use them in conjunction with relevant laws, other appropriate professional codes and recent developments in evaluation practice.

The AES divides its Guidelines (2002) into three main stages of programme evaluation: commissioning and preparing an evaluation, conducting an evaluation, reporting the evaluation's results. Under each stage the AES identifies ethical principles that evaluators should observe at that stage. It then outlines procedural guidelines that might help evaluators to observe those principles.

The AES (2002) acknowledges that their Guidelines (2002) are complemented by the Joint Committee's Standards (1994) and the AEA's Principles (1994⁴).

The present author compared the AES' Guidelines (2002) with the Joint Committee's Standards (1994 in Owen with Rogers 1999), AEA's Principles (1995 in Owen with Rogers 1999) and Stufflebeam's (2001) definition of meta-evaluation. Fundamentally the Guidelines, Principles and definition promote sound evaluation practice, as characterised by the Joint Committee's standards for utility, feasibility, propriety and accuracy.

Specifically, there seem to be no complementary standards or principles for the AES's procedural guidelines for evaluation clients who are preparing an evaluation brief,⁵ selecting an evaluator and/or disseminating evaluation reports (guidelines 1, 9 and 22 are reproduced in appendix 1). Similarly, there appear to be no standards or principles for evaluators who are responding to an evaluation brief and/or competing for an evaluation contract (guidelines 2 and 8 are reproduced in appendix 1).

In New Zealand, the Treaty of Waitangi between Māori and the Crown is fundamental to the context in which evaluators work. Evaluators need to consider how they can collect good quality information with, from and about Māori. Te Puni Kōkiri's (1999) *Evaluation For Māori: Guidelines for Government Agencies* assists evaluators to do this. Thomas (2002) has also developed a framework for assessing the cultural appropriateness of programmes and services.

It seems then that it is appropriate for meta-evaluators to apply the Joint Committee's Standards (1994 in Owen with Rogers 1999), the AEA's Principles (1995 in Owen with Rogers 1999) and Stufflebeam's (2001) definition in New Zealand, because they promote generic characteristics of sound evaluation practice: useful, feasible, ethical, accurate. This comes with the proviso that meta-evaluators have criteria to judge whether an evaluation has met its Treaty of Waitangi obligations. Meta-evaluators also need to include the appropriate AES (2002) procedural guidelines for evaluation clients, and for evaluators responding to evaluation briefs and competing for evaluation contracts

⁴ Stufflebeam (2001) and Owen with Rogers (1999) cite the AEA's 1995 Guiding Principles. The AES cites the AEA's 1994 Guiding Principles.

⁵ A client commissioning an evaluation identifies the evaluation's purpose, key questions, preferred approaches, intended audiences and stakeholders in an evaluation brief (Australasian Evaluation Society 2002). An evaluation brief is not a written agreement between the client and an evaluator. Consequently, it does not come under the Joint Committee's standard 'P1 Formal Obligation'. Similarly, a brief is not a negotiated project plan that is initiated by an evaluator. Therefore, it is not incorporated in the Association's principles' 1 and 2 under 'Integrity/honesty' (Owen with Rogers 1999).

3. Who should conduct a meta-evaluation?

Evaluation authors agree that all evaluators, or multi-disciplinary evaluation teams, must possess the knowledge, abilities, skills and experience required to undertake an evaluation. Evaluators should not practice beyond the limits of their competence (Australasian Evaluation Society 2002, Stufflebeam 2001, Sanders 1994 in Owen with Rogers 1999, Shadish et al. 1995 in Owen with Rogers 1999).

These authors are debating, however, the basis of an evaluator's credibility and objectivity. Evaluators are described as 'external' or 'internal'. External evaluators are 'outside' an organisation responsible for an evaluand⁶. Internal evaluators are psychologically or physically 'inside' the organisation (they might have worked on the evaluand). Owen with Rogers (1999: 137) describe internal evaluators as "...in tune with the organisation and...willing to provide evaluative advice designed to improve its functions."

Cummings (1988 in Owen with Rogers 1999) suggests that credibility and objectivity are the greatest differences between external and internal evaluators. People believe that if external evaluators are outside an organisation responsible for an evaluand, then they are objective, and if external evaluators are objective, then they are credible. Internal evaluators 'earn' their status as credible and objective practitioners by having a history of producing sound evaluations for their employer.

The links between externality and objectivity, and objectivity and credibility, might be the basis for Owen with Rogers' (1999) statement that external evaluators should conduct summative meta-evaluations. Wadsworth (1997) argues that objectivity is a false basis for seeing external evaluators as credible. She thinks that external evaluators' "fresh perception" (1997: 20) is their basic strength. Wadsworth (1997) suggests that internal evaluators can counter audiences' perceptions that they are biased by overtly questioning their own assumptions, ideas, understandings and practices.

More fundamentally, Guba and Lincoln (1989) challenge people's belief in an objective reality. Instead they assert that multiple, socially constructed realities exist. They believe that individuals construct their own realities as they attempt to understand their experiences. Consequently, no one is objective because no one has an independent, unbiased perception of 'reality'. Guba and Lincoln (ibid) portray evaluators as working in partnership with clients and audiences to create better informed and more sophisticated constructions of their own, their clients' and their audiences' experiences of evaluands. They base credibility on how closely clients' and audiences' constructed realities match the realities that evaluators' construct and attribute to their clients and audiences.

Patton (2002), Schalock (1995) and Stufflebeam (2001) also appear to equate credibility with competence, honesty and sound evaluation practice. Patton (2002) indicates that evaluators build their credibility through a history of intellectual rigour, professional integrity and methodological competence. More specifically, Schalock

⁶ The term 'evaluand' refers to the object of an evaluation. The evaluation being evaluated is the evaluand in a meta-evaluation (Owen with Rogers 1999).

(1995) believes that evaluators can establish their credibility with a client and audience by:

- learning about a programme before evaluating it (this includes learning about the programme's context, history, purpose, content, structure, providers and users)
- planning the evaluation with the client and audience
- being competent
- being honest.

Stufflebeam (2001: 194) is pragmatic. He acknowledges that an evaluation's context will affect whether an internal or external meta-evaluator conducts the meta-evaluation:

Sometimes the evaluator cannot or need not engage even a single independent metaevaluator - especially when the target evaluation is internal, small scale, and informal. Even then, the evaluator can usefully self-assess evaluation plans, operations, and reports against pertinent professional principles and standards.

Stufflebeam's pragmatism might be based on his belief that professional standards are an objective basis for judging whether an evaluation is sound. He (1998: 293) also states that evaluators can keep their own personal biases in check by using "...multiple sources of evidence, multiple measures, multiple evaluators, multiple evaluations, professional standards, and metaevaluations⁷."

Owen with Rogers (1999) appear to suggest that an evaluation's purpose influences whether clients use internal or external evaluators. Clients commission internal evaluators to conduct evaluations for organisational learning, programme improvement and 'low-level' accountability. They commission external evaluators when there is a strong need for accountability and/or a perceived need for special expertise and 'objectivity'.

Patton (1997: 352 his emphasis) proposes that:

The more politicized the context in which an evaluation is conducted and the more visible an evaluation will be in that politicized environment, the more important to credibility will be an independent assessment of evaluation quality.

It appears then that all meta-evaluators should work within the limits of their competence and strive to produce sound meta-evaluations. The basis of a meta-evaluator's credibility is, however, under debate. Some evaluation authors portray external meta-evaluators as objective, and therefore credible, because they are outside an organisation responsible for an evaluand. Others indicate that internal and external meta-evaluators' credibility needs to be based on a history of producing honest, sound meta-evaluations. Internal evaluators can minimise the risk of being perceived as biased by, for example, openly questioning their own practices. But the extent to which perceived conflict of interest could undermine a meta-evaluation's

⁷ The present author assumes that Stufflebeam is referring to formative rather than summative meta-evaluations, although, evaluators might be more aware of controlling bias if they think that their completed evaluations will be evaluated.

purpose might guide a client's decision to use internal or external meta-evaluators. In some contexts the messenger is as important as the message. A client might need, for example, to provide external funders with information that will help them judge whether a favourable evaluation of a costly programme is sound.

4. How should a meta-evaluator conduct a meta-evaluation?

Stufflebeam (2001) has developed a general methodology for conducting meta-evaluations. It requires meta-evaluators to:

- demonstrate that they, or the multidisciplinary meta-evaluation team, possess the required expertise to conduct the meta-evaluation
- have a written agreement with their client
- interact with the meta-evaluation's client, primary audience and secondary audience
- collect relevant information, analyse it and judge the soundness of the evaluation
- disseminate the meta-evaluation's findings
- help the client and the primary audience to use the meta-evaluation's findings to improve future evaluations (if required and feasible).

Stufflebeam's (2001) requirements are explored below (except for his first requirement which was discussed in section 3).

4.1 Written agreement

The AES (2002), Stufflebeam (2001) and the Joint Committee (1994 in Owen with Rogers 1999) advise meta-evaluators to have a written agreement with their client. Such agreements help clarify understandings and resolve disputes between the parties. An agreement would specify:

- the meta-evaluators' tasks
- the meta-evaluation's budget and time frame
- the available resources and who has access to those resources
- the procedures for protecting privileged information
- the procedures for storing and disposing collected information
- the procedures for resolving disputes
- who is able to edit, publish and release the meta-evaluation report(s)
- who owns the meta-evaluation materials
- who is able to use the meta-evaluation materials after the meta-evaluation is completed (Australasian Evaluation Society 2002).

4.2 Interacting with clients and primary audiences

Clients and primary audiences are more likely to use a meta-evaluation's findings if they have worked with the meta-evaluators throughout the meta-evaluation process (Stufflebeam 2001, Owen with Rogers 1999, Guba and Lincoln 1989). Patton (2000: 426) reasons:

...intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they have been actively involved; and by actively involving the primary intended users, the evaluator is training users in

use, preparing groundwork for use, and reinforcing the intended utility of the evaluation every step along the way.

Stufflebeam (2001) acknowledges that a client's organisational culture, however, influences who is involved in a meta-evaluation and how they are involved. A meta-evaluation's context will affect how meta-evaluators identify and interact with clients and primary audiences.

4.3 Data management

4.3.1 Key questions

There is a dynamic interplay between a meta-evaluation's key questions and selected criteria. Sometimes a key question will lead to a criterion. Other times a criterion will lead to a key question. Stufflebeam (2001) states that the meta-evaluators' key questions should be based on the selected criteria so they can judge whether an evaluation is sound. A key question based on the Joint Committee's (1994 in Owen with Rogers 1999) utility standard U1 would, for example, be "did the evaluators identify all of the audiences involved in or affected by the evaluation?"

The meta-evaluation must answer the client's and primary audience's key questions also. These might or might not be based on criteria. A client might want to know, for example, whether an evaluation affected participants' opinions of the client. This acknowledges that meta-evaluators, clients and primary audiences value different things and so require different forms of evidence to judge the evaluation's soundness (Patton 2000, Stufflebeam 1998, Guba and Lincoln 1989). It also ensures that the meta-evaluation's findings are relevant to the primary audience and meet their information needs. This increases the likelihood that the audience will use the findings (Patton 2000, Owen with Rogers 1999, Stufflebeam 1998).

4.3.2 Selecting and prioritising criteria

Meta-evaluators need to work with a client and primary audience to select criteria for a meta-evaluation, according to Stufflebeam (2001). They might prioritise particular criteria that the evaluation must meet to be judged sound. Stufflebeam (ibid) advises meta-evaluators to document the reasons for such decisions so that audiences can scrutinise the values underlying the selection and prioritisation of criteria.

4.3.3 Data collection methods

Stufflebeam (2001) suggests that meta-evaluators attempt to answer a meta-evaluation's key questions by collecting and analysing existing, relevant, written information. This information includes agreements, evaluation plans, data collection tools and reports. If the existing information does not meet the primary audience's information needs, then meta-evaluators should use observation, interviews and surveys to collect new information about the evaluation being evaluated.

From this, it seems that primary audiences should consider carefully the effect of only specifying information needs that can be met by analysing existing information. This is especially so if people are involved in, or affected by, an evaluation's evaluand. The evaluand's users, deliverers, administrators and funders (and the evaluation's audiences, clients and evaluators) could be valuable sources of

undocumented and unanticipated information about the evaluation's strengths and weaknesses. A meta-evaluation's primary audience should either allow for undocumented information when specifying their information needs, or the meta-evaluators should acknowledge that the meta-evaluation is based on existing information only.

4.3.4 Analysing the data and judging an evaluation's soundness

Stufflebeam (2001) uses checklists based on the selected criteria to judge an evaluation. The evaluation is usually scored against each criterion and then graded using categories such as 'met / partially met / not met', or 'excellent / very good / good / poor'. Each piece of information used to judge an evaluation is referenced against a criterion. This enhances the meta-evaluation's credibility because the client and audiences can review the evidence for Stufflebeam's judgments.

Similarly, the Joint Committee (Sanders 1994) has developed a checklist for indicating whether an evaluation 'addressed' a standard fully, partially, or not at all (or whether a standard did not apply to the evaluation). The Committee has countered the risk of meta-evaluators using the checklist as a 'scorecard' by suggesting that they:

- list an evaluation's strengths, weaknesses and suggestions for improvement in regard to a standard
- use the listed information to judge the extent to which the evaluation addressed the standard and tick the appropriate box on the checklist
- use the checklist to judge the extent to which the evaluation addressed the applicable standards under each of the four characteristics of sound evaluation: utility, feasibility, propriety, accuracy
- summarise ways of improving the evaluation's soundness.

4.4 Reporting and dissemination

Meta-evaluators, clients and primary audiences need to agree on a reporting process when planning a meta-evaluation. Stufflebeam (2001) notes that meta-evaluators usually prepare a meta-evaluation plan, interim reports linked to important aspects of the meta-evaluation, and a final report. He suggests that for each report, meta-evaluators submit a draft to their client and primary audience, present the draft orally at a workshop and gather 'feedback' on it, and then submit a final version.

A client and/or primary audience could refuse to disseminate a meta-evaluation's findings if they were unfavourable. In such circumstances the client, primary audience and meta-evaluators would be bound by the clause in their written agreement (see section 4.1) that specified who was able to edit, publish and release the meta-evaluation report(s).

4.5 Follow-up services

Stufflebeam (2001) recommends that meta-evaluators and their client consider whether the meta-evaluators will assist the primary audience to interpret and use the meta-evaluation's findings constructively. He believes that such follow-up services strengthen a meta-evaluation's impact. Uusikylä and Virtanen (2000: 52) also argue that:

...the continuous and reflexive interpretation of evaluation findings is the only way to enhance organisational learning and thus increase the utilization of evaluation results.

Three key issues were explored in previous sections of the present paper: meta-evaluation criteria, internal and external meta-evaluators, meta-evaluation methodology. The effect of these issues on a proposed summative meta-evaluation of a communication skills evaluation will be discussed in a forthcoming section. The communication skills evaluation will be described next to provide context for that discussion.

5. What will the proposed summative meta-evaluation evaluate?

The Royal New Zealand College of General Practitioner's (the College) General Practice Education Programme prepares doctors for general practice in New Zealand. It has two stages:

- Stage I General Practice Education Programme (Stage I GPEP)
- Stage II Advanced Vocational Education (Stage II AVE).

Participants attain Fellowship of the College (FRNZCGP) after successfully completing both Stages of the General Practice Education Programme. College Fellows can apply to the New Zealand Medical Council for vocational registration. Vocationally registered general practitioners are able to practice independently under New Zealand's Medical Practitioners Act (1995).

The proposed summative meta-evaluation will focus on a communication skills evaluation that the Stage I GPEP Committee commissioned⁸.

5.1 Stage I General Practice Education Programme

Stage I GPEP assists doctors to develop practical knowledge, skills and appropriate attitudes for providing effective general practice health care (Royal New Zealand College of General Practitioners 2000). Putting patients at the centre of general practitioner consultations is a characteristic of effective care because it is associated with improved patient health outcomes. These include reduced levels of concern, reduced levels of discomfort, improved mental health and increased patient satisfaction (Stevenson 2002, Stewart et al. 2000, Kinnersley et al. 1999).

General practitioners must be able to communicate well with their patients if the practitioners are to build consultations around their patients' needs. Good communication underpins Belle Brown et al.'s (2001) patient-centred clinical method. This method requires a general practitioner to respond to a patient's cues by moving backwards and forwards between:

- exploring the patient's disease and their unique experience of the illness
- understanding the patient's disease and illness experience in the context of their life setting and stage of personal development
- working with the patient to define their problem, set management goals and identify the patient's and general practitioner's roles

⁸ The Stage 1 GPEP Committee is the evaluation's client and primary audience.

- finding opportunities to assist the patient to prevent diseases and improve their health
- building a long-term relationship with the patient
- working within the limits of their own capacity and the consultation's time constraints.

The Stage I GPEP aims to teach doctors the communication skills required to conduct a patient-centred consultation and so provide patient-centred care (Royal New Zealand College of General Practitioners 2000). The Stage I GPEP Committee wanted to know whether the Programme was achieving this aim. They commissioned a multidisciplinary team to evaluate doctors' communication skills. The proposed summative meta-evaluation will focus on this evaluation.

5.2 Evaluating doctors' communication skills

The Stage I GPEP Committee and evaluation team designed an outcomes-based Impact evaluation⁹ of the Programme delivered in Auckland¹⁰. They used a 'before and after' method, and existing data collecting instruments,¹¹ to measure any changes in doctors' communication skills between the beginning and end of the Auckland Programme.

5.2.1 Collecting data at the beginning of the Auckland Programme

Ten doctors in the Auckland Programme volunteered to participate in the communication skills evaluation. Each participant was asked to complete two self-assessment questionnaires in the first week of the Auckland Programme (January 2002). Groel et al.'s (1990) self-assessment questionnaire measured the participants' attitudes towards patient-centredness. Jenkins' (2000) questionnaire measured their perceptions of their consultations in terms of patient-centredness. Groel et al.'s (1990) questionnaire was chosen because they showed that it was reliable. Jenkins (2000) showed that his questionnaire had construct validity.

Each participant was also asked to audiotape five consultations with real patients in the first week of the Auckland Programme. Three College Fellows¹² used The Measure of Patient-centered Communication (MPCC) (Belle Brown et al. 2001) to score the participants' consultations. The MPCC was chosen because Boon and Stewart (1998) found that researchers had used the patient-centred method (now known as the MPCC) to assess doctor-patient communication skills in general practices. None of the other reviewed tools appeared to be more reliable, valid, or practical than the MPCC.

⁹ The University of Auckland's ethics committee approved the communication skills evaluation.

¹⁰ The Committee and the evaluation team decided that evaluating doctors' communication skills at all five Programme delivery sites was too costly. They chose Auckland because the Programme has 'settled' there (Owen with Rogers 1999); the Programme has been delivered in Auckland since 1974 and its current syllabus was formalised in 2000.

¹¹ The existing data collection instruments were two self-assessment questionnaires and an 'instrument' for measuring patient-centred communication (see section 5.2.1).

¹² These Fellows were chosen because the evaluation team wanted to compare the Fellows' experiences of reviewing videotaped consultations with their experiences of assessing audiotaped consultations. The Fellows were trained to use The Measure of Patient-centered Communication (MPCC) (Belle Brown et al. 2001).

5.2.2 Auckland Programme

Communication skills were an implicit part of all the seminars and workshops in the Auckland Programme. Specifically, two, two-day communication skills workshops were held.

The first communication skills workshop was in the second week of the Auckland Programme (after the participants were asked to complete the self-assessment questionnaires and audiotaped consultations¹³). It aimed to:

- introduce the concept of patient-centredness
- allow participants to practise communication skills in a safe educational environment
- enable participants to peer review recorded consultations.

Five kinds of educational methods were used:

- instruction
- modelling
- skill practise
- feedback
- small group discussion on communication skills.

The second workshop was in June 2002. It aimed to:

- consolidate the patient-centred skills already learnt
- teach more advanced communication skills such as motivational counselling skills, giving bad news and dealing with angry patients
- introduce actors and simulated consultations to practise the new skills
- enable participants to review videos of real consultations with peers and reflect on how to improve the consultations.

The previous educational methods were used along with role play and video review.

5.2.3 Collecting data at the end of the Auckland Programme

In the last six weeks of the Auckland Programme (September and October 2002) each participant completed the same self-assessment questionnaires they completed in the first week. Each participant also audiotaped a further five patient consultations.

5.2.4 Analysing the data

The evaluation team used statistical tests to compare the participants' before-Programme and after-Programme data. They only found one statistically significant difference; participants appeared to rate themselves more strongly towards having patient-centred consultations after completing the Auckland Programme (paired t test = 3.5, $p = 0.0067$).

¹³ Some participants audiotaped patient consultations in the first and second weeks of the Auckland Programme. This meant that they completed the audiotaping after the first communication skills workshop.

5.2.5 Reporting and Dissemination

The evaluation team will report their findings and conclusions to the Stage I GPEP Committee. The findings will be disseminated through journal articles and Conference presentations.

This section of the present paper provided the context for the proposed summative meta-evaluation of the communication skills evaluation. The next section explores how the three aforementioned key issues will affect that meta-evaluation (namely meta-evaluation criteria, internal and external meta-evaluators, and meta-evaluation methodology).

6. How will the key issues affect the proposed summative meta-evaluation?

The proposed summative meta-evaluation's purpose is to judge the soundness of the communication skills evaluation. The meta-evaluation's findings will assist the Stage I GPEP Committee¹⁴ to decide whether the College should undertake a nationwide evaluation of doctors' patient-centred communication skills.

6.1 Meta-evaluation criteria

The commissioned meta-evaluator and Stage I GPEP Committee will identify key questions. The Joint Committee's Standards (1994 in Owen with Rogers 1999) will be used to judge the evaluation's soundness because these Standards reflect the fundamental characteristics of sound programme evaluation: useful, feasible, ethical, accurate. Given that the evaluation team only found one statistically significant difference (see section 5.2.4) the Joint Committee's (ibid) cost effectiveness feasibility standard F3 will be prioritised.

It seems appropriate for meta-evaluators to apply the Joint Committee's Standards (ibid) in New Zealand, providing they have criteria to judge whether an evaluation meets its Treaty of Waitangi obligations. Criteria to judge whether the evaluation was culturally appropriate for Maori doctors and Maori patients will be developed.

The Stage I GPEP Committee commissioned an internal evaluation team to conduct the evaluation. They did not prepare an evaluation brief, or use a competitive tendering process. Consequently, the corresponding AES (2002) procedural guidelines 1, 2, 8 and 9 are inappropriate for the meta-evaluation.

The AES's (2002) procedural guideline 22 is also inappropriate because the evaluation team will release the final evaluation report, not the Committee. Thus, the Committee should not have an opportunity to breach the report's integrity.

The meta-evaluator will avoid only collecting information that is relevant to the Joint Committee's Standards (1994 in Owen with Rogers 1999), because this would prevent them from collecting unanticipated information about the evaluation's strengths and weaknesses. The Standards will guide (rather than prescribe) the types of information that are collected so that any other relevant information is gathered.

¹⁴ The Stage 1 GPEP Committee is the meta-evaluation's client and primary audience.

6.2 Internal and external meta-evaluators

The Stage I GPEP Committee has commissioned an internal evaluator to conduct the meta-evaluation. Audiences could perceive the meta-evaluator as having a conflict of interest because they are part of the team currently conducting the evaluation. The meta-evaluator will counter this perception by working with the Stage I GPEP Committee to produce an honest, methodologically sound meta-evaluation. The meta-evaluator will achieve this by:

- declaring their potential conflict of interest in the final meta-evaluation report
- documenting the reasons for selecting and prioritising particular criteria (see sections 4.3.2 and 6.1)
- documenting the evidence used to judge the evaluation's soundness (see sections 4.3.4 and 6.3.3).

An internal meta-evaluation seems appropriate given that the meta-evaluation's purpose is to judge the soundness of an internal evaluation for an internal client/primary audience. Its findings will contribute to:

- internal decision-making about the evaluation's soundness
- personal and organisational learning about outcomes-based Impact evaluation
- improving internal evaluations.

An internal meta-evaluation also seems appropriate for the meta-evaluation's context. The Stage I GPEP Committee will use the meta-evaluation's findings to help it decide whether the College should undertake a nationwide evaluation of doctors' patient-centred communication skills. The Committee will probably decide that the evaluation was not cost effective because only one statistically significant difference was found (see section 5.2.4). Consequently, it might decide to conduct an improved evaluation nationwide, or not to conduct one at all. The Committee wants to base its decision on sound meta-evaluation findings in case audiences challenge the decision. Otherwise the meta-evaluation's context does not seem highly politicised.

6.3 Meta-evaluation methodology

6.3.1 Written agreement

The meta-evaluator and Stage I GPEP Committee will write a meta-evaluation plan that includes the meta-evaluation's purpose, method, reporting strategies, ethical issues, budget and timeline. They will agree on that plan and any subsequent changes.

6.3.2 Interacting with clients and primary audiences

The meta-evaluator will work with the Stage I GPEP Committee throughout the meta-evaluation. The College supports this type of working arrangement because it results in organisational learning about evaluands and evaluation practice.

6.3.3 Data management

Different data collection methods will be used. This will minimise the limitations and errors associated with a single method (such as low response rates and data entry errors in postal surveys) (Patton 2002). The meta-evaluator will analyse the

evaluation team's proposal, ethics committee application, data collection tools and reports. Questionnaires about the evaluation process that the evaluation team posted to the participants in December 2002¹⁵ will also be analysed. Information about the evaluation's strengths and weaknesses, and ways of improving it, will be gathered from the assessors, evaluators and in a 'self-report'¹⁶.

The meta-evaluator will compare their own thoughts with the participants', assessors' and other evaluators' responses, and with information from the analysed documents. This will test the extent to which different data collection methods yielded essentially the same information. If the information is very consistent, the meta-evaluation's findings will reflect a shared understanding of the evaluation's soundness. If the information is very inconsistent, the findings will show the complexity of judging the evaluation's soundness (Patton 2002).

The meta-evaluator and Stage I GPEP Committee will develop a checklist and process similar to the Joint Committee's (see section 4.3.4). This means the information will be collected and analysed systematically, thereby reducing the likelihood of error and increasing the meta-evaluation's accuracy. The Stage I GPEP Committee would be able to review documented evidence for the meta-evaluator's judgements (Scriven 2000). Stufflebeam's (2000) guide to developing checklists will be used to create the checklist.

6.3.4 Dissemination

Stufflebeam's suggested reporting process (see section 4.4) is similar to the College's standard process. There is one difference. College staff email draft interim reports for comment (rather than present them orally at workshops) because Committee members live throughout New Zealand and only meet face-to-face occasionally. Consequently, the meta-evaluator will email draft interim reports to the Stage I GPEP Committee, gather 'comments' on them, and then email a final version.

The meta-evaluator will email the draft final report to the Committee two weeks before they meet face-to-face, present the report orally at the meeting and gather 'feedback' on it, then email a final version.

6.3.5 Follow-up services

The meta-evaluator will help the Stage I GPEP Committee make the best use of the meta-evaluation's findings over time.

7. What can be concluded?

The present paper began with an analysis of three key issues in the meta-evaluation and evaluation literature: meta-evaluation criteria, internal and external meta-evaluators, meta-evaluation methodology. Then the paper discussed the effect of these issues on a summative meta-evaluation of a communication skills evaluation.

¹⁵ This was after the participants completed Stage I GPEP and stopped participating in the evaluation.

¹⁶ The meta-evaluator's 'self-report' will contain their perception of the evaluation.

This analysis and discussion mean that the paper can conclude with a methodologically sound approach that will enable the proposed summative meta-evaluation to meet the Stage I GPEP Committee's information needs and contribute to their decision-making. The meta-evaluator will:

- agree on a written evaluation plan with the Committee
- identify key questions, select criteria and develop a checklist for the meta-evaluation with the Committee
- use different data collection methods
- use the aforementioned checklist to analyse the collected data
- report to the Committee throughout the meta-evaluation process via email and in a face-to-face meeting
- help the Committee to use the meta-evaluation's findings to decide whether the College should undertake a nationwide evaluation of doctors' patient-centred communication skills, and to improve future outcomes-based Impact evaluations.

8. Appendix 1

Australasian Evaluation Society’s Guidelines 1, 2, 8, 9 and 22

<p>Guideline 1 – Briefing document</p>	<p>“Those commissioning an evaluation should prepare a briefing document or terms of reference that states the rationale, purpose and scope of the evaluation, the key questions to be addressed, any preferred approaches, issues to be taken into account, and the intended audiences for reports of the evaluation. The commissioners have an obligation to identify all stakeholders in the evaluation and to assess the potential effects and implications of the evaluation on them, both positive and negative.”</p>
<p>Guideline 2 – Identify limitations, different interests</p>	<p>“In responding to an evaluation brief, evaluators should explore the shortcomings and strengths of the brief. They should identify any likely methodological or ethical limitations of the proposed evaluation, and their possible effect upon the conduct and results of the evaluation. They should make distinctions between the interests of the commissioner and other stakeholders in the evaluation, and highlight the possible impacts of the evaluation on other stakeholders.”</p>
<p>Guideline 8 - Compete honourably</p>	<p>“When evaluators compete for an evaluation contract, they should conduct themselves in a professional and honourable manner.”</p>
<p>Guideline 9 – Deal openly and fairly</p>	<p>“Those commissioning an evaluation and/or selecting an evaluator should deal with all proposals openly and fairly, including respecting ownership of materials, intellectual property and commercial confidence.”</p>
<p>Guideline 22 – Do not breach integrity of the reports</p>	<p>“In releasing information based on the reports of the evaluation, the commissioners have a responsibility not to breach the integrity of the reports.”</p>

(Australasian Evaluation Society Inc. 2002: 6, 8, 12)

9. References

Australasian Evaluation Society Inc. (2002). *Guidelines for the Ethical Conduct of Evaluations*. Lyneham, ACT: Australasian Evaluation Society Inc. Downloaded from www.aes.asn.au on 9 May 2003.

Belle Brown, J., Stewart, M. and Ryan, B.L. (2001). *Assessing Communication Between Patients And Physicians: The Measure Of Patient-centred Communication (MPCC)*. London, Ontario Canada: Thames Valley Family Practice Research Unit and Centre for Studies in Family Medicine.

Boon, H. and Stewart, M. (1998). Patient-physician communication assessment instruments: 1986 to 1996 in review. *Patient Education and Counseling*, 35, 161-176.

Curran, V.R. (2000). An Eclectic Model For Evaluating Web-based Continuing Medical Education Courseware Systems. *Evaluation & The Health Professions*, 23(3), 318-347.

Grol, R., Maeseneer, J., Whitfield, M. and Morkink, H. (1990). Disease-centred versus patient-centred attitudes: comparison of general practitioners in Belgium, Britain and The Netherlands. *Family Practice*, 7(2), 100-103.

Guba, E.G. and Lincoln, Y.S. (1989). *Fourth Generation Evaluation*. Newbury Park, California: Sage Publications, Inc.

Jenkins, M. (2000). Doctors' understanding of their consultations in terms of doctor/patient centredness. *Education for General Practice*, 11, 36-45.

Joint Committee On Standards For Educational Evaluation. (1981). *Standards For Evaluations Of Educational Programs, Projects And Materials*. New York: McGraw-Hill Book Company.

Kinnersley, P., Stott, N., Peters, T.J. and Harvey, I. (1999). The patient-centredness of consultations and outcome in primary care. *British Journal of General Practice*, 49, 711-716.

Owen, J.M. with Rogers P.J. (1999). *Program Evaluation: Forms and Approaches*. 2nd edition. St Leonards, NSW: Allen & Unwin.

Patton, M.Q. (1997). *Utilization-Focused Evaluation: The New Century Text*. 3rd edition. Thousand Oaks, California: Sage Publications, Inc.

Patton, M.Q. (2000). Utilization-focused Evaluation. In: D. L. Stufflebeam, G. F. Madaus and T. Kellaghan (Eds.), *Evaluation Models: Viewpoints On Educational And Human Services Evaluation*. 2nd edition. Boston/Dordrecht/London: Kluwer Academic Publishers, 425-438.

Patton, M.Q. (2002). *Qualitative Research & Evaluation Methods*. 3rd edition. Thousand Oaks, California: SAGE Publications, Inc.

Royal New Zealand College of General Practitioners. (2000). *General Practice Education Programme: Stage 1 Intensive Clinical Training Programme (GPVTP): Syllabus*. Wellington: Royal New Zealand College of General Practitioners.

Sanders, J.R. (1994). Applying the Standards. In: The Joint Committee on Standards for Evaluation. *The Program Evaluation Standards*. 2nd edition. Thousand Oaks, C.A.: Sage. 8-20.

Schalock, R.L. (1995). *Outcome-Based Evaluation*. New York and London: Plenum Press.

Scriven, M. (2000). *The Logic And Methodology Of Checklists*. Downloaded from <http://www.wmich.edu/evalctr/checklists/cdc.htm> on 12 May 2003.

Stevenson, A.C.T. (2002). Compassion and patient centred care. *Australian Family Physician*, 31(12), 1103-1106.

Stewart, M., Belle Brown, J., Donner, A., McWhinney, I.R., Oates, J., Weston, W.W. and Jordan, J. (2000). The Impact of Patient-Centered Care on Outcomes. *The Journal of Family Practice*, 49(9), 796-804.

Stufflebeam, D.L. (1998). Conflicts Between Standards-Based and Postmodernist Evaluations: Toward Rapprochement. *Journal of Personnel Evaluation in Education*, 12(3), 287-296.

Stufflebeam, D.L. (2000). *The Checklists Development Checklist (CDC)*. Downloaded from <http://www.wmich.edu/evalctr/checklists/cdc.htm> on 12 May 2003.

Stufflebeam, D.L. (2001). The Metaevaluation Imperative. *American Journal of Evaluation*, 22(2), 183-209.

Te Puni Kōkiri. (1999). *Evaluation For Māori: Guidelines for Government Agencies*. Wellington: Ministry of Māori Development.

Thomas, D.R. (2002). Evaluating the cultural appropriateness of service delivery in multi-ethnic communities. *Evaluation Journal of Australasia*, 2(2), 50-56, December.

Uusikylä, P. and Virtanen, P. (2000). Meta-Evaluation as a Tool for Learning: A Case Study of the European Structural Fund Evaluations in Finland. *Evaluation*, 6(1), 50-65.

Wadsworth, Y. (1997). *Everyday Evaluation On The Run*. 2nd edition. St Leonards, NSW: Allen & Unwin.