CRANFIELD UNIVERSITY

SHIRINE VOLLER

THE ROLE OF PROGRAMME EVALUATION
IN ORGANISATIONAL DECISION-MAKING
ABOUT MANAGEMENT AND LEADERSHIP DEVELOPMENT

SCHOOL OF MANAGEMENT

MSc BY RESEARCH THESIS

CRANFIELD UNIVERSITY

SHIRINE VOLLER

THE ROLE OF PROGRAMME EVALUATION
IN ORGANISATIONAL DECISION-MAKING
ABOUT MANAGEMENT AND LEADERSHIP DEVELOPMENT

SCHOOL OF MANAGEMENT
CENTRE FOR EXECUTIVE LEARNING AND LEADERSHIP

MSc BY RESEARCH THESIS
ACADEMIC YEAR 2008-2010

SUPERVISOR: DR. DAVID DENYER
OCTOBER 2010

Abstract

A number of models of evaluation exist in the education and training domain, yet most are rarely cited in studies of the effectiveness of management and leadership development. Given the vast sums of money that continue to be invested by organisations in management development, we must assume that organisations are perceiving value and that this is identified, in the main, through processes other than formal evaluation. Relatively little attention has been paid to how organisations use evaluation models and data, and a specific challenge has been raised about the usefulness of evaluation models in informing decision-making.

To begin to address this challenge, this study set out to document the role of programme evaluation in organisational decision-making about management and leadership development. Eight customised programmes aimed at "high flyers" and middle to senior managers were studied. In each case, the way in which decisions were made and the role played by evaluation were explored from multiple stakeholder perspectives.

The study found that programme evaluation did contribute to decisions made about management and leadership development, but only to a limited extent, and only to a subset of decisions. Central to the organisational decision-making process were stakeholder preferences and experience. This included anecdotal data and impressions gained informally by key stakeholders who came into direct contact with a programme and its participants. External influences, unrelated to the programme itself, were also important.

Given these findings, it is suggested that there may be value in re-framing evaluation as part of an Evidence-Based Management (EBMgt) approach. EBMgt recognises the contribution of four elements of evidence to effective decision-making. Currently, decision-making about management and leadership development programmes does not draw fully on all four elements. Evaluation, both context-specific and casting light on external evidence, is under-utilised, resulting in potentially sub-optimal decisions. Thus, future research to explore how to better integrate these elements would be valuable.

Acknowledgements

Professor Cliff Bowman started me on the journey of this MSc, and although I have probably cursed and thanked him for it in equal measure over the past two years, I will always be immensely grateful for his belief in me and concern for my development. His ability to provide incisive comments and original ideas at the drop of a hat has helped me through many a sticky moment.

The urge to grab a pencil and draw a picture is, I have come to realise, a sign of clear and conceptual thinking, and a habit shared by Cliff and my supervisor, Dr. David Denyer. I have walked away from numerous supervisions clutching a diagrammatic gem which invariably helped to clarify my own thinking. I would also like to thank David also for his relaxed attitude to supervision and practical approach to research.

I benefited from the expertise and support of Cranfield's and Ashridge's librarians, Anita Beal, Anna Bosch, Lorraine Oliver and Heather Woodfield, in all aspects of literature searching and sourcing. Their direction was particularly valuable in the early stages of the review process, and in their tireless sourcing of material throughout.

Dr. Colin Pilbeam and Professor Kim Turnbull James sat on my review panel and provided pertinent comments and critique at key stages along the route toward writing the final thesis. I would like to thank them for the time and attention they gave.

Numerous other colleagues at Ashridge and Cranfield, including members of my Research Methods Course, provided ideas and much-needed support. I would specifically like to thank the client directors who provided access to their clients and agreed to be interviewed, and, of course, the clients without whom I would have had no data to work with.

Finally, I would like to thank my long-suffering partner, Tom, for bearing with me throughout what has been a topsy turvy two years.

CONTENTS

LIST OF APPENDICES

LIST OF TABLES AND FIGURES

CHAPTER 1. INTRODUCTION

Management and leadership development remains a key priority for many organisations (Alimo-Metcalfe and Lawler, 2001) and a sizeable industry – estimates put annual global spend on leadership development alone at over £30 billion (Reade and Thomas, 2004). Even the recent recession only temporarily suppressed the market's appetite, and 2011 is projected to bring an increased emphasis on such development (Duke Corporate Education, 2010).

Over the past fifty years, measuring the effectiveness and impact of management and leadership development has been a challenge for both the practitioner and academic communities. Various attempts have been made to address the question: 'does management and leadership development work?' through empirical studies and by constructing syntheses of existing research in the field through the use of meta-analytical techniques. These studies have revealed moderate levels of effectiveness with regard to participants' reactions to programmes and self reports of learning and behavioural change (Burke and Day, 1986; Collins and Holton, 2004). However, they provide limited evidence for organisational impact.

The systematic review for this thesis builds upon and updates the existing syntheses. Endorsing their findings, it finds that attempts to make the connection between development interventions and organisational performance continue to result in, at best, moderately helpful measures. The research emphasis is moving toward using evaluation to understand the dynamics of specific programmes and what contributes to or inhibits their effectiveness. This implies a shift in the dominant evaluation purpose from *proving* value to *improving* and *learning* (Easterby-Smith, 1994).

However, for those commissioning management development, ascertaining whether their investment has been worthwhile remains relevant. Many practical challenges exist and, despite the range of theoretical models of evaluation available, they are rarely used in practice. The exception is Kirkpatrick's four-level framework (1998) which still dominates the evaluation discourse, though examples of its application at the level of measuring organisational results or systemic impact are rare (Alliger et al., 1997).

Recently, researchers have challenged the utility of evaluation models, arguing that most do not support the ways in which organisations make decisions about management and leadership development and this is why they are rarely used (Holton and Naquin, 2005). The concern that evaluation, ultimately, should support improved decision-making is not new (Hesseling, 1966; Warr et al., 1970; Burgoyne and Cooper, 1975), yet relatively few researchers have set out to reconcile these two domains. Holton and Naquin (2005) call for "naturalistic decision research to understand how HRD decisions are really made in organisations" (2005, p.277). Such research, they argue, would inform how evaluation might be re-conceived of as a decision-making process and thereby enhance its use and standing.

This thesis attempts to make a start in answering their call by providing a theoretically informed and empirically grounded account of how decisions are made about

management and leadership development programmes and the role of evaluation data in the decision-making process. 'Programme evaluation' is considered here to relate to the planned and purposeful collection of data about a programme, and this working definition does not extend to include informal feedback processes.

Eight customised programmes aimed at "high flyers" and middle to senior managers were included in the study. In each case, the way in which decisions were made and the role of evaluation were explored from multiple stakeholder perspectives and using secondary data pertaining to the programmes in question.

The study found that programme evaluation does contribute to decisions made about management and leadership development, but only to a limited extent, and most substantially to the decision to modify programme design. More important seems to be the impressions gained by key stakeholders through direct exposure to a programme, its participants and their actions. It is suggested that those responsible for running programmes should therefore seek to maximise opportunities for direct stakeholder contact and acknowledge the importance of informal feedback processes in influencing decision-making. They should also seek to improve the quality of the data used to inform the decision-making process. One lens through which to achieve this, and where further research would be beneficial, is that of Evidence-Based Management (EBMgt). Specifically, future research should focus on how to better integrate all four elements of EBMgt, including evaluation, in order to improve the quality of decision-making about management and leadership development programmes.

The thesis is divided into this and seven subsequent chapters:

Chapter 2 introduces the field of management and leadership development evaluation and encompasses definitions of evaluation, its purpose, evaluation models, designs and methods and what evaluation studies have revealed in terms of the effectiveness of management and leadership development.

Chapter 3 presents the key findings of a systematic review of the last 10 years of literature on the effectiveness of management and leadership development, and how that effectiveness is assessed in terms of approaches to evaluation. It concludes with a statement of the research questions identified for the empirical work, and an explanation of why these are felt to be important.

Chapter 4 outlines the research design, setting and methods of data collection and analysis for the empirical work.

Chapter 5 presents the findings of the study. It comprises a summary of evaluation methods used across the eight management and leadership development programmes, a descriptive 'vignette' of each programme, and a synthesis of how key decisions are reached at each stage in the lifecycle of a programme and how evaluation contributes to these.

Chapter 6 interprets the findings in the form of a discussion. Salient aspects of the research are discussed and suggestions put forward which might explain the results.

Chapter 7 draws out the implications raised by the study, and acknowledges its limitations.

Chapter 8 makes a short conclusion based on the empirical work.

A list of references follows the conclusion, and the body of the thesis is adjoined by a number of appendices, including a substantial part of the systematic literature review.

CHAPTER 2. LITERATURE REVIEW PART I

The literature review comprised two distinct stages. Part I mapped the key domains of literature in relation to the evaluation of management and leadership development. It is the subject of this chapter and includes definitions of evaluation, its purpose, evaluation models and designs and an assessment of management and leadership development effectiveness as determined by prominent meta-analyses of the field. Part I provides an analytical frame for part II, an in-depth, systematic review of the most recent literature on the effectiveness of management and leadership development, and the ways in which evaluation is deployed in organisations.

2.1 The evaluation of management and leadership development programmes

A variety of conceptions of evaluation of management and leadership development programmes exist. One definition sees evaluation as the systematic collection of data to determine success in terms of quality, effectiveness or value (Goldstein, 1986; Hannum and Martineau, 2008). It occurs when specified outcome measures are conceptually related to intended learning objectives (Kraiger et al., 1993). Evaluation arguably could encompass the total value of a training intervention in social *and* financial terms (Talbot, 1992) and as such, be distinguished from validation, which is limited to assessing achievement of pre-determined objectives. However, this definition and distinction are not commonly agreed upon, and do not serve to reflect the breadth of what evaluation can offer. In this thesis, evaluation is considered to relate only to the planned and purposeful collection of data about programmes. This definition does not extend to including informal and incidental feedback processes, though these are recognised as being both important and influential, as will be seen later in the thesis.

By and large, in practice, evaluation is distinguished from 'research' in that the emphasis is on measuring results. Research seeks to understand relationships among variables or describe phenomena as a means of developing knowledge that can be generalised and applied (Hannum and Martineau, 2008). However, evaluation can also contribute to theory-building (Burgoyne and Singh, 1977) and a number of models of evaluation attempt to incorporate an explanatory dimension in addition to pure measurement, including so-called systemic models (Warr et al., 1970; Stufflebeam, 1989) and other models incorporating contextual factors (Baldwin and Ford, 1988; Holton, 1996; D'Netto et al., 2008).

In order to make decisions about evaluation designs and methodologies it is important to clarify the purpose of any evaluation. There are a range of reasons why evaluation might be important, succinctly categorised by Easterby Smith (1994) as *Proving, Improving, Learning* and *Controlling* (2008).

Figure 1. Easterby Smith's (1994) varying purposes of evaluation

Purpose of	PROCESS	OUTCOME
SUMMATIVE	CONTROLLING Is it going according to plan?	PROVING Is it achieving what was intended?
FORMATIVE	IMPROVING Is there a better way of doing what we are trying to do?	LEARNING Can we re-visualise what we are trying to do?

Whether evaluation is to 'prove' or 'improve' is at the heart of the distinction between summative and formative evaluation, where summative evaluation is about proof and formative evaluation is about improvement (Michalski and Cousins, 2001).

Burgoyne and Cooper (1975) articulate a distinction between evaluation as 'pure' research that adds to the body of theory about the efficacy of different educational methods, and evaluation in a specific context to provide timely data about an intervention. Both, they argue, should provide feedback to support decision-making. Other researchers have also argued that improved decision-making, or control, should be the purpose of evaluation research (Hesseling, 1966; Warr et al., 1970; Hamblin, 1974). These purposes are congruent with Easterby-Smith's (1994) classification scheme.

Specifically, evaluation can be used to demonstrate the potential or realised benefit that individuals, organisations and communities gain from a programme; to modify an existing programme so that it better achieves its objectives; to show how programme participation connects to enhanced organisational performance or societal improvement; to promote participant-reflection in order to enhance learning or; to encourage greater discussion of what works and why (Hannum et al., 2007).

Collins and Denyer (2008) argue that asking the pertinent questions: *Evaluation for whom? Evaluation of whom?* (Hesseling, 1966) and *evaluation of what?* might lead to a robust conversation about purpose and pave the way for constructive discussions about evaluation design. Whilst stakeholder-based evaluation (addressing the *Evaluation for whom?* question) is well developed in programme evaluation literature, it is 'barely recognised' in the context of training evaluation (Michalski and Cousins, 2001, p.37). Different stakeholders have different perspectives on the purposes – and consequences - of an evaluation (Hirsh and Burgoyne, 2009), as Michalski and Cousins (2001) set out to demonstrate, yet this may not be fully taken into account by instigators of evaluations if blinkered by their own perspective. Power and politics also play a role, and if evaluation is inherently political (Kim and Cervero, 2007), and if different stakeholders have different power-relations with one another, then these need to be considered in the design and subsequent interpretation of any evaluation.

The purpose of evaluation can be further complicated due to a lack of goal clarity of the intervention itself. Burgoyne and Singh (1977) argue that a vicious circle is created when training lacks clear goals, leading to a lack of certainty as to how and whether the goals have been achieved, and the absence of a coherent evaluation strategy. Each factor reinforces the others and, they argue, possibly the only way in which this circle

can be broken is through the developing a coherent evaluation strategy, which drives the effort to elicit goal clarity.

Clarifying purpose has implications for the choice of evaluation design and methodology. If evaluation is being conducted as research, then the emphasis should be on accuracy and robust methodology in order to allow for the discovery of generalisable relationships and development of empirical theory. However, if the purpose is to provide feedback in a timely and cost-effective manner in order for decisions to be made about a specific training programme, then certain compromises on methodology may be an acceptable sacrifice for pragmatism (Burgoyne and Cooper, 1975; Burgoyne and Singh, 1977). Lack of clarity in turn can lead to dissatisfaction with its results.

An analysis of evaluation from a decision-making perspective (Holton and Naquin, 2005) concluded that most evaluation models fit within a rational-economic framework which decision-making research has shown do not work in practice. Following their logic, if the purpose of evaluation is as a means to making decisions about investment in training and development, then the majority of models are inadequate. Models that incorporate decision-making theory more fully are called for, with Holton's (1996) model and the Evaluative Learning for Inquiry in Organizations (EILO) model (Preskill and Torres, 1999) cited as exemplars. These models may well be more useful for managerial decision-making. However, there remains a lack of basic research evidence exploring how decisions about management and leadership development programmes happen in practice, and what contribution evaluation processes and data make to these.

2.2 Approaches to evaluation

Many models of evaluation from the training domain have been theoretically derived from extant literature, and adopted and converted into workable evaluation designs to varying extents, for which a range of data collection and analysis methods are employed. Three key perspectives are identified: a taxonomy, or levels of impact, perspective; a transfer perspective; and a systems perspective. Each of these approaches is underpinned by a set of beliefs and assumptions.

2.2.1   The taxonomy or levels of impact perspective

The approach most often adopted by organisations is the Kirkpatrick model. His four-step evaluation model has been in circulation since the late 1950s (1959a; 1959b; 1960a; 1960b) and dominates the field of evaluation (Holton, 1996; Alliger and Janak, 1989) in terms of its direct use and its influence on other models and frameworks. Despite acknowledgement of its limitations (Alliger et al., 1997) and attempts to move the field of evaluation forward as exemplified by the perspectives outlined below, alternative models are often derivations of Kirkpatrick (Swanson and Holton III, 1999; Burgoyne et al., 2002). The Kirkpatrick model has become a 'global taxonomy' to enable people to think about evaluation criteria (Alliger and Janak, 1989). In practice, the approach provides a universally applied model for evaluation designs.

The model has four 'steps' – more often referred to as 'levels': *reaction, learning*, *behaviour* and *results*, with a fifth level sometimes added to encompass the ultimate value of training in terms of organisational success criteria (Holton, 1996). Authors have variously conceptualised this in terms of economic benefits or human good (Hamblin, 1974), societal value (Kaufman and Keller, 1994) or financial return on investment (Phillips, 1995). Kaufman and Keller (1994) are not alone in also proposing modifications to existing levels, such as expanding the remit of Kirkpatrick beyond training to human performance interventions in general.

The use of the model tends to be hierarchical and is underpinned by three implicit assumptions (summarised in Alliger and Janak, 1989) which Kirkpatrick may not originally have intended. That is, levels are arranged in ascending order of the value of information provided; the levels of evaluation are causally linked; and each succeeding level is correlated with the previous level and correlations among all levels are positive. However, caution should be applied in using findings from one level to infer causal links and conclusions about another since only limited evidence of correlations between the levels has been found (Alliger et al., 1997; Alliger and Janak, 1989).

Early industry studies (Catanello and Kirkpatrick, 1968; Kirkpatrick, 1978) showed a clear prevalence of level one evaluation, whilst a later meta-analysis (Alliger and Janak, 1989) found that most published studies focused on a single level of evaluation, but not necessarily level one. This single-level focus is not, however, conclusive. Other studies (Collins, 2001; Collins, 2002; Russon and Reinelt, 2004), have found evidence of leadership development programmes being evaluated on multiple levels. Russon and Reinelt (2004) found that the main focus was on individual outcomes (knowledge, skills, attitudes, changes in behaviour, values and beliefs), but programmes were also evaluated at the organisational, community, field or systems level. Collins' meta-analyses (2001; 2002) found that organisational performance was more frequently becoming a goal for leadership development, but that the proportion of studies measuring results-level outcomes was still small.

Alliger et al (1997) developed an augmented version of Kirkpatrick's four-level framework to facilitate their meta-analysis. This modified framework segmented two of the four levels, subdividing level one into *Reactions as affect*, *Reactions as utility judgements*, and a category that encompassed both, *Combined reactions*. Level two, learning, was similarly subdivided, into *Immediate post-training knowledge*, *Knowledge retention*, and *Behaviour/skill demonstration*. Hirsh et al's' (2002) framework for the performance impact of business school education is another that appears to be informed by Kirkpatrick's framework, though less overtly, and visualises a value chain from business school inputs through to global economic performance and social benefit.

2.2.2 The transfer perspective

Whilst not concerning evaluation as such, Baldwin and Ford's (1988) conceptual framework of the learning transfer process is fundamental to the development of later models of evaluation, and to the understanding of the influences on training effectiveness. Transfer, the degree to which trainees effectively apply the knowledge,

skills and attitudes gained in a training context to the job (Newstrom, 1984, in Baldwin and Ford, 1988) is argued to be a function of more than the original learning (Atkinson, 1972), and for transfer to have occurred "learned behaviour must be generalised to the job context and maintained over a period of time on the job".

Baldwin and Ford synthesise an extensive body of existing research to specify three constituent phases: *Training inputs, Training outputs* and *Conditions of transfer*. Within *Training inputs* they discriminate between trainee characteristics, training design and work environment. The *Training outputs* phase deals with learning and retention, whilst the phase *Conditions of transfer* is to do with generalisation and maintenance.

The Holton model (1996), informed by Baldwin and Ford's (1988) model of transfer of training, proposes an integrative evaluation model that claims to account for the impact of primary and secondary intervening variables. Like Baldwin and Ford, Holton notes that factors external to training design will affect the impact of training. Holton's model proposes three primary outcome measures: *learning, individual performance and organisational results*, which show similarities with Kirkpatrick's levels two, three and four. However, unlike Kirkpatrick, Holton recognises the potential influences of trainee reactions, motivation to learn and ability on *learning outcomes*, and motivation to transfer, transfer conditions, or environment, and ability on *performance outcomes*.

Whilst conceptually appealing due to its explanatory power, Holton's model is also complex and relatively untested, perhaps due to its recent development, but perhaps also due to the demands in its implementation (Holton and Naquin, 2005).

Whilst not referring to Holton's (1996) or Baldwin and Ford's (1988) models, D'Netto et al (2008) cover similar territory with their *Predicted model of management development effectiveness*. They divide management development into a three-stage process which comprises needs assessment, the development programme, and evaluation. They focus purely on what happens before and after the programme, including *antecedent components* and *post-programme components* (p.4). Antecedent components comprise organisational learning culture, individual initiative, top management support and the programme's link to corporate strategy, whilst post-programme components include evaluation, line manager support and opportunities to use the skills learned. With just one empirical study to have operationalised this approach (D'Netto et al., 2008) the empirical evidence in support of the model is thin. However, as with Holton's (1996) and Baldwin and Ford's (1988) models, it demonstrates the attention that needs to be paid to a range of factors outside of the delivery of a programme that are likely to affect its impact. Collins and Denyer (2008) argue that evaluation is frequently based on a deficit model of training whereby a programme is administered to a population and an assessment is made of whether or not, and by how much, the intervention has worked. They suggest that such approaches cannot answer crucial 'why' questions which would help to explain the intricacies and contextual factors that lead to variation in the success of a given training programme. This is critical if the purpose of evaluation is either *improving* or *learning* (Easterby-Smith, 1994). Arguably, models such as Holton's (1996) and D'Netto et al's (2008)

incorporate an explanatory element in terms of specifying potential confounding variables beyond the design and delivery of a programme itself,

2.2.3 The systems perspective

Several so-called systems models of evaluation have been developed which emphasise the importance of context. These include the CIRO (Context, Input, Reaction, Outcome), CIPP (Context, Input, Process, Product) and CIMO (Context, Intervention, Mechanism, Outcome) approaches (Warr et al., 1970; Stufflebeam, 1989; and Denyer et al., 2008 respectively).

The use of the term 'context' is variable. Both CIRO and CIPP embrace the initial phases of identification of need and programme initiation: CIRO focuses on programme conception, considering 'context' to be the problems in the organisation from which specific training needs and learning objectives can be derived (Easterby-Smith, 1981). CIPP emphasises the importance of the target audience and how their needs are aligned with the resultant design, before moving on to the effectiveness of implementation and results. Both CIRO and CIPP thus see 'context' in terms of needs assessment (Martineau, 2004), which is often weak (Baker and Gorman, 1978). Another perspective sees 'context' as the factors which could affect implementation *after* a programme's purpose and objectives have been determined (Thurley et al., unpublished, cited in Easterby-Smith, 1981).

Denyer et al (2008) developed a further systemic model, leveraging the work of Pawson and Tilley (1994), and informed by design science. In their CIMO (Context, Intervention, Mechanism, Outcome) framework the authors propose that: 'to understand outcomes (O) it is important to focus on certain hypotheses about mechanisms (M) through which an intervention seeks to bring about change as well as recognising the contextual conditions (C) (Collins and Denyer, 2008, p.170). This addresses concerns that an explicit programme theory/theory of change has been found to be wanting in some leadership development programmes, and desired impact is disconnected from teaching and learning activities (Russon and Reinelt, 2004).

Perhaps the levels of impact perspective might be best aligned with Easterby-Smith's (1994) purposes of *proving* and *controlling*, the transfer perspective with *improving*, and the systems perspective with *learning*. From a managerial decision-making standpoint, Kirkpatrick remains the dominant model in use, perhaps due to its longevity combined with conceptual simplicity, which belies the challenges inherent in its implementation.

2.3   The design and operation of programme evaluation

Research design deals with the practical aspects of how an evaluation is conducted and, alongside magnitude and criteria collected, is one of the core components of an evaluation strategy . Magnitude is a trade-off decision: Extensive evaluations are more costly and the data more reliable than smaller evaluations. Small evaluations cost less,

but the data is likely to be less reliable than an extensive evaluation study. Research design can be simple through to complex, with simplistic designs being easy to implement, yet yielding data of limited value, and complex designs having the potential to deliver insightful findings, but being met with the challenges of cost, participant resistance and difficulty of implementation. There used to be a geographical difference between the US and Europe in this regard, with Americans favouring a 'pure' research approach, whilst European studies tended to favour greater pragmatism and timeliness, at the expense of compromised data quality (Burgoyne and Cooper, 1975).

A hierarchical classification of research design places pre-experimental, quasi-experimental and experimental designs (Tannenbaum and Woods, 1992) in increasing order of sophistication. Pre-experimental designs include case study work or post-training comparison of two groups (a control and the experimental group).

Experimental designs, in contrast, require random assignment of individuals into training or control group and, often, the collection of pre- and post-training data, a design referred to as Pretest-Posttest With Control (PPWC) or, if only retrospective data is collected, Posttest Only With Control (POWC) (Collins and Holton, 2004). In practice, experimental designs have significant problems of implementation and run the risk of 'the tail wagging the dog' whereby the stringent requirements of evaluation take precedence over, for example, the needs-based selection of participants and individuals who could potentially benefit from a programme are excluded because they are allocated to the control group (Kearns, 2005, Preskill and Torres, 1999).

Quasi-experimental designs provide rigour over and above that enabled by pre-experimental designs, but do not go so far as random assignment. They may involve collecting data from participants who are yet to attend a programme and comparing this with participant data, and/or collecting data at multiple time-points before, during and after a programme, a design referred to as Single Group Pretest-Postest (SGPP) (Collins and Holton, 2004). Academics aligned with a positivist philosophy who favour experimental designs have historically criticised SGPP designs due to their weak controls and threats to internal and external validity. In practice, however, they are often used to evaluate training programmes (Carlson and Schmidt, 1999).

Surveys or so-called 'happy sheets' are the most widely used method of collecting data on in-programme, or immediate post-programme participant feedback (Tannenbaum and Woods, 1992) and capture Kirkpatrick's level one: Reactions. This is not fully reflected in the academic literature, which privileges more sophisticated studies of evaluation for their potential to contribute to theory development (Alliger and Janak, 1989; Arthur Jr et al., 2003). Questionnaires, semi-structured interviews, with participants and other key stakeholders, focus groups (Hannum, 2004) and a wide range of bespoke and pre-existing measurement instruments (see Collins, 2002) are frequently used methods whilst observation (Collins and Holton, 2004; Collins and Denyer, 2008; Hayes, 2007), 360 degree feedback (Rosti and Shipper, 1998), journal analysis, site visits (Russon and Reinelt, 2004), ethnography (Tanton and Fox, 1987) and repertory grid are amongst other methods that are deployed.

For outcomes at the system level, various objective measures may be used, such as employee retention, turnover and business results, alongside other measures collected as part of standard HR process, for example, employee satisfaction surveys and balanced scorecards (Sirianni and Frey, 2001). Triangulation of data sources is not uncommon, and is likely to yield more robust results.

Resource availability, capability, timescale and purpose all influence the choice of evaluation design and methods employed in any given circumstance. In turn, design and methods influence the degree of utility that any evaluation has in informing different types of decision.

Burgoyne (1975) argues that if the function of evaluation research is to improve decision-making, it should be designed as a process of feedback in a control system. In this conceptualisation, micro-level feedback concerning a specific programme should set out to generate evaluation data that can be captured and reported in a timely fashion. This may result in pragmatic, efficient designs that make methodological compromises. The second type of feedback should provide information to "check, test and improve...assumptions about things which are enduringly true of the education process" (p.54), contributing to a body of knowledge and helping to build a macro-level picture of effectiveness of different educational methods.

2.4   Meta-analyses

In addition to empirical studies reporting the effectiveness of individual management and leadership development programmes, several meta-analyses have been conducted that quantitatively aggregate the results of primary studies with the aim of assessing whether or not management and leadership development works, i.e. their primary purpose is to provide *proof* of impact (Easterby-Smith, 1994). They also allow relationships to be identified that are not apparent from the original studies (Arthur Jr et al., 2003).

A meta-analysis of organisational training (Arthur Jr et al., 2003) found a medium to large effect size. Whilst concerned with training in general, two of the research questions, regarding the relationship between training effectiveness and level of evaluation criteria applied, and between needs assessment and training effectiveness, are pertinent to evaluation of leadership and management. Training effectiveness was found to decrease substantially from learning measures through to behavioural and results measures, which was attributed to the social context and environmental favourability of the post-training environment. The findings with regards the relationship between needs assessment and training effectiveness were unclear.

Arthur et al.'s (2003) study and others, (Collins and Holton, 2004; Collins, 2001; Collins, 2002) followed Burke and Day's (1986) *A cumulative study of the effectiveness of managerial training,* which is widely regarded as the principal empirical support for the effectiveness of management and leadership development programmes (Collins and Holton, 2004). The study set out to address three related questions, about the overall

effectiveness of managerial training, the relative effectiveness of different training types (content/subject areas) and the relative effectiveness of different training methods.

The results were somewhat ambivalent, and led to the overall conclusion that 'managerial training is, on the average, moderately effective' (p.243).

Six managerial training types and seven training methods were specified. These were assessed according to four criterion measures, namely: subjective learning, objective learning, subjective behaviour and objective results. A summary of the results is shown in table 1 below.

Table 1. Effectiveness of managerial training: results of Burke and Day's (1986) study

Criterion measure	Training type effect	Training method effect
Subjective learning	Overall moderate effect across all training types; reasonable effectiveness for *Human relations/Leadership* and *Self-awareness* programmes	Positive effect for *Behavioural modelling*, *Sensitivity training*, *Lecture with discussion and role play/practice*, *Multiple techniques*
Objective learning	Overall moderate effect; *Improving motivation/Values* quite effective	Moderate effect for *Lectures*, and various combinations thereof
Subjective behaviour	Overall moderate effect; similar effect across all training types	Positive effect for *Lectures with discussion, role play or practice*, *Leader match* and *Behavioural modelling*
Objective results	*General management* training very effective; *Human relations/Leadership* training very effective, but caution advised given small number of studies (n=3)	Insufficient studies available

Collins and Holton (2004) built on Burke and Day's (1986) meta-analysis, reviewing 83 studies of managerial leadership development published between 1982 and 2001. They included Single Group Pretest Posttest (SGPP) designs in their study, and created a new categorisation of outcomes, building on the objective-subjective (Burke and Day, 1986) dimension and performance- or learning- level (Swanson and Holton III, 1999) distinction. This resulted in a six-type classification as shown in table 2 below, and was the first meta-analysis to attempt to measure behaviour objectively.

Table 2. Collin's and Holton's outcome categories

Outcome level	Meaurement type	
	Subjective (S)	Objective (O)
Knowledge (K)	K-S	K-O
Behaviour/expertise (BE)	BE-S	BE-O
System results/performance (SP)	SP-S	SP-O

Overall, the effectiveness of managerial leadership development programmes was found to vary widely. Some programmes had very large effects, whilst others "failed miserably" (p.232). There was more evidence of evaluation at the systems level than in

Burke and Day's (1986) meta-analysis, but learning outcomes were still the dominant measure used.

Knowledge–objective outcomes were found to be highly effective. There was a drop in effect size from knowledge outcomes to expertise and system outcomes, and the inadequacy of organisational/system-level evaluation methods, coupled with a small sample, led the researchers to question the validity of the system-outcome effect size. Expertise-objective outcomes were found to be moderately to highly effective across the research designs – posttest only with control group (POWC) and pretest posttest with control group (PPWC) were moderate, whilst single group pretest posttest (SGPP) designs gave highly effective results. Expertise-subjective findings were moderately effective across all three research designs. The overall moderate effectiveness of expertise-objective and -subjective outcomes mirrored Burke and Day's (1986) findings. System-objective measures showed a moderate effectiveness of training, but lower than the effect determined by Burke and Day (1986). No system-subjective results were reported and only a small percentage of studies reported outcomes at the organisational level.

Unfortunately, despite conducting a rigorous and expansive search, the number of empirical studies meeting Burke and Day's (1986) and Collins and Holton's (2004) inclusion criteria was relatively small. The analysis of results was thus hampered, particularly when categories were subdivided, and confidence in the findings is somewhat limited. The effects of moderating variables, such as quality of trainer and training content and organisational climate, was not incorporated into the analyses. Despite the competent use of meta-analytical techniques to enable comparison of studies and aggregation of data, the degree of explanatory power of the meta-analyses was small. The very nature of stripping down empirical studies to the extent that their findings can be integrated into a meta-analysis inevitably requires the sacrifice of richness of information. Thus meta-analyses are useful for *proving* (Easterby-Smith, 1994) overall effectiveness and relative effectiveness of different training types and methods, but make a limited contribution to answering 'why' questions (Denyer et al., 2008). A further limitation necessitated by the meta-analytical approach was the restriction on research design criteria. It is felt that a consideration of more qualitative designs would be a useful addition, albeit a difficult one to reconcile with the techniques of meta-analysis. It would enhance the depth of understanding about what works and why in different contexts (Denyer et al., 2008).

Despite their limitations, the two seminal meta-analyses described above were valuable for two main reasons. Firstly, they provided, for the first time, aggregated data about the effectiveness of managerial and leadership development programmes. For those commissioning such programmes, the moderate effect size found constitutes at least some evidence to support the decision to invest in management and leadership development interventions. Secondly, the meta-analyses provided templates for categorising variables such as intervention type, training method, research design and outcome levels. This is useful for other researchers in structuring future evaluation research studies.

2.5   Summary and focus for systematic review

Although a variety of models for assessing the effectiveness of management and leadership development exist, the field of evaluation is still dominated by Kirkpatrick's four-step model. Several of the alternative models attempt to reflect the complexity of evaluation in this arena by taking a holistic or systems approach whereby a whole range of variables are factored into predictive models. The diversity of models available reflects, to an extent, the varying purposes that evaluation can serve. Many researchers argue that the ultimate purpose of evaluation should be to support improved decision-making or a process of control.

Several attempts have been made to construct a synthesised view of the effectiveness of the field of management and leadership development through the use of meta-analytical techniques applied to disparate studies. To date, these have shown moderate evidence of effectiveness. Burgoyne (1975) argues that if the function of evaluation research is to improve decision-making, it should be designed as a process of feedback in a control system. In this conceptualisation, meta-analyses provide one type of feedback: macro-level data about educational processes used in management and leadership development in general. The other type of feedback relates to micro-level, context-specific programme evaluation.

As we enter the second decade of the 21st century, organisations globally continue to invest in the development of their employees, and the need for evidence about the effectiveness of management and leadership development remains – at both the macro and micro level. This need, coupled with changes in the emphasis of management and leadership development programmes, and the appearance of sophisticated evaluation models, suggests that a review of the most recent literature on the effectiveness of management and leadership development is timely in responding to macro-level questions. The systematic review that follows in Chapter 3. aims to achieve this and to provide a comparison of the latest literature with earlier studies. It also assesses the extent to which the field of management and leadership development evaluation has moved forward, and what salient areas for empirical research remain or have emerged.

CHAPTER 3. LITERATURE REVIEW PART II: SYSTEMATIC REVIEW

The two questions addressed by the systematic review were:

1. *How is the effectiveness of management and leadership development programmes assessed?*

   Which was compartmentalised into the following sub-questions:

   a. *What types of programme are subject to evaluation?*
   b. *What types of outcomes are considered important to measure?*
   c. *What evaluation designs are used, and what is the prevalence of the different types?*
   d. *What methods of data collection and analysis are employed in different situations?*

And:

2. *Are management and leadership development programmes effective?*

3.1   Scope

The systematic review focused exclusively on empirical studies published in 2000 and later. It can be seen as an extension of the seminal meta-analyses conducted by Burke and Day (1986) and Collins and Holton (2004).

3.2   Review protocol

The objective of a systematic review protocol is to set out a robust and transparent framework for the systematic investigation of the literature relating to the topic of interest (Tranfield et al., 2003): in this case, the effectiveness of management training and development programmes, and how evaluation is used. The review protocol also details the system used to select and appraise studies and the data extraction form where pertinent information about individual articles is captured. As such, the systematic review protocol is considered to be a helpful frame by which to organise the locating and storing of useful reference material.

The systematic review built on the foundations of two earlier meta-analyses that reviewed the effectiveness of management and leadership development. Burke and Day's (1986) *Cumulative Study of the Effectiveness of Managerial Training* is widely considered the 'principal empirical support for the effectiveness of managerial training and leadership development programmes'. Collins and Holton (2004) started their search of the literature where Burke and Day's meta-analysis left off, reviewing articles from January 1982 to December 2001.

Burke and Day's (1986) meta-analysis and Collins and Holton's (2004) follow-up study each had a slightly different focus, scope and selection criteria. Whilst following closely the search methodologies used in the previous meta-analyses, this review also had a slightly different focus. That is, it focused on formal management and leadership development programmes, as opposed to wider management development experiences, and it did not include or exclude articles on the basis of evaluation design as both previous analyses did.

The starting point for the search was 2000, i.e. a year before the end of Collins and Holton's (2004) search. The small overlap was a precautionary measure in case the most recently published articles when Collins and Holton conducted their search were not available.

The systematic review is a five-stage process (Tranfield et al., 2003). These five stages are further broken down into a series of steps, as outlined in table 3 below.

Table 3. Systematic review process

Stage 1 – Planning the Review
Step 1 – Forming a review panel
Step 2 – Mapping the field of study
Step 3 – Producing a review protocol
Stage 2 – Identifying and evaluating studies
Step 4 – Conducting a systematic search
Step 5 – Evaluating studies
Stage 3 – Extracting and synthesising data
Step 6 – Conducting data extraction
Step 7 – Conducting data synthesis
Stage 4 – Reporting
Step 8 – Reporting the findings
Stage 5 – Utilising the findings
Step 9 – Informing research
Step 10 – Informing practice

Source: RMC course notes on systematic review, and Cranfield School of Management doctoral web portal, accessed 31 May 2009.

Part I of the literature review comprised Step 2 in the planning of the review. The members of the review panel, formed during Step 1, are listed in appendix A. The

paragraphs below explain briefly how Stage 2, *Identifying and evaluating studies,* and Stage 3, *Extracting and synthesising data,* were implemented. They set out how the systematic search was conducted (Step 4) and studies were evaluated for quality (Step 5). The data extraction form (Step 6) is shown in appendix G. The review findings are the result of Step 7, conducting data synthesis, and Step 8, reporting the findings. They informed the final stage, *Utilising the findings,* through the empirical study (Step 9) which led to implications for practice (Step 10).

The search strategy was to use a rigorous approach to identify and select a relatively small number of key references pertinent to the field of study. This short-list of key references comprised the core literature body for the systematic review, and the identification of additional references either by cross-referencing from this core or through other means was legitimised through the application of the same quality standards that were applied to the references that comprise the core.

The databases identified for the systematic review were ABI Trade & Industry & Global/ Proquest, Business Source Complete (EBSCO) and CSA Social Sciences to access ERIC and PsychINFO. Databases were selected based on their coverage of key journals relevant to the field of management and leadership development. These databases also have extensive coverage of journals which, whilst not necessarily core to management and leadership development, cover fields where evaluation work has been conducted. PsychINFO, for example, is particularly valuable for articles relating to health care, a field where extensive research around evaluation of management and leadership development has been conducted, whilst ERIC specialises in education. Appendix B lists the databases used, with a brief description of each.

Four search filters were applied during the search: peer reviewed/scholarly articles only, published 2000 and later, abstracts only and not articles where the search terms appeared in the publication title, but not in the article title or abstract (Appendix C).

Two search strings were employed, the second to slightly broaden the range of articles returned by the first (Appendix D). The abstracts were then subjected to a series of inclusion/exclusion criteria (Appendix E) in order to refine the list that went through to the quality appraisal stage.

Articles that passed the tests of relevance and satisfied the inclusion criteria were taken forward for in-depth critique to assess their quality. Four aspects of each article were graded for quality: theory and context, methodology, results and analysis and contribution to knowledge (Appendix F).

Once the core list of articles that passed the quality appraisal process was identified, the reference lists in these articles were reviewed and further unique, i.e. not already found through database searches and potentially relevant articles identified. These articles

were sourced and subjected to the same quality appraisal as articles identified through the initial searches.

Endnote was used for storing data about selected references, its fields modified to include standard citation information, quality appraisal criteria and codes used in earlier studies by Burke and Day (1986), Collins (2002) and Collins and Holton (2004) and a framework for assessing evaluation outcome categories (Appendix G).

3.3   Summary of review findings

Table 4 below summarises the findings of the systematic review in relation to the review questions posed. This is followed by a discussion of the findings in section 3.4. Details of numbers of articles involved at progressive stages of the short-listing and quality appraisal process, where these articles came from and when they were published are shown in Appendix H, whilst a thorough and fully-referenced review is available in Appendix I.

Table 4. Systematic review questions and findings

Systematic review question	Theme	Finding
What types of programme are subject to evaluation?	Sector	Just under half of studies focused on the health sector; public sector featured more prominently than private sector, even with healthcare excluded.
	Customised v. open	Customised programmes in greater prevalence than open programmes, designed either for a single organisation or for a particular sector.
	Subject area/focus	Programmes focused on leadership development in all but three cases, and comprised a range of skills, knowledge and behavioural components.
	Programme design	Considerable variation in programme design and duration, but typically face-to-face workshops interspersed with activities to support and embed learning, such as project work and action learning; emphasis on applied and active learning.
What types of outcome are considered important to measure?	Type of measure	Subjective measures far more prevalent than objective measures for knowledge, behaviour/expertise and systems results/performance outcomes.
	Prevalence of outcome measures	Subjective behaviour/expertise most prevalent of the six possible outcome categories. Over half of studies measured systems level outcomes subjectively, but only a small number of studies attempted to measure objectively at this level.
	Number of outcome levels assessed	Large majority of studies measured more than one outcome category, most commonly two or three. No study attempted to measure all six outcome categories.
What evaluation designs are used, and what is the prevalence of different types?		Two thirds of studies employed Single Group Post-test Only (SGPO) designs; a quarter used Single Group Pre-test Post-test (SGPP) designs. Typically, SGPO designs were multi-method, and

		several were explicitly described as case studies. Only four studies used a control group, two in a Pre-test Post-test situation (PPWC), and two in a Post-test Only design (POWC).
What method of data collection and analysis are employed in different situations?	Data collection	Qualitative and quantitative data collection methods were used, often in combination. Qualitative approaches included interviews, focus groups and participant observation. Quantitative methods were primarily surveys and assessment instruments, a number of which were pre-existing and were often used for two purposes: participant learning and evaluation. Many of the evaluation studies were longitudinal, a feature inherent in Pre-test Post-test designs, but also featured in Post-test Only designs where data was collected at several points in time and at varying lengths of time after a programme.
	Data analysis	Data analysis methods congruent with methods of data collection. Generally, qualitative data was analysed thematically, with or without the use of software, whilst a range of statistical tests were applied to quantitative data.
	Design limitations	Limitations and compromises on the evaluation designs included: inability to isolate attribution, lack of clear measures of programme success, limits on generalisability and transferability, lack of long-term measurement and impact, over-reliance on self-report data, positive respondent bias and unclear influence of learning transfer factors outside of the evaluation remit.
Are management and leadership development programmes effective?	Positive results	Most of the studies showed a positive impact, with some showing high levels of programme effectiveness where, for example, all programme objectives were achieved or all assessment measures elicited positive results.
	Neutral and negative results	No programmes reported overall negative effects, though there were a couple of instances of negative elements within an overall neutral or positive impact, and a couple of studies resulted in inconclusive findings.
	Effectiveness relative to outcome level	Some evidence to suggest that lower levels of outcome, i.e. those related to knowledge and behaviour/ expertise elicited greater success than higher levels, relating to systemic outcomes.

3.4   Discussion of systematic review findings

A review of the post-2000 literature on the evaluation of management and leadership development programmes shows dominance of the public sector across published peer-reviewed papers and, specifically, the health sector. Just under half (n=14) of the review papers were concerned with leadership in health care. This sector skew is thought to be a result of the disposition in the health sector toward evidence-based decision-making, a requirement to demonstrate value for public investment and for evidence to be made available in the public domain, and perhaps the orientation of

researchers toward academic outputs. The private sector was dramatically under-represented. This could be due to the lack of incentive to publish evaluation studies whose intent is business-focused, e.g. to justify ongoing investment in a programme, or modify future iterations, and is not to seek acclaim from publication. Further to this, commercial sensitivity may be a good reason not to publish evaluation findings (Burgoyne et al., 2003), and when evaluation is conducted to satisfy a business need, drivers such as pragmatism and resource minimisation, perhaps coupled with shortage of research skills, may mean that the methodology and results fall short of the standard required for academic publication in any case. Finally, private sector underrepresentation raises a question about the extent of evaluation conducted in this sector in wider practice, and what sorts of research designs and methodologies are used to deliver *fit-for-purpose* evaluation (Hayward, 2009).

There was a dominance of customised programmes over open programmes. Despite the industry trend for more customisation and fewer open programmes, this is not thought to be the primary explanation for the difference. Rather, it seems that the trend toward single case-based designs comes into play. Many of the studies were conducted with a single programme under scrutiny and the focus of evaluation was on the extent to which that programme was effective, and in what ways, relating it back to the theoretical underpinnings of its design and delivery. This narrow focus is beneficial if the goal is to understand in detail what works and why in a particular situation, but it seems that an opportunity is being missed for comparative evaluations. Providers of management and leadership development should be concerned with the impact of their programmes, the advancement of pedagogic theory and practice, and in a strong position to evaluate the relative merits and impact of different programme types and designs. However, only two studies were conducted by business schools that made comparisons between programmes. Other studies were undertaken by academics that explored a particular theory or phenomenon where the primary purpose of the evaluation was theory testing and/or theory building, but they were not conducted in a comparative context.

The above discussion raises the issue of the role of stakeholders in influencing evaluation. Michalski and Cousins (2001) argue that whilst different groups might value a similar set of training outcomes, the relative importance placed on these outcomes as criteria for evaluation varies. Kim and Cervero (2007) go further, describing evaluation as inherently political. They further emphasise the significance of stakeholders, stating "Given the political nature of evaluation, the role of evaluators is likely to be entangled in power struggles among stakeholders" (p.7). In the review studies, a range of stakeholder dynamics was evident, but generally little, if any, attention paid to the relationship between the authors of a study, the commissioners, the evaluators and the programme designers and deliverers. In the minority of cases, all these relationships were made explicit. There were examples of programme deliverers taking responsibility for the evaluation and the writing up of a study in close liaison with the programme's commissioners, or academics working in partnership with commissioners to research a programme's impact. In a couple of situations educational researchers approached an organisation and requested access in order to evaluate a programme. In other cases the request was made from the other direction: commissioners of a programme solicited independent evaluators to report back on a

programme's impact. The stakeholder group which was, by and large, excluded from the design process was programme participants. Whilst participants may have benefited through enforced reflection or feedback from repeat 360s, this seems to have been a by-product rather than intentional evaluation outcome. The exceptions were the few studies that took an action inquiry stance, purposefully embracing participant reflection as a meaningful outcome of evaluation. It is suggested that more attention should be paid to making explicit the relationships between and involvement of different stakeholder groups in the evaluation process as a way of justifying an evaluation design or explaining the way data is interpreted.

An additional reason to pay more attention to stakeholders comes from Holton and Naquin's (2005) critical analysis of evaluation models. They argue that one of the reasons that evaluation models are not widely used is that most are derived from a rational-economic framework and that rational-economic models do not work in practice for making decisions. They set bounded rationality models in contrast, which have as "a foundational assumption that decision makers have neither the time not resources to conduct complex ROI-type evaluations" and "individuals use limited pieces of information to find a satisfactory resolution rather than an optimal decision" (p.265). Also favoured by Holton and Naquin are naturalistic evaluation models, involving collaborative, participatory and learning-oriented approaches, which, through the close involvement of organisational stakeholders, are most likely to result in a decision-making process which is natural for the organisation and leads to change. Evaluations which most closely resemble the idealised type for decision-making as described by Holton and Naquin (2005) involve multiple stakeholder groups in their design and delivery through various action inquiry, participative inquiry or action research approaches. There were several in this review.

The broad review of the evaluation literature that preceded this systematic review identified a range of models of evaluation. Perhaps unsurprisingly, given Holton and Naquin's (2005) critical analysis of theoretical models of evaluation, virtually no reference was made in the review articles to any of these models, or others. Only Kirkpatrick, the realist stance taken by Pawson and Tilley (1994) and Harper and Beacham's (1991) impact evaluation were explicitly discussed, and only in a handful of cases. Rather than aligning their design with a particular *model* of evaluation, other studies focused on a specific research approach, such as appreciative inquiry, intrinsic and instrumental case study or action learning, arguing that it could capture the complexity of a situation. It appears that whilst theoretical models, which attempt to disentangle complexity into the suite of variables which impact upon the success of a programme, have largely been ignored, certain methodologies applied in practice cover more or less the same ground. This suggests that the translation of theoretical models into practical methodologies requires more attention, or indeed that so long as the evaluators are proficient researchers and cognisant of the complexity of the situations they are investigating, the use of a theoretical model is not a prerequisite of a robust evaluation.

Research designs involving a control or comparison group were in the small minority, and the most popular design choice by far was Single Group Posttest Only (SGPO).

Whilst Single Group Pretest Posttest (SGPP) designs – and by implication, to an even greater extent, SGPO designs - have been criticised for their inability to control for maturation, history, testing and instrumentation effects (Campbell and Stanley, 1963), the review confirmed earlier research that showed that, in practice, such designs are often used (Carlson and Schmidt, 1999; Sackett and Mullen, 1993). It is suggested that a couple of reasons underpin this finding: Firstly, control/comparison designs are derived from a positivistic experimental mindset, whereas single group designs focusing on the study group only are more closely aligned with realist philosophy and explanatory intent. It has already been seen that most of the studies in the review were concerned with a particular programme, thus it is not surprising that SGPP and SGPO designs were in the majority. Secondly, there are substantial practical difficulties of manipulating a situation into an experimental one: ethical considerations of involving individuals who are not intended to benefit from a programme, and the nonsense of assigning managers at random to control or experimental groups without taking individual and organisational needs into account (Kearns, 2005). It is interesting to note that had the same evaluation design restrictions been applied as in Burke and Day's (1986) meta-analysis, only four of the 31 studies in the review would have been considered, or using Collins and Holton's (2004) selection criteria, only 12. This suggests that even in published evaluation studies, notwithstanding wider practice, there is a trend to move away from experimental and quasi-experimental evaluation designs. Whether practical considerations or a philosophical shift are the primary driver of the change, recent evaluation studies most commonly involved only programme participants and other stakeholders associated with the programme and no control or comparison group.

Programmes in the review were most commonly evaluated at two or three outcome levels (of a possible maximum of six) where levels were categorised as relating to knowledge, behaviour/expertise and systemic results/ performance and were either subjectively or objectively assessed (Collins and Holton, 2004). Amongst the handful of studies evaluating at a single level, none relied solely on knowledge outcomes. This implies two things: firstly, knowledge acquired is not a sufficient outcome of a management or leadership development experience and, secondly, knowledge acquisition is not a reliable proxy for programme effectiveness, demonstrating an appreciation of the derailing effects that stand between learning and change at behavioural or systemic levels. This implication supports the ambivalence arising from meta-analyses by Alliger and Janak (1989) and Alliger et al. (1997) that found at best moderate correlations between Kirkpatrick's levels of outcome. As stated above, studies in this review most commonly evaluated at multiple outcome levels. This contrasts with several earlier meta-analyses that found a single level of outcome analysis to be most common. Perhaps this is again related to the intent with which recent evaluations have been approached, with more emphasis on holistic designs. Alliger and Janak (1989) and Alliger et al. (1997) found that whilst a single level of evaluation was most common, this was *not* level one, knowledge. Collins and Holton (2004) in turn found that behavioural outcomes were found in the majority of studies. These findings are supported by this review, but are in conflict with industry studies conducted in the 1960s and 70s (Catanello and Kirkpatrick, 1968; Kirkpatrick, 1978) which showed a clear prevalence of knowledge evaluation only. The discrepancy we are

seeing here is further reinforcement of the suggested gap between those research studies which get published in academic journals and the wider field of practice.

Subjective outcome measures were twice as prevalent as objective measures. In situations where programme objectives are unclear, and there is only a vague notion of the correlation between training and performance, it may be easier to opt for subjective accounts of improvement or change than objective ones which require respondents to be precise about the extent to which specific attitudes or behaviours have changed. Thus the demands on programme commissioners and designers are less onerous if subjective measures are to be used. Perhaps evaluators are doing them a disservice in the long run by not challenging them up front to articulate the specific objectives of a programme and how its design is intended to deliver these. On a practical point, often the subjective assessments are made by participants, an easier group to access due to existing relationships and contact points, than a wider group of stakeholders. Subjective measures are not, however, necessarily easier to implement and analyse than objective ones. They are often more time-consuming to interpret and synthesise data from because they tend to be more open and wide-ranging. They can also provide more valuable and relevant data. For example, objective measures were frequently made using standard measurement instruments but there are questions to be considered about the extent to which these instruments are relevant to a specific programme and whether, by using an instrument with a specific focus, other important developmental aspects of a programme are ignored.

The review identified that a multiplicity of evaluation approaches, designs and methods have been used in recent literature on the effectiveness of management and leadership development programmes. In each case, the resultant evaluation is a consequence of its purpose, who the stakeholders driving it are and what resources are available, which lead into choices about design, what outcomes are measured and what methods are used to collect and analyse data. This reflects the complex and context-dependent nature of management and leadership development evaluation.

3.5   Comparing effectiveness with earlier studies

A substantial majority of the management development programmes reviewed had positive effects, confirming Burke and Day's (1986) conclusion that "managerial training is, on the average, moderately effective" (p.243) though less specific than Collins and Holton's (2004) assertion that "practitioners can attain substantial improvements in both knowledge and skills…" (p.217). A small number of studies provided neutral or inconclusive evidence, or negative elements within an overall positive outcome, but none resulted in negative outcomes. This contrasts with Collins and Holton's (2004) meta-analysis that showed a wide range in the effectiveness of management and leadership development programmes, with some programmes that "failed miserably" (p.232) and had a negative effect size measure. It cannot be said for certain why this difference occurred. Possibly programmes are becoming better; the evaluation designs and subjective outcome assessments used enabled positive findings to be identified, even if they weren't the ones the programmes set out to achieve; and/or perhaps studies with inconclusive or negative results were not written up, the "file

drawer problem" (Rosenthal, 1979), or looked favourably upon by journal reviewers and were never published.

The review provided limited evidence that knowledge/learning-objective outcomes were highly effective and effect size dropped when moving to behaviour/expertise and system results/performance outcomes. This supported Collins and Holton's (2004) findings. The review also, mirroring Collins and Holton's experience, makes cautious claims about system-level outcomes due to the small sample size and inadequate measures of system results/performance outcomes. Having said this, the number of studies attempting to evaluate at organisational level had increased since Collins and Holton's (2004) work. The raw numbers indicate a growth from 11 to 16. In this review, most of the systems level evaluation attempts used subjective measures, often relying on qualitative data, so it might be that this explains the growth since Collins and Holton's work: they found no examples of system level subjective measures, perhaps because they couldn't reconcile the use of qualitative data to assess subjective systems outcomes with their meta-analytic method. Notably, given the lack of definitive evidence in this review and Collins and Holton's (2004) meta-analysis, Burke and Day (1986) found objective systems/results criteria to be very effective for certain training types, such as general management, though they warn caution in the interpretation of other training type categories given the small number of studies included.

Putting too much weight on comparisons with Burke and Day's (1986) and Collins and Holton's (2004) meta-analysis should be avoided since the methodology evolved from study to study. Burke and Day, for example, subcategorised managerial training into six types, a division which would have been redundant in this study where articles focused almost exclusively on leadership. Burke and Day also identified different training methods, which again would have been a meaningless categorisation since almost all studies involved multiple teaching and learning methodologies. Further, the outcome categories used by Burke and Day were expanded by Collins and Holton to incorporate two additional categories: objective behaviour and subjective results, and it was Collins and Holton's categorisation that was used in this systematic review. Perhaps the most significant difference between the studies, however, is the way in which effect size was measured. Both earlier studies used meta-analytic techniques, whilst in this review a subjective assessment was made of the extent to which outcomes had been achieved. Whilst meta-analysis confers clear benefits, the danger of using such techniques is, as previously discussed, that high quality research designs that do not fit an experimental mindset are excluded, and that any analysis has limited explanatory power since it reduces complex phenomena down to numbers.

Particularly if the trend toward more holistic, explanatory evaluation designs is real, it is suggested that the time for meta-analysis of the sort used by Burke and Day (1986) and Collins and Holton (2004) has passed, and the focus from a research perspective going forward should be on increasing the explanatory dimension of evaluation. This should support the ongoing deepening of understanding of the relationship between context, intervention, mechanism and outcome (Denyer et al., 2008). Put simply, the relationship between what works and why in different circumstances.

3.6 Implications for empirical work

The systematic review mapped the latest trends in the conduct of evaluation of management and leadership development. It identified the dominance of public-sector studies, particularly health, and a focus on customised leadership development. Subjective outcome measures were found to be twice as prevalent as objective measures and outcomes tended to be measured at multiple levels. Overall, most programmes were found to have positive effects, with stronger relationships apparent between development programmes and knowledge and behavioural outcomes than system-level results/performance. Many studies emphasised using evaluation to understand the dynamics of specific programmes and what contributed to or inhibited their impact, perhaps because repeated attempts to make the connection between development interventions and organisational performance have resulted in, at best, moderately helpful measures.

The review contributes to the body of evidence about the effectiveness of management and leadership development as a whole. This addresses one of the two requirements for evaluation research identified by Burgoyne (1975) if evaluation research is to support improved decision-making about education and training. The other requirement is for timely information about individual programmes. It is a concern that the review was strongly public-sector biased as this leaves open the question as to the extent of evaluation practice in the private sector and the extent to which timely evaluation data is being used to support decision-making.

Also concerning was the lack of attention paid to stakeholder dynamics – and therefore the interests and needs of different stakeholder groups in order to use evaluation data constructively - and the paucity of reference to theoretical models of evaluation. Despite their apparent explanatory power, such models appeared not to be widely recognised nor adopted: Evaluation theory may not be supporting practical action as effectively as possible. One plausible explanation for this is the criticism that whilst evaluation should have decision-making as a core output (Hesseling, 1966; Warr et al., 1970; Swanson and Holton III, 1999; Russ-Eft and Preskill, 2001), most evaluation models are not well-suited to the ways in which organisations make decisions about development interventions (Holton and Naquin, 2005) and this is why they are rarely used. Evaluation models have tended to come from the education domain, paying scant attention to the body of knowledge on organisational decision-making.

Holton and Naquin suggest that one avenue for future evaluation research to address this problem is to "conduct naturalistic decision research to understand how HRD decisions are really made in organisations" (p.277). This echoes a much earlier call (Burgoyne, 1973) suggesting that attention be paid to studying decision and control processes in education systems. Given the gaps apparent from the systematic review, this appears to be a valuable exercise.

The decision was made, therefore, to focus the empirical research for this study on how decisions are made in practice about management and leadership development, and on

what role programme evaluation plays in the process. Given the lack of data from the private sector in recent published evaluation studies, a further decision was taken to limit the scope of the empirical work to the private sector. At a time when demand for management and leadership development remains high, but alternatives to 'traditional' programmes are emerging as a result of technological, economic and environmental change, identifying how evaluation might better support informed decision-making about investment in this field was felt to be an important endeavour. It was hoped that the empirical work might also reveal whether Holton and Naquin are right about the limitations of existing evaluation models for decision-making, and whether, perhaps, evaluation needs to be re-conceptualised.

The research questions for the empirical work were:

- *How are decisions about management and leadership development programmes made?*
- *How does evaluation contribute to these decisions?*

CHAPTER 4. RESEARCH DESIGN, SETTING AND METHODS

Research design is about *informational adequacy* (Zelditch, 1962), i.e. "organising research activity...in ways that are most likely to achieve the research aims" (Easterby-Smith et al., 2002, p.43) and *efficiency*, whether the research plan allows for adequate and efficient data to be collected given practical constraints of time, access and cost (Zelditch, 1962). Underpinning research design is the researcher's philosophical position. In this study, the researcher holds a relativist ontology situated in the constructivist paradigm (Denzin and Lincoln, 2000). In practical terms, this confers a preference for understanding the perspectives of key research informants, recognising that 'facts' are dependent on their position, whilst assuming that particular practices and structures exist and that it is possible to identify these (Easterby-Smith et al., 2002).

Choices made in the research design were to recognise a degree of researcher involvement, bounding this by using standard protocols for data collection, to use relatively small numbers of research informants in a fieldwork setting and to work toward theory development (Easterby-Smith et al., 2002). The purpose of the study was descriptive and explanatory (Marshall and Rossman, 1995). It sought to understand how processes of decision-making occur in organisational settings, and to understand the specific role of evaluation in these processes. In order to achieve this, an abductive research strategy (Blaikie, 2007, p.10) was employed, whereby the perspectives of research subjects about evaluation and how they believed it influenced decisions made about management and leadership development programmes were explored. This 'bottom-up' approach was followed by a 'top-down' approach of relating the research subjects' perspectives to extant evaluation theory (Blaikie, 2007).

The study focused on a number of individual management and leadership development programmes, in each case incorporating multiple key stakeholder perspectives in order to strengthen validity. The design was intended to allow common themes and patterns to be identified within the study, whilst recognising a limitation on the universal generalisation of its findings.

4.1   Setting and sample

The unit of analysis for the study was individual management/ leadership development programmes. In total, eight programmes were studied. The rationale for investigating at the programme level was that it would enable depth of understanding to be developed about each programme by triangulating data from two or three stakeholders, each of whom would bring a different perspective. Their individual perspectives could then be compared and contrasted to see where there was congruence and where there was difference, and what might underpin these differences. The stakeholder-based approach is similar to that used by Kim and Cervero (2007), whereby this approach "views that evaluation is generally organised around the questions posed about the programme by those who request and commission the evaluation and other pertinent stakeholders" (p.7).

Programmes for inclusion in the study were identified through an iterative process of seeking and awaiting approval from client directors at Ashridge to approach their clients.

Ashridge is the researcher's employer and was selected due to the access afforded to data subjects. It is an independent business school and one of the leading providers of management development in the UK. Ashridge offers open and customised programmes, organisation consulting and post-graduate qualifications ranging from Certificate through to Doctoral level. Provision of private sector customised programmes is the biggest segment in Ashridge's portfolio.

A degree of consistency was sought by only approaching client directors of private sector companies that had spent a minimum of £50,000 on a management or leadership programme with Ashridge in the past two years which involved more than one cohort of participants. The sector choice was made on the basis that evaluation as part of the audit trail was felt to dominate the rationale for evaluation in the public sector, and there has been a paucity of recent published evaluations from the private sector. The time frame of the last two years was applied so that interviewees would be more likely to be able to recall details about the programmes in question and identifying and reaching them in the first place would be more feasible.

40 client organisations were included in the selection phase, of which 10 reached the shortlist after either the client director or key client contact of the other 30 declined to participate. Of these, two were excluded from the analysis – in one case because only one interview was achieved and in the other because the programme was too early in its lifecycle for interviewees to provide sufficient data, and this had not been detected prior to interview. 21 interviews relating to the eight remaining programmes were included in the analysis.

The programme name, the letter assigned to each client company and the job roles of the research subjects are shown in table 5. below.

Table 5. Interview schedule

Company	Programme Title	Provider interview	1st client interview (key client contact)	2nd client interview
A	Leading as a [Company A] Manager	Client director	Programme manager	Senior business executive
B	European Leadership Programme	Client director	Programme sponsor	-
C	Leaders of the Future	Client director	Programme manager	Senior business executive
D	Navigator	-	Programme owner	Programme manager

E	Power 2 Lead	Client director	Programme owner	Programme manager
F	Executing Strategy	Client director	Programme manager	Head of learning and development
G	Aspiring Leaders	Client director	Programme manager	-
H	Strategic Leadership Programme	Client director	Senior business executive	Former programme manager

4.2   Methods

4.2.1   Data collection

For each programme, data was collected through semi-structured/ non-standard (Healey and Rawlinson, 2004) interviews with key stakeholders associated with the programme, supported by document analysis. This type of interview is suited to exploratory and explanatory research (Saunders et al., 2007). Three interviews for each programme were sought, as far as possible, with: the key client contact, usually the programme manager of the programme in question, the Ashridge client director and a second key stakeholder of the programme as identified by the key client contact. Two semi-structured interview protocols were developed, with a substantial degree of overlap; one for clients and one for the Ashridge client director, to reflect their different positions in relation to a programme (see appendices J and K). The protocols were used to retain a reasonable level of consistency across the interviews, but were used flexibly to accommodate different levels of interviewee knowledge about a programme and to allow for rich seams of inquiry to be explored in more depth. Most interviews were conducted by telephone, due to the international location of many of the interviewees and practical constraints of time and budget, though some were conducted face to face at Ashridge. Interviews lasted between 30 and 45 minutes. All interviews were recorded and transcribed verbatim by the researcher.

Documentary secondary data includes written data in the form of meeting notes, minutes, reports and can also include non-written materials, such as recordings of meetings (Saunders et al., 2007). For this work, documentary secondary data was limited to written documents. Documents were used to 'corroborate and augment' (Yin, 2009) the interview data and were read as part of the interview preparation process. This also minimised the time required during interviews on gathering background information and helped establish the researcher's credibility with interview subjects as it demonstrated thorough preparation prior to the interview. Documentary secondary data included tender documents, signed annual programme contracts between client and provider, programme timetables, participant lists and evaluation reports. This documentation was gathered through the electronic shared filing system held by the provider, Ashridge. In a few cases, documents relating to evaluation that had not been

available on the Ashridge shared electronic filing system and were discussed during interviews were sent afterwards by the interviewee.

4.2.2   Data analysis

21 interview transcripts were imported into NVivo 8 qualitative analysis software. As advised (Tagg, 2010) a small number (n=4) transcripts were free coded initially, selected on the basis of reflecting variation, i.e. different stakeholder roles and representing different programmes. The free nodes were used to construct a skeletal coding tree, and the free coding was subsequently removed from the four transcripts. The coding tree was then used as the framework for coding eight transcripts, and was added to with additional nodes as coding progressed. At this point, the tree was re-structured and coding categories re-grouped to better reflect the data.   Categorisations from the evaluation literature were used to structure two of the coding categories. The Kirkpatrick (1998) four-level framework for evaluation level and Easterby Smith's (1994) classification of evaluation purpose. These categories were used after identifying that themes around evaluation level and purpose were emerging from the data, i.e. the data confirmed and justified the use of existing frameworks. All other categories were derived exclusively from the data. The coding tree is shown in Appendix L.

A vignette about each of the programmes under consideration was written as a descriptive, multi-perspective synthesis of each programme. The vignettes were developed by integrating the themes and using illustrative quotes from the two or three interviews and supporting documentation about each programme. Multiple stakeholder perspectives added validity to the findings and allowed for differences in stakeholder perspectives to be identified and shown. These are not positioned as 'case studies' because, whilst multiple sources of data were used (interviews and secondary documentary evidence) as advised by case study researchers (Yin, 2009, p.114) the breadth of evidence gathered was insufficient to justify classing the vignettes as cases.

4.3   The role of the researcher in the study

The researcher's stance was as an 'outside learner' (Blaikie, 2007, p.11). At the point of contact with research subjects, the focus was on recording and understanding their perspectives rather than asserting pre-existing views. Inevitably, the researcher interpreted what research subjects said through a filter of experience: almost ten years working in the field of management and leadership development. This experience was considered to be beneficial to understanding the context of the interviews and it created a good rapport with research subjects. To maximise the chance of research subjects providing honest responses to questions posed, the researcher explained how their data would be used, assured all research subjects anonymity and confirmed that their company would not be identified. Researcher bias was controlled for by working within the framework of a standard interview protocol and by meticulous transcription and methodical analysis of data.

CHAPTER 5. FINDINGS

The findings are presented in four sections: Section 5.1 presents descriptive data about the eight programmes included in the study and summarises the evaluation methods deployed across these programmes; section 5.2 is a series of vignettes which describe the detail of how decisions are made and the role of evaluation in each programme; section 5.3 draws together the findings from the eight programmes and discusses how decisions about management and leadership development programmes are reached; and section 5.4 focuses on the role of evaluation in contributing to these decisions.

5.1   Descriptive data and summary of evaluation methods

5.1.1   Descriptive data

The programmes ranged from accelerator programmes for high potential relatively junior managers, through to senior management programmes. Group size was typically in the mid twenties, with a range from 18-30 participants. In terms of design, some programmes were delivered in a single residential block, whilst others were multi-modular spread over a period of up to nine months, with the exception of one programme which spanned an 18-month period. Several programmes involved inter-modular learning with varying degrees of structured support for this, and work-based projects either during or after the residential element had finished. The newest programme was first run in March 2010 whilst the longest established had been running since December 2005. There were variations in the number of programmes run each year, from one to five.

Table 6. Programmes included in data collection process

Company	Programme name	Target audience	Group size	Programme design	Programme history
A	Leading as a [Company A] Manager	Senior management	20-25	5-day programme	First programme January 2006, 5 in 2006 & 7, 4 in 2008, 3 in 2009, 2 in 2010
B	European Leadership Programme	High potentials moving from management to leadership roles	21-28	Was 2 modules. Now 3 modules: 2 x 3 days followed by 3.5 day live case	First programme started December 2005, and run approximately annually since
C	Leaders of the Future	High potentials	24	3 x 3-day modular programme	First run in 2009, one programme scheduled for 2010

D	Navigator	Middle - senior management	24	18-month programme: development centre followed by individual phone coaching	First programme started December 2007, one programme in 2008, three in 2010
E	Power 2 Lead	Senior management	25-30	Modules 2 and 4 5 days and held at Ashridge. Preparatory work (module 1) and between-module work (module 3)	Pilot in 2006 - 2007, two programmes in 2007 and 2008 and one in 2009. No programme in 2010
F	Executing Strategy Programme	Middle -senior management	18- 21	Pilot: 3-day then 2-day module; Now 2 x 3-day modular programme	Piloted December 2009, first programme February 2010, 5 cohorts to run in 2010
G	Aspiring Leaders Programme	Middle management	c. 25	7-day programme	First programme March 2010, second programme scheduled November 2010
H	Strategic Leadership Programme	Senior management	24-25	2 x 5-day modular programme	Started in 2007. Last run in 2008. Intention to run again in 2010/11

5.1.2   Summary of evaluation methods

As described in detail in section 5.2 below, each of the eight programmes was subject to some formal evaluation. At a minimum - and the case for all programmes - an end of programme review form was distributed to participants to gather reaction-level feedback about a programme. A range of other evaluation methods were also employed, encompassing post-programme interviews with participants and line managers, reporting back by HR participants, focus groups, repeat 180 or 360 feedback processes, project presentations and noting of participants' career progressions. A number of planned or aspirational future evaluation processes were also described, typically toward the higher end of the range of outcome categories, based on an adaptation of Collins and Holton's (2004) categorisation of outcome categories.

The only evaluation theories or models cited were Kirkpatrick's four-level framework and return on investment (ROI), though no specific reference was given for ROI.

Table 7 summarises the actual and intended (I) formal evaluation processes in place for each programme and the levels of outcome assessed. It can be seen that reaction-level feedback was captured in all cases, and between one and three additional categories of outcome were measured, with intentions for more levels to be incorporated.

Table 7. Evaluation methods and levels of outcome assessed

Company	End of module/ programme review forms	Post programme interviews with participants	Post programme survey	Post programme interviews with line managers	Repeat 180 or 360	Presentations on work-based projects	Focus groups	Participant promotion	HR participants report back	Future plans for evaluation	Reaction	Knowledge - subjective	Knowledge - objective	Behaviour/expertise subjective	Behaviour/expertise objective	Systems results/performance subjective	Systems results/performance - objective
A	✓		✓							Study impact of all leadership programmes on company over longer time frame	✓	✓		✓		I	I
B	✓	✓		✓						More structured data gathering process	✓	✓		✓			
C	✓								✓	Retention/turnover, job satisfaction, promotion, better integration with appraisal process	✓					I	✓
D	✓	✓				✓					✓			✓	✓	✓	
E	✓	✓	✓	✓					✓	Participant interviews, performance assessment, 'leadership index' assessment, promotion staff satisfaction	✓	✓		✓	I		I
F	✓									employee engagement survey, direct reports' feedback, interviews with participants at 3 & eight-10 weeks post programme, control group to compare business plans	✓	✓		I		I	I
G	✓					✓				Ongoing implementation of learning, management & leadership competence, potential for moving into more senior roles	✓	✓	I	✓	I		✓ I
H	✓					✓					✓			✓	✓		

33

5.2 Programme vignettes

Short descriptive accounts are provided below of each of the eight management and leadership development programmes included in the study. The 'vignettes' cover the decisions made throughout the programme lifecycle from conception through to commissioning, delivery, review and subsequent repeat, or postponement, with a particular focus on evaluation and how it was used. An exemplar of the coding structure used, applied to the vignette for Company A, is shown in Appendix M. A summary of the themes emerging from the company examples is presented in section 5.3.

5.2.1 Leading as a [Company A] Manager, Company A

The CEO of Company A chairs a Capability Council, of which one of the four capabilities in focus is leadership. Beneath the Capability Council is the Leadership Forum, comprised of senior HR and business leaders, which sets the overall agenda for leadership development and has been the instigator of a new leadership curriculum. Individual programmes are overseen by a reference or steering group which reports to the Leadership Forum and, ultimately, the Capability Council. The programme manager for *Leading as a [Company A] Manager* has the wider role of Global Leadership Capability Manager and is responsible for all programmes in the leadership curriculum.

These changes are relatively recent, and at the time that *Leading as a [Company A] Manager* was instigated, company A's Leadership Academy would commission programmes on request where a gap or business need had been identified. In the case of *Leading as a [Company A] Manager* a senior HR manager *"did the rounds" [client director]* with a colleague to identify potential providers. This was followed by a formal tendering process and a presentation to HR and line managers, from which a preferred provider was selected. The same senior HR manager who had led the provider identification process was influential in this decision.

The current curriculum of leadership programmes is designed around leadership 'elements' (competencies) for leaders at specific transition points in the organisation. *Leading as a [Company A] Manager* was identified as a programme suitable for supporting leaders as they transition into roles where they will be leading other managers and/or global functions. As a consequence, the specificity with which candidates are selected has increased, although the population from which programme participants are drawn is largely the same as before. Another reason for greater emphasis on the selection process is that the programme had become highly popular, and a new HR Vice President recognised that requests to participate were driving selection, rather than it being a strategic choice. The selection policy was changed when the original programme sponsor retired.

"Individually, these people got a lot out of it, but they were perhaps not the right people any more...They are not part of a group that has a lot of future. They are good performers but they don't have necessarily the potential to go on."[Client director]

The programme manager, together with a steering group, decides what the programme should focus on and any substantive re-design issues. Minor alterations are dealt with by the programme manager and client director. The steering group comprises HR and business managers, including a senior business sponsor, and has become less formal than when it was first set up. The senior HR team is kept informed of any changes.

The first iteration of the programme was run as a pilot. After the pilot, and subsequent iterations of the programme, changes were made based on participant feedback from the programme review forms, conversations with participants, and the personal experience of the programme manager and faculty team who were present throughout the programme. More feedback was solicited from participants of the pilot than of later programmes, though the same range of data sources were used. The programme manager or another member of the HR team always attends the programme. In the early iterations, a representative of the Global Leadership team and an executive board member were present throughout. Now, senior executives attend and take a role in delivering the programme. The CEO, amongst others, has commented that this is the *"best programme ever" [Client director]* and past participants who have been promoted to more senior roles in the company have also been strong advocates.

Evolving business needs are anticipated to be an ongoing reason for adapting the programme. Two reasons were given for why the programme might be terminated in the future: the participant population has been exhausted, or an alternative training programme or resource requirement has taken priority. In terms of deciding to keep running the programme, it appears that the visibility of and to senior managers is very positive, and important:

"It makes a big impact. It's a very natural flow....That is part of the success: the fact that we have senior management and they open up." [Client director]

Evaluation

"Happy sheets" [Programme manager] are distributed at the end of each programme. The provider's standard programme review form is used, which is similar to the form company A uses for other training programmes. Participants rate how well the programme met its stated objectives, the value to them as an individual and the extent to which they would recommend the provider. They are also asked to rate individual sessions and they have an opportunity to write comments. Participants' scores are aggregated and a summary produced which is distributed by the programme administrator to the faculty team and the programme manager.

The results from the *"happy sheets" [Programme manager]* give the programme manager confidence that the programme is going well. When the scores are high, as they have typically been, little is done with the results. It appears that closer attention is

paid when other factors, e.g. the recent focus on transitions as part of the new leadership strategy, prompt a need to redesign the programme:

"When we went to redesign I went into the survey results to see where the highlights are and where we can see that it's not that good, the results are a little bit lower and we see what we can do, see where we can change it." [Programme manager]

The programme manager also has informal conversations with participants during and after the programme about what they thought about the topics covered and how they might take them further.

A survey was conducted in 2007 of participants on all five 2007 iterations of the leadership programme. This was initiated by the programme manager at the time. Participants were asked questions about the impact of the programme on their self-perception, their behaviour in specific areas and on the company; and specific questions about post-programme contact with line managers, colleagues and other course participants. It was not clear how the results of this evaluation were used and what, if any, influence they had.

There is a desire from the HR side to pay more attention to demonstrating value:

"What we do need to do, I think, is to have a stronger relation about how there is value for the business, the return on investment for programmes like this...As a whole we need to look at the effect that the [leadership] curriculum has on the whole organisation and the development of the organisation over a longer time frame". [Programme manager]

From other quarters there is scepticism about the purpose and value of evaluation:

"I also think...these people have to give themselves a reason for existence. And they think that having a lot of papers, a lot of evaluation helps. But I think it's true: it's when the CEO meets someone who's been on the programme and says, 'so how was it?' I think that's the more valuable feedback." [Client director]

Summary

Evaluation of *Leading as a [Company A] Manager* comprises feedback forms distributed at the end of the programme seeking reaction-level feedback and a survey focusing on impact was conducted in 2007, though it is not clear how the data from this was used or how influential it was. The feedback forms are used to check that participants are satisfied with the programme, and to support design modifications. Other contributors to design modifications are changes in the leadership strategy, informal participant feedback and firsthand experience of the programme. The programme has high visibility with senior managers in company A and one of its features of perceived success is the involvement of these senior figures in delivering the programme. The programme sits within a well integrated support and management structure.

5.2.2   European Management Programme, Company B

When Company B approached the provider to tender for the work it was a new spin-out company, having split off from its long-established parent company. With plants located across Europe, a need was identified for a programme that would develop potential plant and functional managers and broaden their horizons:

"We wanted... [to] get people thinking about the company strategy, the company's views on leadership as well as external best practice and get them a bit of exposure outside their own function or plant and working with people in different parts of the company." [Senior sponsor]

The relationship the client director has with the European President with and HR sponsor, has been and continues to be important. When a new President joined Company B a couple of years ago, the client director developed a strong relationship with him in order to persuade him of the potential to use the EMP to *"set the standards, set the tone" [Client director]* in the company, and rather than developing generic skills, to focus on addressing company issues. The client director is committed to trying to shift away from the concept of 'a course' to the proposition that working with the provider can provide a genuine opportunity for change in the client organisation. He believes that raising awareness with any client of what is possible, and then designing a solution that delivers value and is visible at the highest level in the organisation is the key to client loyalty:

"...either in the development process – if we're lucky – or once we start, we say, 'There's an opportunity to be much more useful to your organisation', and that's when we get loyalty." [Client director]

"First of all it's awareness. This is a massive point. Most people are not aware for the potential of a training course for their managers to help them change the organisation." [Client director]

The programme underwent a major re-design in light of the client director's influence on the President and is now focused around the strategic priorities of the organisation. There has also been a shift toward emphasising the leadership element since a *"more consistent path" [Senior sponsor]* emerged about the strategic direction of the company and the HR sponsor recognised that high potentials were struggling when put into senior jobs without being equipped with the necessary leadership skills.

From the HR sponsor's perspective, participant feedback about the programme is very important. This, in combination with line manager feedback, business issues and discussion with the client director influence ongoing modifications to the programme design. He suggests that more structure around what evidence is used in deciding to modify the programme would be desirable:

"I'm probably ashamed to tell you it's not a very scientific process at the moment.
Maybe as we're staffing up we'll get a bit more structure and process and fact base."
[Senior sponsor]

The client director believes that the criteria clients use to decide whether a programme is delivering value depends on how the programme is positioned and perceived:

"If they go on a course, they might say, 'Thank you very much, that was a good course',
and that's it. But if they go through a course with somebody who's curious about their
business and obsessed about delivering more value...then we might be able to persuade
the client to do more than the course and then they kind of discover religion and start
doing it that way. But I think we've still got lots of clients that go on 'a course' and it's
really vulnerable to being pulled in tough economic times". [Client director]

In the case of the EMP, senior managers are involved in delivery and discussion on the programme, and participants make individual and organisational commitments for how they are going to make changes back at work. The client director believes that this visibility and demonstration of value to the top management of the organisation has been central to the decision by Company B to keep running the programme:

"It comes much closer to something that can deliver return on investment, and they say
'this is fantastic: let's keep going!' And ever since I've done that, it's been an automatic
re-schedule in next year's diary."
Interviewer: "And when people say, 'this is fantastic!' what are they basing that on?"
"They're basing it on their Chief Exec turning up and hearing his people talking
confidently, intelligently and in an informed way about how to improve the business."
[Client director]

When asked what might prompt a decision to stop the programme, hypothesised responses were: a change in the overall programme of training and development, internal customer demands and participant feedback on how useful the programme had been. Any such decision would be made by the HR sponsor together with the learning and development director and senior management team.

Evaluation

The standard provider programme review form is distributed at the end of each module. It captures scores and comments on each session in the programme and an overall score on the extent to which the programme met its objectives and its value to participants. The provider sends the HR sponsor and learning and development director the feedback and the HR sponsor would normally share this with the management team.

At the end of a programme, the HR sponsor also appeals directly to participants for feedback:

"I plead or yell or threat – take your pick! – with the delegates to email me or phone me
or come up and see me with their considered views about what they felt about the

programme and what they'd like to see differently in future programmes for future generations." [Senior sponsor]

The HR sponsor asks informally for feedback from delegates and their line managers, about changes in participants' thinking and behaviour, three to six months post-programme. He speaks to the team that has been involved in delivering the programme and pays a lot of attention to the client director's input. He would, however, like to add more structure to the feedback gathering process in the future:

"That [gathering feedback] is one of the areas I'd like to formalise a little bit more." [Senior sponsor]

Though not positioned as formal evaluation, the response of the management team of the plant hosting the 'live case' suggests the programme's impact:

"The management team hosting the live case found the project findings useful, as well as the exposure to the talent." [Senior sponsor]

Summary

Company B uses programme review forms to evaluate its *European Management Programme* and make changes to the design accordingly. Informal conversations between the HR sponsor and client director, participants, their line managers and with the senior managers involved in the programme are also important influences on programme design and on how the EMP is perceived to generate value. Value to the organisation is something the client director is passionate about delivering and demonstrating. This is manifest through the live case, and the CEO and senior managers being closely involved in the programme and seeing for themselves how participants are talking about and planning for change in service of business improvement. The client director believes that the way in which programmes are perceived, whether as 'a course' or as an enabler of competitive advantage, is central to the way in which the client decides whether a programme is delivering value.

5.2.3   Leaders of the Future, Company C

Leaders of the Future was developed after a new Corporate Training Manager joined the business and reviewed existing provision for management and leadership development. Working with an existing company concept of three levels of management, she recognised the need for a programme for the first level: First time, high potential managers. The need was seen partly as an investment in talent management and partly for succession planning.

The Corporate Training Manager and her manager, Head of Corporate Training and Development, visited a number of business schools to gain a general impression. During these visits, the provider's faculty helped the clients to clarify their thinking around what they needed and discussed potential solutions. The provider – among other

business schools - was then invited to tender. This invitation was based on a range of factors:

"If you go back to the basics, it was a lot about credibility, chemistry, facilities."
[Programme manager]

For the final selection, the Corporate Training Manager invited a wider group of HR heads to participate in the decision-making process. These individuals would ultimately be the customers of the programme; the people nominating and paying for participants, and the corporate training manager felt it was important that they felt involved and engaged in the choice of provider and programme design. The corporate training manager nevertheless had a significant influence over the final choice. She presented the decision back to the top management, inviting them to comment, but suggesting that the programme should be progressed and developed.

She worked with the provider's faculty to develop the programme design. As part of the design phase, she led a round of interviews with key stakeholders from the business to identify where the *Leaders of the Future* programme should focus and to ensure their expectations were aligned with the design.

The corporate training manager is responsible for the ongoing smooth running of *Leaders of the Future*, amongst a range of other programmes. She approaches the executive committee directly with requests for help when she needs to, and keeps her manager informed. She is well networked and leverages the lack of rigidity in the way the company operates:

"They don't seem to be so process-formalised. It's fairly informal. She is the contact person; she knows a lot of people. If something happens behind the scenes, it's probably informal." [Client director]

A quota of places is allocated to each of the business streams within Company C. The HR head of each stream nominates participants after discussion with the overall stream and other senior managers. Participant nomination is tied into the succession plan for the business stream.

The Corporate Training Manager attended the first programme as a participant and drew on her experience during the review meetings with the client director after each module as well as using data from the programme review forms to inform design modifications. The first module went well and the review process was straightforward:

"We had a good review. It went very well, extremely well, with very good ratings in every respect...There was no major disaster, and therefore no major discussion."
[Client director]

Module two didn't go as well, for a variety of reasons, and the weak points were reflected in the review forms and the experience of the client director and corporate training manager:

"I was in the classroom watching, observing and I was listening to other people, sharing the opinions." [Programme manager]

This put more pressure on the provider for the third and final module:

"There was a decision point...so we all knew that if we did not perform, there would be a problem." [Client director]

Module three went very well, *"like a dream" [Client director].* The client director and Corporate Training Manager then reviewed the whole programme. The programme review forms informed the review, with particular attention paid to comments where a participant had given a low overall score. The client director also introduced some new ideas for content.

In preparation for running the second iteration of the programme, the corporate training manager conducted *"update interviews"* with a couple of senior stakeholders. She did these to ensure continued alignment of the programme with current business issues and organisational changes, and to keep key stakeholders engaged. She attended the second and third programmes as an observer. The monitoring and review process was similar to that used to review the first programme, i.e. personal experience and participant feedback informing a review meeting after each module and at the end of the programme.

Even during the recent challenging economic times, the corporate training manager was given the green light by the executive team to continue with the programme. She believes that attracting and retaining talent is particularly important in the industry in which Company C operates and that being an *"employer of choice" [Programme manager]* involves maintaining a focus on staff development. There are no plans at present to stop running the programme, and hypothesised reasons include a change in personnel and loss of the personal relationship between the client director and corporate training manager, or the participant population reaching saturation:

"What tends to happen is personalities change. Particularly on the client side, that's when the programme stops. Or it's run its natural course: you've got 300 managers and two or three years later you've covered the population." [Client director]

Evaluation

Programme review forms are distributed at the end of each module. These assess participants' reaction to individual sessions and their overall impression of the programme.

In terms of longer term measures, the corporate training manager has been paying attention to job promotions:

"We started as 24 [participants], and by the middle of the programme...we had already nine moves within the organisation..." [Programme manager]

"Now we are keeping in touch with these people informally, because I wanted to put up some sort of statistics." [Programme manager]

She is clear that these job moves are associated with the appropriate selection of high potential candidates for the programme, rather than being causally related to participants having attended the programme. However, she believes that it may have helped them on their way:

"If you pinpointed the right quality of people at the initial stage, it's really a career boost." [Programme manager]

In terms of additional evaluation, she sees the long term progress of programme participants and their loyalty to staying with the company as key indicators of programme success:

"For me, the next indicator of success would be if these people actually stay with the company, if they are happy in their current jobs or their new jobs and if they are viewed as potential successors." [Programme manager]

The senior HR manager of one of the business streams speaks to participants on an informal basis after their attendance on the programme and expects to see an improvement in on-the-job performance, though this is not formally captured or reported. He has a vision for a more formalised process for integrating evaluation with the appraisal process, not just for *Leaders of the Future*, but for other development programmes, though this has not yet been put into practice.

Summary

Great attention is paid to stakeholder engagement, relationship development, alignment with the business strategy and participant selection for *Leaders of the Future*. Evaluation comprises end of module review forms, complemented by the personal experience of the corporate training manager and client director when it comes to reviewing the programme. Longer term indicators of programme success are anticipated to be the retention of participants in the company and their promotion and contentment with their work. Participant promotion is thought to reflect selection of the right calibre of candidate, rather than the assumption being that the programme caused their promotion, though it is felt that the programme is a contributing factor in their success.

5.2.4   Navigator programme, Company D

When Company D slipped out of the Sunday Times Top 100 Companies listing a few years ago, there was a concerted effort to address and improve people management skills. The listing confirmed results from staff satisfaction surveys that people

management capability could be improved. The Managing Director set up a working party tasked with looking at people management, and this group concluded that there was a particular issue for engineers who had finished their professional qualifications and become chartered. There was no structured career support at this point, and many staff left or lost motivation.

The focus of a management development programme was initially to address this population directly. However, a new Group Learning and Development Manager re-joined the organisation, and was adamant that it was necessary to address management at the top of the organisation if programmes lower down were to have any effect. As such, two programmes were developed: *Navigator*, for more senior managers, and *Horizons*.

A handful of top UK business schools were invited to tender for the *Horizons* programme. The Group Managing Director and Group HR Director made the final choice. The decision to choose the same provider to deliver *Navigator* followed as a consequence, to ensure that the two programmes were closely aligned.

The Group Managing Director, Group HR Director, Group Learning and Development Manager and programme manager formed a steering group for *Navigator* and *Horizons*. They were joined later by a newly appointed board member who had taken on responsibility for Learning and Development. The steering group was heavily involved in designing the programmes and in their internal advocacy. The role of the steering group has since receded, meeting only when big decisions need to be taken, and the programme manager deals with day-to-day programme management in conjunction with the provider's senior account manager.

When *Navigator* was designed there was an expectation that it would have a limited shelf life and be curtailed when the target population had been through the programme. This is still the plan. With *Horizons* being available to more junior managers, there was also the expectation that this would provide managers with the people management skills they would need as they progressed through their careers. By the time they reached the level at which they would be eligible for *Navigator*, they would already have the requisite level of people management competence.

Evaluation

"Reactionnaires" are administered at the end of the initial development centre which are *"pored over" [Programme owner]*. Informal conversations with participants, the direct experience of the programme manager and other observers and current business issues in Company D are also used to inform design modifications to the development centre.

For the pilot, gathering data was taken particularly seriously, not only to improve the development centre for the next cohort, but to maintain the buy-in of participants:

"It's really important if you want to make sure that the programme will be successful that the first group, if they've experienced anything they don't like, feel that they are completely involved and know what's happening and what things are going to be changed. They're the budget holders and otherwise they won't agree to spend the money." [Programme owner]

There were a number of formal and informal discussions with the pilot group after the development centre both to gather feedback and to inform them of design changes for future cohorts.

In summer 2009 an evaluation aiming to demonstrate the business impact of the first two iterations of the programme was conducted by Company D's learning and development department. Telephone interviews were held with participants who had completed the programme three and six months previously to ascertain what they had done differently as a consequence of being on the programme. Participants shared their action plans, discussed how much they had actioned and were asked about their perception of the programme's business impact. All participants interviewed reported having made major behavioural changes and a number had made structural changes in the way their department operated. Aggregate results from the repeat 360 were also analysed and the overall data suggested that participants had *"moved forward" [Programme owner]* but the exercise was primarily designed for development, not assessment. Up until this point, it had been difficult to assess impact due to the confidential and individual nature of the programme.

The impetus for the *"impact evaluation" [Programme manager]* came from the Group Learning and Development Director. A one-page summary of the evaluation was presented to the Board by the board-level Learning and Development sponsor. No feedback was received, nor expected:

"We will obviously have to evaluate it properly one day but at the moment everybody seems to be happy with anecdotal evidence, actually and it was like that from the beginning when we were designing it." [Programme manager]

As a consequence of this evaluation, there has been a design change in an effort to improve the support for participants. Participants will now be invited to participate in telephone action learning sets. *Horizons* uses action learning, and it has been seen to work successfully there. Some participants of *Navigator* are action learning facilitators on *Horizons* and, based on their experience, were advocates of introducing action learning to *Navigator*.

The programme manager has developed her personal view of the programme's success based on personal contact with participants:

"Just from chatting to people all the time. I bumped into someone a couple of weeks ago and he told me that he'd actually changed his life...It's all individual evidence, but we've had feedback such as: People are listening more, people are having better

communication with their teams, people are meeting more, some people have re-structured parts of their business..." [Programme manager]

"There's a collective gut feel that it's working, but we don't have any... quantitative data to support it, only qualitative data." [Programme manager]

Summary

The confidential and individual nature of the *Navigator* programme meant that initially evaluation was limited to feedback forms about the development centre and anecdotal feedback from participants. The development centre was modified on the basis of this feedback, as well as the direct experience of the programme director, other observers and the need to incorporate current business issues. An evaluation was conducted in summer 2009 to explore the impact of the programme. The findings prompted the introduction of action learning to the programme. The decision to end the programme is expected to be taken when the target population has been exhausted, which should coincide with the next wave of managers coming up through the organisation who have already had an opportunity to develop people management skills through the *Horizons* programme.

5.2.5   Power 2 Lead, Company E

Company E comprises the corporate holding and four subsidiaries of which three are product-focused and the fourth is called Solutions and contains the learning and development department. In 2004, the current programme owner, who is situated in the corporate holding, initiated and managed a discussion with senior managers in the individual businesses to decide whether there should be an advanced international management development programme and, if so, what it should look like. Having reached agreement that there was a need, a paper was prepared for the executive board outlining the costs and how it fitted into the company's career development strategy. The board approved the proposal and several business schools were invited to present their recommendations for the design. HR and business managers from each of the subsidiaries were involved in the process. Three business schools were shortlisted and asked to give another presentation to HR and the business. The presentations were followed by reference checks, and, after some time, different providers were selected for two programmes: *Power 2 Lead* and *Power 2 Manage*. Representatives of Company E then requested to meet more faculty and staff and sit in on a number of teaching sessions, and only at the end of this, formally appointed the provider. One of the influential board members at the time had not heard of the provider and this is thought to be, at least in part, a reason that a different, known, provider was chosen to provide one of the programmes.

A new programme manager for *Power 2 Lead* joined the learning and development department of Company E in 2007 and took on the day to day management of the programme. She works with the programme owner from the corporate holding who retains a substantial involvement in the content and running of the programme and attends at least a day of each *'venture'* (iteration). The programme owner is the key

client contact for the provider and her wider remit is as Director of Management Development Programmes.

After each module, there is a feedback and review session with the client director, a second key faculty member, the programme owner, her assistant and the programme manager. These review sessions have become shorter and less intensive as the programme has become more established.

Power 2 Lead will not run in 2010, primarily due to insufficient participant nominations, which may be an extension of the managers of business units – which pay for participants' places - wishing to save money. It is expected that there will be sufficient nominations for the programme to run again in 2011. The programme owner will discuss the future running of the programme with the head of HR and the executive board, with the intention of securing a formal signoff from the board.

It is anticipated that the programme will stop when the target group has been through the programme, and/or if the company changes and its priorities change. A new CEO is starting in 2011 – the board member who had not heard of the provider when *Power 2 Lead* was first started – and his influence is an unknown factor.

Evaluation

Feedback forms are distributed at the end of modules two and four. The data is collated by the provider and sent to the client. The feedback form covers the between-module learning activities, module content, module learning environment, quality of administration and service and the e-learning portal.

Company E's learning and development department also sends a feedback form two months after the programme and interviews participants to inquire into their overall satisfaction with the programme and impact of the learning in the workplace. Participants' line managers are debriefed at the end of the programme, and interviewed a couple of months later about whether they have noticed any improvements in participants' work. The results of these interviews are sent to the programme owner who discusses them with the senior HR team and addresses any areas of programme underperformance, discussing this with the client director where applicable:

"We expect high satisfaction and whenever there is feedback that you would say is on the average not that good, we then focus on these topics and see what we can improve."
[Programme owner]

The programme owner also speaks to participants:

"I also personally talk to some of the participants in every venture...to see what has happened and what the atmosphere was and in the overall evaluations if there are any ratings that are not satisfactory we focus and discuss what we can improve. Continuous improvement actually." [Programme owner]

When an HR participant attends the programme s/he is expected to keep a diary with notes about his/her experience and discuss this afterwards with the programme team.

A summary of the programme evaluation is presented to the board along with evaluation of *Power 2 Manage* for the purposes of demonstrating that the programmes are running well, are relevant and to keep the buy-in for future investment in management development. The board requested that participants write down their key learnings from the programme, but the value of this exercise seems to be in some doubt:

> *"Hardly anyone wrote critical feedback...This [the key learnings summary] was presented to the board. I doubt they had time to read it." [Programme manager]*

Prior to the current evaluation process, there was an even more extensive evaluation component, but this has since been reduced:

> *"We found out that actually this kind of state-of-the-art and very intensive evaluation is, we didn't have the resources to do all that, and we found out that a little less is also adequate." [Programme owner]*

> *"One question is always if there is a balance between the effort and the outcome...For this programme it made sense, because it was so extensive, but now we've reduced that, because really it was a great effort to organise all this and co-ordinate it." [Programme manager]*

The programme owner also recognises the importance of informal feedback processes:

> *"I mean, usually formal and informal evaluations go in the same direction, they have the same outcome actually, and the informal [laughs] are very often quicker." [Programme owner]*

She uses the end of module feedback results and review meetings with the client director both to make ongoing improvements to the programme, but also so as not to be caught by surprise by feedback that has gone up the organisation informally and come back. The programme manager also values the informal:

> *"Even with all these evaluation processes I still think it's very important as well to have informal talks. Because in a personal conversation with participants, you get many details out of it." [Programme manager]*

There is an expectation that in the future there may be some additional measures put in place to track whether programme participants have become *"more professional on managing on an international level and international cultures" [Programme owner].* There is an expectation that interviews might be used, along with performance assessment of senior managers and the company's externally assessed *"leadership index" [Programme owner]* - which is a pre-requisite for any manager seeking promotion to a top management position. Other data sources anticipated to be used are

staff satisfaction surveys, statistics on staff promotions and the impression of the senior management of changes in the leadership culture.

Summary

Every iteration of *Power 2 Lead* has been subject to considerable evaluation, in the form of comprehensive end-of-module feedback forms and interviews with participants and/or their line managers. Although formal evaluation is still taken very seriously, the intensiveness of the evaluation has been reduced. It was felt that good data could still be generated with less resource investment, especially after the early iterations gave evidence of the programme's success. Informal feedback is also important, and felt to reflect the formal, but is generated more quickly.

5.2.6   Executing Strategy programme, Company F

Several parallel strands of work led to the creation of the *Executing Strategy* programme: a move to develop a consistent approach to leadership development across Company F, the identification of strategy implementation as an area of potential weakness, and an initiative on levels of work which set out to identify what was expected from each level in the organisation. It was decided to focus a development programme on the level in the organisation whose prime responsibility it was to deliver strategy. The strategic planning process was driven by the chief executive, working closely with the HR Director.

The programme manager for leadership then took responsibility for scoping the programme and identifying potential providers. He started with a diagnostic phase, during which he reviewed existing data, e.g. an employee engagement survey and a leadership behaviour survey, and interviewed the direct reports of a number of the candidates likely to attend the programme.

The programme manager and his boss had worked in the past with the two providers who were ultimately commissioned to co-deliver the programme, and approached both organisations, and others, informally to discuss their ideas for a programme. This was followed by a formal tendering process and an online 'reverse auction'. Two providers were selected and invited to jointly design and deliver the programme. The decision-makers in the selection phase were the programme manager, his boss – the head of learning and development, and her boss - the head of organisation development. The decision was signed off by the head of HR and senior HR managers were kept informed throughout. The same group worked closely with the providers on the detailed design and redesign that followed the pilot.

The programme manager sits within the learning and development function. He manages the programme with the client directors from the provider organisations and keeps his boss involved. The new head of OD signs off any substantive design changes.

The first module of the pilot programme did not go well and feedback forms elicited some "*pretty dreadful*" scores, *"the worst I've ever seen" [Client director].* The

48

programme manager and client directors felt that some of this was due to a bad atmosphere in the business at that time due to other changes, and not to do with the programme itself, but they nevertheless took the feedback very seriously. One of the client directors also thought the atypical composition of participants, unusual MBTI mix, didactic style, inexperience of co-delivery and problems around contracting with participants contributed to the poor feedback. His views were partly based on firsthand experience, partly on the feedback forms, and partly on feedback during and after the programme from observers – including the head of learning and development who had attended the programme as a participant.

In light of the feedback and experience, the programme manager and client directors carried out focus groups with participants to understand their feedback from the first module in more depth and inquire into what they wanted to focus on in module two. They used the focus group data to adapt the design of the second module, and to completely redesign the first module before the next iteration of the programme.

Company F has committed to running the *Executing Strategy* programme for at least three years so that it becomes embedded in the organisation. Factors that might cause the programme to be stopped include: poor participant satisfaction scores, cost, budget cuts, changing priorities for leadership development, whether value is being seen from the programme and the parent company instigating a different approach to leadership development. The decision to stop the programme would be made by the head of OD, head of HR and the CEO.

Evaluation

Feedback forms, which are an amalgamation of the two providers' programme review forms, are used at the end of each module. The forms gather reaction-level feedback about individual sessions and tutors and prompt participants to reflect on their key learning for the day and how they intend to action it.

Company F has also commissioned one of the providers to design and deliver a comprehensive, multi-modal evaluation of the programme, expected to be implemented over 12 months. The initiative for this came from the programme manager, supported by the head of learning and development and head of OD. The evaluation will include an employee engagement survey, direct reports' assessments of behavioural change, interviews with participants at three weeks and eight-10 weeks post programme, and use of a control group for comparison with the study group to compare the quality of business plans produced. An online closed-question survey may replace interviews in some cases, depending on cost.

It is anticipated that this evaluation will be used to demonstrate to the executive team that the programme is delivering value: that there is improved staff engagement, the benefits of line manager support and improvement in the quality of business plans. Currently, the executive team is not as closely connected with the programme as the head of learning and development would like, and there is no regular reporting to the team. Another purpose is to send a signal to participants of the importance placed on the

programme, and the expectation that they will act on their learning. The head of learning and development wonders whether participants should, in fact, be held more accountable for improving their leadership performance, once the relevance and quality of the programme has been demonstrated through existing evaluation mechanisms.

The programme manager is experienced in evaluation and an advocate of its importance:

"I think it's hugely important. People do always wonder why L&D, not just leadership development, is one of the first things to get cut because people don't invest enough time in the ROI piece for me." [Programme manager]

He is curious and open to finding out if there are other, unanticipated, outcomes resulting from the programme:

"I'm also a firm believer that it may reveal something completely different." [Programme manager]

In addition to the formal evaluation, the head of learning and development is noticing changes in behaviour:

"We've started to see individual groups in the programme starting to take initiatives to connect with people...which was one of the objectives of the programme...And we've had, but again it's anecdotal feedback, evidence of people talking about behavioural changes that they have seen in people coming back from the programme." [Head of L&D]

She feels, however, that whilst anecdotal data is often sufficient, and people are generally realistic about the limits of what can be measured, formal evaluation is important:

"Beyond the post programme evaluation, the feedback and evaluation is currently quite ad hoc and anecdotal, and that's why [the programme manager] is doing this work to formalise that, because if we're going to continue to get the mandate to continue to do it, you have to prove the value, that this investment is worthwhile." [Head of L&D]

Summary

Company F's *Executing Strategy* programme got off to a poor start. Feedback from end of module feedback forms, post-module focus groups with participants and the direct experience of the programme delivery team led to a major re-design. Since then, feedback scores have improved dramatically and anecdotal evidence is beginning to indicate participants making behavioural changes subsequent to the programme. A comprehensive evaluation has been commissioned by the client that will be implemented over a 12-month timeframe. It is expected that this evaluation will be used to demonstrate the value of the programme to the executive team and be used to

make the case for continued support for the programme, and may elicit other, unanticipated, outcomes.

5.2.7   Aspiring Leaders programme, Company G

The Global Head of Training and Development (GHTD) joined company G 18 months ago with the remit of setting up a global career development framework, to include management development. This part of the framework is called the *Core Management Initiative*, and is sponsored by a member of the company's executive committee. The GHTD established a steering group of 12 senior managers from around the world to look at how management development should be implemented. She is highly influential in guiding the steering group. The steering group defined three distinct levels of management: introductory, middle and senior. They decided to start by focusing on the middle management population. The GHTD wrote a brief and invited six business schools to tender. Two were shortlisted and one won the final *"beauty parade" [Client director]*. Prior to this, several managers had been on an open programme run by this provider and their positive feedback gave the provider credibility and company G confidence in their capability to deliver a practical and high quality programme.

Participants are nominated to the programme by the Chief Executives of each region, in liaison with the regional HR teams, based on a set of clear eligibility criteria and the expectation that the candidates have the potential to become *"leaders of the future" [Programme manager]*.

The programme has high level sponsorship, with the Chief Operating Officer chairing the steering group and reporting back in great detail to the Chief Executive and board. On a day-to-day basis the GHTD makes decisions about the design and running of the programme.

Factors anticipated to cause the programme to be stopped are: if Company G merged with another company, if a new board member with strong ideas about learning and development joined the company, environmental concerns and changes in travel policy, and if the programme did not live up to expectations in terms of participant reaction and outcomes:

> *"If we didn't deliver a superb programme [it might be stopped]. There's such a spotlight on the programme. It gets evaluated looking downwards, upwards, sideways...I think the pressure will be on us in terms of inspiring people around management...Persuading [the professional staff] to move into management is a core theme of the programme, I would say [the chief executive] would be looking for a change in the business around that this year and I think we will come under pressure if the programme is not delivering that." [Client director]*

Evaluation

An extensive online evaluation survey was sent to participants toward the end of the programme, which the global head of training and development strongly encouraged them to complete, and a 100% response rate was achieved. The survey was designed and the process managed by Company G, rather than the provider. The survey covered reaction-level feedback about each session in terms of pace, quality, content and the delivery standard of each faculty member, and then posed questions about how participants intended to use the content and how the learning would change their daily work. At the end were questions about the service experience, including food, logistics and the venue. The rationale for such a detailed evaluation was that this was the first time the programme had been run:

"It was quite a labour intensive form. But because it was the first time we were running it, I needed to get that initial feedback". [Programme manager]

The programme feedback was excellent, with a couple of exceptions, and these areas have been addressed for the next programme by the Global Head of Training and Development in conjunction with the client director. The survey results were presented back to the steering group by the GHTD and were sent to the chief executive.

The next stage of evaluation will be when participants report back to the steering group in September 2010 on their team projects. Both the success of the project and the performance of the individuals will be assessed by the steering group.

The steering group is also considering how to assess participants' ongoing implementation of learning from the programme, their management and leadership competence and their potential for moving into more senior roles. The group will meet after the second iteration of the programme in November, to agree how to do this. The GHTD intends to impress on HR managers the importance of line manager follow up with participants. She suggests that line managers should be discussing how the participant has changed in his/her role, and reporting this back:

"You'll be able to see if someone has changed in their role. And then you'll be able to see, say in five years time, do they come through and are they our leaders of the future?" [Programme manager]

The GHTD recognises that feedback forms are useful in the short term:

"I think in some ways the happy sheets and how people came away is quite a telling sign of how the programme went",

whilst longer term evaluation in the form of tracking participants' progress is more for personal satisfaction and for future reference:

"I think you just build up a knowledge bank."[Programme manager]

In addition to the feedback forms and project review, the CEO requested a meeting with the client director to discuss how the programme had gone and what the expectations were of how participants would benefit as a result. Prior to this meeting, he had spoken to a sample of participants and compared their feedback with what the client director said. The client director was also asked to write a report outlining what the programme is delivering and how participants are benefiting, which will be reviewed by the steering group and then the executive committee and CEO. The client director anticipates that a high level of scrutiny and being held to account by the CEO and steering group will continue for future iterations also. She also senses that the conversational aspects of reporting – to the steering group and CEO - are more influential than the data from feedback forms for a programme of this level.

Informal feedback processes have also generated evidence about the programme's impact,

"We've had people email our global chief executive to say what an amazing programme it was, we've had people come up to me, we've had people who've gone up to more senior people in the business to say it's the best thing they've ever done. It's as though the rumour went round quicker in terms of how brilliant this programme was."
[Programme manager]

Already there have been almost twice as many nominations for the next programme than there are places on it, again suggesting the reputation of the programme has become established.

Summary

The *Aspiring Leaders* programme has been developed and managed with a high level of professionalism from the outset. Both the programme and the individual participants are subjected to a high level of scrutiny, from the top of the organisation, in an attempt to ensure that value is being delivered to the organisation. Feedback is captured by an online survey at the end of the programme, the client director has a conversation with the CEO and had to write a report for the steering group, executive team and CEO after the programme, and participants will report back to the steering group on the outcomes of their group projects six months after attending the programme. There is also an expectation to track longer term participant performance and promotion.

5.2.8　Strategic Leadership Programme, Company H

Prior to the current provider running the *Strategic Leadership Programme*, a similar programme had been run by a different provider for a number of years, but had lapsed. The decision to re-start the programme came at a time when the company was growing internationally and the top management team wanted to develop a more internationally focused and better networked group of senior managers. They also wanted to achieve better alignment of strategic thinking, and a shared understanding of leadership and management concepts.

Two HR managers visited a number of business schools, including the provider - at which several managers had attended open programmes in the past. The provider was favoured because of its more applied and less academic delivery style and because it offered the best degree of customisation. A third influencing factor was that a senior HR manager had studied her MBA there and was a strong advocate of the organisation. Three senior HR managers were responsible for selecting the provider.

The client director and programme manager from Company H conducted a diagnostic process to inform the design of the programme, which included speaking to prospective participants about what should be included. This feedback informed and was informed by conversations with the executive team about priorities for the programme, and senior executives were asked to contribute if the priorities coincided with their areas of interest and expertise. The HR business partners were also asked for feedback where applicable.

In terms of selecting participants, an HR business partner from each of the three business divisions liaised with senior executives in their respective division to put forward candidates. The programme manager engaged with other senior executives to select candidates who were not located in one of the business divisions, with the senior vice president of Global HR making the final decision on the selection priorities for this group.

The ongoing fine-tuning of subsequent iterations of the programme resulted from written participant feedback, the judgement of the programme manager and client director based on observations during the programme, and informal conversations with participants. It was also strongly influenced by the programme manager's ideas and the current needs and priorities of the business:

> *"They weren't reflective at all. 'What are the main business issues this year?' 'What are the main strategic themes this year?', 'These are some of the things that the programme should try to achieve'. So, very much prospective rather than retrospective." [Senior sponsor]*

The *Strategic Leadership Programme* has not run since 2008, though the Executive Committee has committed to running it again in 2010 or 2011, possibly because the new organisational strategy launched in January 2010 has 'talent' as a core focus area. The Executive Committee decided to postpone the programme due to the global economic crisis and the need to cut costs in all areas of the business, and also partly for symbolic reasons as the programme was the most expensive of all Company H's leadership programmes:

> *"Symbolically, they...said, 'well, if we're asking everyone to tighten their belts and we're laying off 10,000 people...we have to share the pain". [Client director]*

Senior managers had also given feedback that whilst they would love to attend the programme, the time commitment required was too much at a time when they needed to be focusing on urgent business challenges.

The programme manager (who is no longer working for Company H) believes that instability in corporate HR due to organisational re-structuring, staff changes, and a consequential loss of history about the programme and lack of advocacy for it, could serve to stop it altogether. He thinks this is problematic not just for the *Strategic Leadership Programme*, but for the overall alignment of strategy and leadership. He fears that there may not be the level of engagement between HR and the executive committee to push forward the learning and development agenda in service of delivering the new organisational strategy effectively.

Evaluation

The provider's standard programme review form was used at the end of each module, incorporating individual session ratings, ratings of the degree to which the programme met its objectives and was of value to the individuals overall and inviting comments. The programme administrator collated the scores and comments and sent them to the client director and programme manager. During the post programme review meetings, this feedback was used to inform the discussion.

In addition to the programme review form, Company H designed a two-part questionnaire focusing on learning transfer to be used as a 180-degree feedback process. The questions related to eight - 10 areas in which impact was anticipated to occur from the programme. These areas were also aligned with the company strategy and leadership competences. The first part was distributed to participants immediately after the programme and the questions focused on anticipated application of learning. The second part was sent out three-four months after the programme and asked participants to answer the questions on the basis of the extent to which they had improved their skills in the same eight - 10 areas. Their line managers also answered the questions and both sets of data were presented back to the participant and line manager for them to use in their *"continuous dialogues" [Former programme manager].*

The results of this evaluation were reported to the Leadership Board, chaired by the programme manager and comprising senior managers from the business units and the service/operational functions of the organisation. Board members also had an opportunity to input their ideas for the programme going forward.

The results were also used to check that the programme was delivering outcomes to an agreed minimum level. The one area - strategic communication - that scored lower than the set standard was discussed by the programme manager and client director and the programme design was changed to improve this area.

The programme manager invited participants to come and talk to him during and after the programme and made time to debrief and solicit feedback from the members of the executive committee who delivered sessions. He believes that both formal and informal evaluation approaches are important, particularly in a global organisation where cultural norms affect what people are prepared to write down:

"I would say that we couldn't do without either of the two, because there are certain things you can't capture in a formalised evaluation and that you will only get by engaging with people." [Former programme manager]

Another senior stakeholder designed a generic evaluation form that was used with all of Company H's leadership programmes. It asks participants to give an overall rating of the programme, ratings for programme content, facilitation and hospitality services and to state whether or not they would recommend the programme. He combined data from the *Strategic Leadership Programme* with data from other programmes to give overall ratings of the various leadership programmes run by Company H and constructed a multiple regression model to show which factors had the greatest and least influence on participants' overall perception. He believes that the implementation of this evaluation form has changed the way in which internal trainers and consultants work, and rather than competing, they have begun to collaborate as they are now assessed on teamwork and team performance rather than individually.

Whilst significant effort was invested in evaluation, the impression of two individuals closely involved with the programmes was that there wasn't a huge demand for evaluation data from the senior management of the organisation, although it was linked to performance-related pay for some training and development staff:

"I didn't get a strong sense that there were really deeply interested in strict metrics and ROI from the programme. I think it was more if the content looked good, and the delegates had a good time." [Client director]

"I don't think any of the decision-makers involved in this have shown any interest in the former evaluations right now. Not at all." [Senior sponsor]

There was an expectation that informal feedback processes would work effectively, because of the close relationship of participants with the senior management:

"To be honest, because they [the participants] were so senior, that's a fairly efficient kind of quasi market. If the programme is crap, or there's a big gap that's missing or that's wasting time, I think those conversations would have gotten into the ears of the key decision-makers pretty quickly." [Client director]

Summary

Company H's *Strategic Leadership Programme* has lapsed since 2008, when it was postponed due to financial pressures on the business. Since then the programme manager has left the organisation, and ownership of the programme – including evaluation data from previous iterations - is unclear. The programme had been evaluated in several different ways, and this information used to inform design changes to the next iteration. The impetus for evaluation came from the learning and development function, rather than senior management, who seemed to place greater emphasis on ensuring relevant content and objectives for forthcoming programmes than on reviewing the impact of past programmes.

5.3 Key decisions and how decision-making about management and leadership development programmes happens

Typically, the lifecycle of all the programmes followed a process of: identify need, select provider, design detailed programme, implement, review, modify and repeat and, in some cases, postpone/terminate. The vignettes in section 5.2 above describe how, from the perspectives of two or three key stakeholders, decision-making happened for each of the programmes included in the study. Sections 5.3.1 – 5.3.6 below summarise the salient features of the decision-making process for each of the key decisions across the different programmes at each stage in the programme lifecycle.

5.3.1 Deciding to commission a programme

Two main strategies dominated in the way in which programmes were initially conceived of, labelled as business-led and HR-led. Company A, for example, exemplified the business-led approach:

"Usually...there comes initiatives from the organisation, 'This is an area where we need to develop, this is where we want to go' and we help them to set up learning activities to be able to reach their target." [Programme manager, Company A]

In contrast, Company C pursued more of an HR-led strategy:

"It's a three-level management concept...You have to deal with first time managers who are just starting to do something, mid-level managers who have been doing their job pretty well but still you see if there are areas for improvement, and then managers who are very well advanced in their careers..." [Programme manager, Company A]

Often the identification of need for a programme coincided with new staff joining the organisation, whether a learning and development professional tasked with the job of setting up a more structured management development process and keen to make an impression, or senior management whose joining prompted a strategic review and re-consideration of the business needs to be met by management development.

Characterising all the programmes was the senior level at which discussions to create and commission a programme were held. At this initial stage, the executive team was involved in sign-off, if not prompting the decision to commission a programme:

"The CEO brought in this group of HR people and said to them: you need to put HR processes into this organisation that help us manage it smarter and better. And one of those HR processes was management development." [Client director, Company G]

There was no evidence that stakeholders involved in commissioning programmes were basing their decision upon extant data about the effectiveness of management and leadership development, either from previous evaluations within their own company, or the wider field of evaluation. There was some evidence to suggest that evaluation of the

current programme would serve to inform the commissioning and design of future development programmes, e.g.:

"I think you just build up a knowledge bank." [Programme manager, Company G]

However, this was not a strong theme, and other responses suggested that retrospective data was unlikely to inform future decision-making:

"I don't think any of the decision-makers involved in this have shown any interest in the former evaluations right now. Not at all." [Senior sponsor, Company H]

5.3.2   Choosing the provider

Typically, a couple of HR or learning and development managers visited a range of business schools on an informal basis prior to releasing a formal tender document and inviting providers to tender. Provider reputation was one of the factors informing the initial choice of which providers to approach. There was no evidence that reputation was directly linked to evaluation results from previous programmes run by the provider. However, what qualified a provider as having a good reputation or being a *"top provider" [Programme manager, Company C]* may have been linked to its performance in terms of factors such as turnover, scale of operation, quality of client list, external ranking, international penetration and word-of-mouth. One might assume that the delivery of high quality products, including effective management and leadership development programmes, as assessed by evaluation would underpin strong performance across these factors. However, there was no evidence to assert this link.

Potential providers presented to a panel, usually involving HR and business managers, who made the final selection, although there were often one or two key drivers/influencers of the decision. In additional to the provider's reputation, the personal rapport they established with panel members prior to formal selection and the impression they created during the presentation/selection process were key factors at the point of decision. Prior personal experience of company staff having worked with the provider or attended one of their programmes in the past were also influential both at the invitation and selection stages:

"The attraction to us was that we would tailor more, and the other key was Kerstin [pseudonym], who had done her MBA here years before and had a very good experience and she was very senior in HR and development..."
[Client director, Company A]

The selection process was co-ordinated by the HR/learning and development function:

"I've got the L&D background, to be able to steer the business when they're coming up with these wild and wacky ideas of how they want to develop their people!"
[Programme manager, Company G]

The involvement of business managers in provider selection was often seen as part of

the stakeholder engagement process.

5.3.3   Working up the detailed design

By the time that providers were approached and one selected, client companies had identified a perceived need for a programme, a target audience and some desired outcomes. During the selection process, providers articulated their ideas for programme design to deliver on the clients' aspirations. The combination of ideas was typically then combined in a joint design process between client and provider, led by the programme manager/owner on the client side and client director on the provider side, and involving a wider group of stakeholders. In a number of cases the programme manager/owner led a diagnostic process, sometimes with the provider, interviewing potential participants about their personal leadership/management challenges, and senior managers about what the priorities of the programme should be.

The degree to which the client became involved in the details of the design was variable between companies:

"At the beginning, [the steering group was] very, very heavily involved, in fact they wrote the programme. [The provider] facilitated and helped. It was a real joint design."
[Programme owner, Company D]

"I don't need to worry about the details, the timing, all I need to worry about is what can we provide from our side." [Programme manager, Company C]

The process of developing a programme design appeared to be strongly influenced by prior experience – on both the provider and client side – of designing and running programmes. However, there was no indication that evaluation findings from previous programmes had played a role in informing ideas about programme design.

5.3.4   Selecting participants

Participant selection typically was managed at the business unit/stream level by the senior business stream director in liaison with the head of HR for that part of the business. This followed guidelines from the central HR/learning and development function on selecting participants and, in some cases, a quota of participant places for each business unit.

"I put together a brief...which I sent to the Chief Executives of our regions...I send them a document outlining how they can nominate people. I say a little bit about the programme, a little bit about [the provider], what the programme would cover, and then the nomination criteria...which is linked to...levels."
[Programme manager, Company G]

The selection process was not without challenges:

"Then it became a success, that a lot of people wanted to come on it, and actually the

last groups, you could wonder: why are these people here? - certainly from a company point of view." [Client director, Company A]

"There's far too much nominating of people who are available rather than people who could do with it. And no one-to-one communication to the person who's been nominated about the fact that they've been nominated!" [Programme owner, Company D]

Some programmes were centrally funded, whilst in other cases business units were charged for each delegate place, or there was a combination process: the business unit was charged if there was a no-show from the participant, otherwise the place was centrally paid for.

In no cases was there an explicit policy of self-nomination, although selection may have been influenced by individuals' requests to be nominated.

For one programme a participant profiling exercise had been carried out. Its purpose was to identify whether previous programme participants were representative of the company population as a whole, and to identify specific demographic groups to target for future programmes. The profiling was not, however, linked to any measure of participant learning, behavioural change or success, and thus fell short of using evaluation data to identify whether certain groups appeared to benefit more from the programme than others. In other programmes, where a quota was applied to achieve a mix of participants from across the business, this did not appear to be linked to evidence from evaluation that demonstrated the benefits of mixed groups.

5.3.5   Deciding to continue and/or modify the programme

The first iteration of a programme was typically run as a pilot, and assessed with a greater level of scrutiny than later iterations. The programme manager and client director attended throughout, and often there were additional observers and/or participants from the HR/learning and development function who were expected to provide feedback on how the programme had gone. After the programme, and after each module where a programme was modular, there was a review meeting between client director and programme manager, and sometimes additional stakeholders from the provider and/or client side. In all cases, evaluation data from feedback forms distributed at the end of a programme/module were incorporated into this review. Other data informing the review included the personal experience of having attended the programme, feedback collected informally from participants during or after the programme and conversations with senior management about the priorities of the business, to be reflected by the programme, going forward.

"Feedback from participants plus I was there for all the programmes to observe, so it was also my judgement based on observations. And informal feedbacks or dialogues with the participants. Plus I had the dialogues with the top management team before and after." [Former programme manager, Company H]

In a couple of instances, additional feedback was sought from participants through formal evaluation processes and used to inform design modifications to the programme. Methods included focus groups prompted by poor review scores and interviews conducted several months post programme. Where a steering group was in place, this group might be involved in the detailed review and programme design modifications, or provided with a report by the programme manager and asked for comments and input to future programmes.

Thus a combination of retrospective and prospective data was used to inform decisions to modify future iterations of a programme, with variation reflected in the degree to which programmes were modified and the extent to which retrospective or prospective data was used:

> *"Post those focus groups and module two the programme group sat down and redesigned the whole of module one and actually a vast chunk of module two."*
> *[Programme manager, Company F]*

> *"They weren't reflective at all. 'What are the main business issues this year?' 'What are the main strategic themes this year?'...Very much prospective rather than retrospective."*
> *[Senior sponsor, Company H]*

In the contracting stage before a programme started there was usually a discussion of programme roll-out and an expectation that more than one iteration would be run, implicitly or explicitly agreed even if not contractually set in stone. In a number of instances the scope was for a target population to be reached, for a programme always to be available to a specific level of management, for a critical mass of people to be reached, or for a business need to be met by training, e.g.:

> *"I want to...begin to drive a collective spirit that as a collection of leaders in the middle part of the organisation we can change some things for the better."*
> *[Programme manager, Company F]*

5.3.6   Programme postponement/termination

In all but two cases, the programmes in question were 'live'. In one case, there had been insufficient nominations the previous year to justify running a programme in 2010. This may have been linked to the fact that the business units were charged for each delegate place, and were seeking to cut costs given the tough economic conditions. In the second case, the financial climate and reluctance of potential participants to take time out given business pressures were given as the reasons that the programme had been postponed. However, this had then been followed by key stakeholders leaving the company and a major organisational re-structuring, which had left the programme ownerless and the HR function without strong advocacy to the board:

"Stability in corporate HR, or you could say instability. I left.., some of my colleagues in top management in HR decided to leave as well, our new boss at the time was fired...there were a lot of changes in global HR. You can say that the history [laughs] of what is it that we hold high on the agenda and why, and how is it that we work with it, how is it that we keep a continuum running, it's slightly endangered... So now it's up to the Executive Committee to remember what needs to be done and how and so on."
[Former programme manager, Company H]

When stakeholders of the other programmes were asked hypothetically what would cause their programmes to be stopped, the most common responses were that the target population had been reached, or budget cuts would threaten the viability of the programme. Other reasons cited were if the company's strategic direction changed and there were new business priorities or if there was a merger or takeover during which a new development strategy was imposed. Poor feedback from participants, the programme not delivering on its objectives and/or the company not seeing value from the programme were also given as reasons. Another reason given was the loss of the personal relationship between the provider and client due to staff changes.

5.4   The role of evaluation in decision-making

Of the decisions concerning the management and leadership programmes studied, the one to which evaluation contributed most substantially was that of modifying the design of future iterations of a programme in order to *improve* it (Easterby-Smith, 1994). The type of evaluation data informing this decision was primarily reaction-level data collected from participants at the end of, or immediately after, a programme.

"On the last day of...each programme I plead or yell or threat – take your pick! – with the delegates to email me or phone me up or come and see me with their considered views about what they felt about the programme and what they'd like to see differently in future programme for future generations."
[Senior sponsor, Company B]

 The decision was not linked to evaluated evidence of learning transfer (Baldwin and Ford, 1988) and impact in the workplace. Evaluation processes did not appear to be

sufficiently sophisticated to allow for those deploying them to learn about what worked and why (Denyer et al., 2008) – i.e. to connect specific programme activities with individual or organisational outcomes. Thus, there was limited, if any, potential for using evaluation as a means to improve future decisions about programme design and delivery.

Evaluation data was not the only, nor indeed the main, influencing factor in the decision to modify programme design. Other factors were the personal experience of the programme manager and/or other key stakeholders who attended the programme, the informal conversations the client director and programme manager/owner had with participants and with senior managers who delivered sessions on the programme, and evolving business needs.

In addition to informing iterative improvements, evaluation processes were also put in place to *prove* the value of a programme to the company that had commissioned it. Methods included administering surveys and conducting interviews with participants and line managers several months after a programme. Such data appeared to influence the decision to continue running a programme to some extent. However, key stakeholders' direct experience of a programme and evidence captured through informal and anecdotal feedback contributed more substantially to demonstrating value to key stakeholders and influencing their decision to continue investing in a programme. Informal processes included chance conversations of the programme manager and senior management with participants about their reactions to a programme and the changes that they had made in their work and behaviour as a result. Line managers also provided feedback informally about changes they had seen in the behaviour of their direct reports after having attended a programme. In one situation, the programme manager looked for evidence of participant promotions as a way of demonstrating, to herself and others, that the programme had been targeted at the right candidates – the company's high potentials.

In several instances, interviewees described their aspirations for more elaborate or comprehensive evaluation than was currently in place – including expressing embarrassment at current practice in order to prove the value of a programme:

> *"I'm probably ashamed to tell you it's not a very scientific process at the moment. Maybe as we're staffing up we'll get a bit more structure and process and fact base."*
> *[Senior sponsor, Company B]*

The speculative, future-oriented nature of these comments implies that evaluation to prove value was difficult to reconcile with here-and-now decisions about a particular programme. Evaluation to prove value may have been best suited to informing decisions to commission and design similar programmes in the future. However, there was only limited evidence to suggest that evaluation data was used in this way.

With one exception, where evaluation of the programme was driven by the Chief Executive, HR/learning and development stakeholders initiated the evaluation process.

They were more interested in the use of evaluation to both to *improve* a programme and *prove* its value–- than the provider or senior management:

> *"The MD was confident enough that we could feel that it's doing a good job. Considering how much money we've spent, it's almost ridiculous."*
> *[Programme manager, Company D]*

> *"A lot of these things like evaluation are done to give people a reason for existence, to say 'we are doing a good job'."*
> *[Client director, Company A]*

An element of exerting *control* (Easterby-Smith, 1994) also seemed to be important to HR/L&D stakeholders, whether to justify their decisions and strengthen their standing or influence over other key stakeholders.

Conducting evaluation for the purpose of *learning*, (Easterby-Smith, 1994) i.e. to reinforce or enhance participants' learning from a programme, was not well evidenced. In one case, data from a repeat 360-degree feedback process intended primarily for participant development was aggregated to give a pre-post comparison of participants' behavioural development. However, this was opportunism: using a learning process to provide data for evaluation to demonstrate the impact of a programme on behavioural change, rather than being evaluation conducted with the purpose of learning.

Although a number of research subjects suggested that poor participant feedback from evaluation processes could influence the decision to terminate a programme, there were no examples where this had been the case. The main reasons speculatively given for programme termination were saturation of the target population and budget cuts. Evaluation data had not been a major contributing factor in the decision to postpone the two programmes that had lapsed. Instead it was driven by insufficient participant nominations in one situation, possibly relating to the cost of the programme, and to budget cuts in the other.

In terms of the mechanisms by which formal evaluation data was shared and who the audiences were for such data, this varied across programmes. Data from programme review forms were most commonly collated by the provider's programme administrator and sent to the client director and programme manager, who might pass them on to a wider group of stakeholders. Data from other evaluation processes, which were initiated in all cases by the client rather than by the provider, were collated within the learning and development team and the programme manager or owner would decide how and to whom they should be reported. In some cases a summary was shared with the programme steering group and possibly sent to the executive board, with limited evidence of such reports impacting on decisions about a programme:

> *"I was never clear of what became of them [evaluation results]...after that point."*
> *[Client director, Company H]*

> *"I don't think any of the decision-makers involved in this have shown any interest in the*

former evaluations right now. Not at all."
[Senior sponsor, Company H]

"You never know what to expect!... It's just reporting to them [the board]. If they want to react they do, but it's to keep them in the know."
[Programme owner, Company D]

Evaluation data may have contributed to stakeholder buy-in and senior support for the decision to continue investing in a programme, though – as indicated above - there was limited evidence supporting this view. Stakeholder engagement is central to several models of evaluation (e.g. Preskill and Torres, 1999) but, notably, a stakeholder evaluation approach (Hirsh and Burgoyne, 2009), involving and engaging the interest of a wide range of stakeholders, was lacking in most of the programmes studied.

CHAPTER 6. DISCUSSION

Empirical data concerning eight management and leadership development programmes showed that the primary contribution of formal evaluation to the decisions made about a programme was in modifying the design after one iteration and before the next to *improve* the programme (Easterby-Smith, 1994). The type of data used was, by and large, reaction-level data (Kirkpatrick, 1998). The mechanism for gathering and using evaluation data, primarily from end-of-programme review forms, to inform design changes was simple, timely and cost-effective, i.e. the barriers to implementation were low. In most cases, the process was led by the provider and had an additional purpose for them, quality management: it served as the means by which comparative assessments were made of programmes and faculty and was widely embedded into the provider's organisational practice. Programme review forms were, however, only one of a number of influences on the decision to modify programme design. Often, additional data was solicited from participants and other stakeholders for this purpose. This was sometimes considered as part of the evaluation process by the programme manager and client director – who were typically the lead stakeholders in any decision to modify programme design – but usually fell outside of the way in which evaluation was conceived of. This type of data is retrospective and programme-specific: it is concerned with what worked, what didn't, and what could be improved on, though often ignoring mechanisms of change by not focusing on why (Collins and Denyer, 2008). The decision to modify programme design was also influenced by external factors, such as new strategic priorities and the business environment. These influences tended to come from senior stakeholders, e.g. the executive board or a programme steering group. In sum, evaluation to improve a programme was widespread, but was only one of a number of influences on the decision to modify programme design.

The decision by a client to keep running a programme was another to which evaluation contributed. However, it is worth reflecting on the way in which earlier decisions about the conception and commissioning of a programme were made, because these impacted on subsequent decisions to continue running, or to curtail, a programme. Significant investment of time, resources and professional pride went into programme conception and commissioning. Two main strategies dominated, here labelled HR/L&D-led and business-led, where HR/L&D-led strategies built on assumptions about the general benefits of management development for particular levels of management whilst business-led approaches focused on specific organisational challenges which a management development programme was intended to address. In both cases, for programmes pitched at high potentials and mid-senior management, there was close engagement of HR and L&D professionals with senior management in the early stages of developing the concept and choosing a provider to run a programme, perhaps because of the significant financial investment these sort of programmes entail and because a positive outcome of the development programme on organisational performance would be of strategic importance. The impetus for developing a new programme was sometimes associated with the recruitment of key stakeholders, e.g. an L&D professional brought in to develop a coherent suite of management and leadership development programmes, or a new board member keen to make an impression. These individuals were heavily invested in the programme as their professional reputation was

linked to its perceived success or failure. There were usually one or two key individuals behind the decision to commission a programme, but the decision-making structure was frequently broadened to encompass a wider group of stakeholders. This may have been partly an attempt to implement a structural model of decision-making (Beach and Connolly, 2005) whereby the inadequacy of any one individual's processing and analysis capabilities is overcome by structures that bring together an appropriate group of people to make decisions. However, it appeared to be largely about the HR/L&D lead orchestrating stakeholder engagement: increasing buy-in, sharing responsibility for the investment, raising awareness and easing the way for programme roll-out.

In the decision to commission a provider to run a programme, formal tendering processes tended to be supplemented or subverted by prior personal relationships and/or the opinions of one or two influential stakeholders. The espoused aspiration appeared to be for a decision to be reached through a rational process, which assumes there are no biases (Lyles and Thomas, 1988), yet decision-makers inevitably bring their own prior experiences and mental models – intrinsic biases (Holton and Naquin, 2005) - into the situation. During the commissioning process, providers were often approached on the basis of their reputation. Performance criteria meriting a 'good reputation' may have included turnover, scale of operation, quality of client list, external ranking, international penetration and word-of-mouth. One may assume that strong performance across these types of criteria correlates with delivery of programmes that generate value for companies, which in turn may have been proven by evaluation data. However, this is a tenuous link, and no data was found to support evaluation data being used in this way.

 In the contractual agreements reached between client and provider there was usually an explicit agreement, or implicit assumption, that multiple iterations of a programme would be run. The first iteration was often positioned as a pilot, with the expectation that whether it went well or badly, the next decision would be whether and by how much to change it before the next iteration, rather than whether or not to continue running it. Thus whilst a programme which hasn't started out as well as planned may appear vulnerable, it is often in the personal interests of providers, programme managers and senior stakeholders to make changes in an effort to ensure the programme's improvement, rather than to cancel it outright. Their decisions are in part influenced by their own stake in the outcome, part of the political backdrop in which evaluation is embroiled (Weiss, 1990).

Beyond the initial piloting stage, the decision as to whether to keep running a programme, albeit with modifications, or to postpone or terminate it was subject to a range of factors. It appeared that factors detached from the perceived impact of the programme, such as departmental budget cuts and changing strategic priorities, or having exhausted the target population, were the most influential in a decision to terminate a programme. Poor participant feedback gathered through evaluation and informally was cited as a reason to stop a programme, but this was not evidenced by any examples and remains an area for further exploration.

In terms of continuing to run a programme, beyond formal contractual arrangements, there appeared in some cases to be an element of status quo bias (Samuelson and Zeckhauser, 1988) i.e. a preponderance to a 'non-decision' of continuing as before, unless there was a notable reason to do otherwise. This was coupled with the belief apparent in several cases that if care was taken during the design and development phase, a programme should deliver value, despite a lack of evidence about effective teaching and learning practices being used to support design in the first place. In other cases, data indicating a programme's positive impact was important. Formal evaluation processes contributed in this respect as 'proof' (Easterby-Smith, 1994), and seemed to be held in particularly high regard by programme managers and owners, who were usually the instigators of more involved evaluations. There seemed, however, to be variation in the extent to which so-called 'impact evaluations' (Harper and Beacham, 1991; Russ-Eft and Preskill, 2001), i.e. those seeking to prove the effectiveness of a programme in changing behaviours and contributing to demonstrable individual and organisational outcomes, actually influenced the decisions made about a programme. There were problems of timing if evaluation data for one iteration only became available after the next iteration of the programme had been delivered. It was also questionable whether the mechanisms for change (Collins and Denyer, 2008) leveraged by a programme were sufficiently well understood for the results of an impact evaluation to point to how to change programme design to increase its impact. Impact evaluations were usually owned by the programme manager/owner from with the HR or L&D function, and perhaps served best the purposes of providing self-assurance, and contributing to their future actions when designing and implementing new programmes. The distinction between pragmatic evaluation in the here-and-now that supports decision-making about the programme being evaluated, and evaluation that contributes to a body of knowledge and is used for future reference, is important. Both types should inform decisions about programmes (Burgoyne, 1973), yet there was little evidence from this study that practitioners use programme evaluation from one situation to inform the next.

An interesting point of detail is the discrepancy between actual and intended evaluation practice: in several instances, unease or embarrassment was expressed about the inadequacy of current evaluation practice and claims were made about plans for more comprehensive evaluation in the future. The details of this tended, however, to remain vague. This was not true for all companies and a minority had extremely well thought-through and integrated evaluation approaches that sought to ensure that the organisation could measure the value being gained from investment in management development. In other situations there may have been an 'evaluation collusion': an espoused desired to do more, and more elaborate, evaluation, but because this it was not in any of the key stakeholders interests, it never happened. One might argue that the provider's motivation is for the client to continue to buy programmes; the programme manager and programme owner are keen that the programme is perceived to be delivering value as this reflects well on them and demonstrates that they are doing a good job, and the senior stakeholders who made the decision to commit resources to the programme in the first place may be satisfied once they have been shown to have invested in the company's people. This is a speculative and deliberately provocative interpretation, but is intended to provide support for the views that different stakeholders have

different interests in the results of evaluation (Michalski and Cousins, 2001) and that the evaluation approach will vary depending on whose interests are being served (Greene, 1997).

In summary, this empirical study found that of the decisions made during the lifecycle of a management and leadership development programme, the one to which evaluation appears to contribute most substantially is that of modifying programme design. There was no evidence that evaluation contributes to early decisions around programme conception and commissioning and limited evidence of the extent to which evaluation affects decisions to continue or terminate a programme. What is notable is the degree to which processes and data other than formal evaluation are influential in the decisions to which evaluation also contributes. Senior managers, the programme manager, programme owner and client director, are all influenced by their direct experience of a programme, by the informal conversations that occur during and after a programme with participants and their line managers and by observing and experiencing changes in participants' behaviours and organisational practices. These tend not to be considered as 'evaluation', because there is no intent to capture and collate data and the process is not systematised nor controlled, yet such mechanisms appear to be highly influential in how key stakeholders of a programme gain an impression of whether or not it is working, and what impact it is having.

There is an opportunity to build on this realisation and for future research to focus on the effective integration of a range of sources of evidence in making decisions about management and leadership development interventions. The aim would be both to promote the contribution of evaluation and to recognise its place alongside other sources of evidence. One frame through which to consider this is that of Evidence Based Management (EBMgt) defined as:

"making decisions through the conscientious, explicit, and judicious use of four sources of information: practitioner expertise and judgement, evidence from the local context, a critical evaluation of the best available research evidence, and the perspectives of those people who might be affected by the decision" (Briner et al., 2009).

Figure 2. Briner, Denyer and Rousseau's (2009) four elements of EBMgt

69
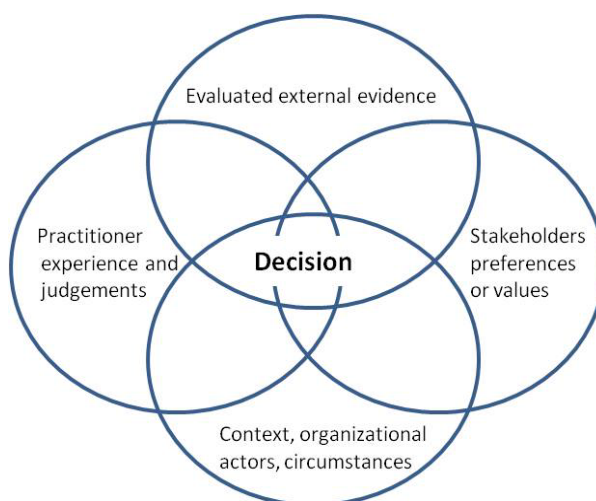
Briner, Denyer and Rousseau (2009) argue that optimal decision-making happens when evidence from all four elements is effectively brought together, whilst recognising that the emphasis placed on each element will vary from decision to decision.

Figure 3. Modified EBMgt model reflecting decision-making about management and leadership development programmes

e.g. Systematic literature reviews, meta-analyses, research on learning design and educational methods, evaluation frameworks.

Including: programme manager, programme sponsor/owner, senior managers, participants, their line managers & colleagues, provider

1. Evaluated external evidence

4. Practitioner experience and judgements

Decision

2. Stakeholders preferences or values

3. Context, organizational actors, circumstances

Informed by: Participant observation, prior experience of running programmes Informal conversations with participants

Further research

Informed by: programme evaluation data, business-drivers, HR-drivers

Those responsible for decisions relating to the eight evaluated programmes in this study were rarely seen to draw on evaluated external evidence about 'best practice' in either learning design or approaches to evaluation (Circle 1). Further, in most of the programmes studied, information about the programme and the organisational or participant context was not gathered in a systematic way, and formal evaluation processes - one component of 'context' – were often not influential (Circle 3). Therefore, decisions were based predominantly on preferences of stakeholders, particularly programme participants (Circle 2) and the largely implicit and subjective experience of the practitioners responsible for commissioning, designing and delivering the programmes (Circle 4).

This is congruent with the bounded rationality perspective advocated by Holton and Naquin (2005) that recognises decision-makers' cognitive limitations and does not assume they operate rationally with optimal information and resources at their disposal. However, in terms of EBMgt, it is a notable gap.

The conclusion drawn is that there is an opportunity to explore how to better integrate all four elements of EBMgt in order to improve the way in which decisions about management and leadership development programmes are reached. If this could be orchestrated, and evaluation conceived of as a valued part of an evidence-based

approach, perhaps evaluation will become a more respected activity and decision-makers will make decisions with greater awareness and better information, which must surely be a positive step.

CHAPTER 7. IMPLICATIONS AND LIMITATIONS

7.1   Implications and suggestions for further research

This work makes its contribution by bringing empirical data to the underexplored area of how programme evaluation is used in support of decision-making about management and leadership development. For practitioners, it provides confidence that evaluation is serving well the purpose of fine tuning programmes and safeguarding participant satisfaction. However, to fully realise its potential to support decision-making, it is suggested that less emphasis should be placed on documenting and reporting evaluation data and more on engaging key stakeholders directly with the programme and its participants, giving recognition to the influence of informal and incidental feedback processes.

For researchers, the study should serve as a starting point for further work to integrate decision-theory with programme evaluation. There is a vast body of literature on organisational decision-making, which was outside the scope of this study, and translating its salient features into ideas for practical research linked to evaluation would be valuable. This could be aligned with testing the lens of Evidence-Based Management (EBMgt) (Briner et al., 2009) as a model for reframing the way in which evaluation is conceived of. Future research could focus first on understanding the barriers to integration of the four elements of EBMgt and second on identifying and testing ways to improve integration. This might elicit whether the model is indeed appropriate to decision-making in this context, or whether it needs to be modified.

7.2   Limitations

Several limitations of the study are recognised, despite attempts to apply consistency and rigour to the literature review process and empirical work.

Firstly, it was difficult to obtain access to client organisations where a programme had recently been cancelled. This may have been partly the timing, coming in the shadow of the global financial crisis when client organisations were focusing their energy on survival, and perhaps also because client directors were reluctant to grant access to companies if they felt a sense of personal failure that a programme had been cancelled and did not wish for this to be explored. As a consequence, there is only limited data about the factors that cause programmes to be cancelled and the contribution of evaluation to this decision. This was partially addressed by asking 'what would...' questions, but it is recognised that hypothetical questions are not necessarily accurate representations of 'what is' or 'what was'.

Secondly, the researcher works for the organisation which provides the management and leadership development programmes studied. There may have been some reserve on the part of interviewees from the client side, fearing that they might reveal too much information about their organisation, and from the provider side in client directors not wanting to damage the relationship with the client. By way of mitigating against reservations about disclosure getting in the way, interviewees were assured anonymity

for themselves and the companies, and were given the opportunity to review the vignettes written about their programme prior to inclusion.

Thirdly, and relating to the point above, there was a double edge to studying programmes delivered by a single provider: on the upside, there was a limit to the degree of variation in the programmes examined, which supported the validity of the work. On the downside, the generalisability of the study is somewhat weakened and it would be interesting to explore whether similar themes would emerge if programmes delivered by other providers were considered. All the programmes included in the study were aimed at "high-flyers" or middle-senior management. They were considered of strategic importance for the investing companies, and stakeholders involved included senior managers, if not the board. It would be interesting to explore decision-making processes and the role of evaluation programmes aimed at lower levels of management.

CHAPTER 8. CONCLUSION

There is limited evidence in the evaluation literature about how evaluation models and data are really used by organisations to inform decisions, despite decision-making being one of the primary espoused functions of evaluation. Further to this, questions have been raised about the usefulness of HRD evaluation models in supporting decision-making and naturalistic decision research called for, whereby decision-making processes in use are studied. This study took inspiration from these ideas. It makes a contribution by exploring the role of programme evaluation in organisational decision-making about management and leadership development.

The study concludes that formal programme evaluation does contribute to decisions made about management and leadership development, but only to a limited extent, and only to certain decisions. The main contribution of evaluation appears to be in the decision to modify a programme with the intent of improvement. To some extent evaluation was also found to contribute to the decision to continue to run a programme, and perhaps to programme termination and to the confidence with which future programmes are designed and deployed. There was no evidence to suggest that decisions about commissioning and designing management and leadership development programmes were informed by evaluation of previous company programmes, or by evaluation data sourced from the wider field of management and leadership development.

Anecdotal data about a programme gathered informally by key stakeholders who come into direct contact with a programme and its participants was found to be integral to the same decisions to which evaluation contributed. For management and leadership development programmes aimed at "high flyers" and senior managers, these stakeholders are typically at the top of the organisation and hold significant decision-making power and authority. External factors, unrelated to the programme itself, were also found to be influential.

There is an opportunity to build on the realisation that formal evaluation is just one of multiple influences on decisions about management and leadership development interventions, and for future research to focus on the conscious and judicious integration of a range of sources of evidence. Evidence-Based Management (EBMgt) provides a useful frame by which to do this. Proponents of EBMgt propose that optimal decision-making happens when evidence from all four elements of the model is effectively brought together. It appears that, currently, decisions about management and leadership development programmes are being made using a combination of elements of the EBMgt model, but not all elements are being used to best effect. Those responsible for decision-making rarely draw on *evaluated external evidence* – one of the four elements - about 'best practice' in either learning design or approaches to evaluation. Further, information about *context, organisational actors and circumstances* is not gathered in a systematic way, and formal evaluation processes - one component of 'context' – often lack influence. Therefore, decisions are based predominantly on preferences of stakeholders, and the largely implicit and subjective experience of the practitioners responsible for commissioning, designing and delivering the programmes.

If ways can be found to successfully integrate all four elements of Evidence-Based Management, it is likely that the utility of evaluation will increase and, ultimately, decision-makers will make their decisions with greater awareness and better information.

REFERENCES

Alampay, R. H. and Morgan, F. T. (2000), "Evaluating external executive education at Dow Chemical: its impact and the Pygmalion effect", *Human Resource Development International,* vol. 3, no. 4, pp. 489-498.

Alimo-Metcalfe, B. and Lawler, J. (2001), "Leadership development in UK companies at the beginning of the twenty-first century: Lessons for the NHS?", *Journal of Management in Medicine,* vol. 15, no. 5, pp. 387-404.

Alliger, G. M. and Janak, E. A. (1989), "Kirkpatrick's Levels Of Training Criteria: Thirty Years Later", *Personnel Psychology,* vol. 42, no. 2, pp. 331-343.

Alliger, G., Tannenbaum, S.,I., Bennett, J. W., Traver, H. and Shotland, A. (1997), "A meta-analysis of the relations among training criteria", *Personnel Psychology,* vol. 50, no. 2, pp. 341-358.

Arthur Jr, W., Bennett Jr., W., Edens, P.,S. and Bell, S.,T. (2003), "Effectiveness of training in organizations: A meta-analysis of design and evaluation features", *Journal of Applied Psychology,* vol. 88, no. 2, pp. 234-245.

Atkinson, R. (1972), "Ingredients for a theory of instruction", *American Psychologist,* vol. 27, pp. 927-931.

Baker, H. K. and Gorman, R. H. (1978), "Evaluating the effectiveness of management development programs", *Public Personnel Management,* vol. 7, no. 4, pp. 248-257.

Baldwin, T. T. and Ford, J. K. (1988), "Transfer of Training: A Review and Directions for Future Research", *Personnel Psychology,* vol. 41, no. 1, pp. 63-106.

Bass, B. M. and Avolio, B. J. (1990), *Manual: The Multifactor Leadership Questionnaire,* Consulting Pscyhologists Press, Palo Alto, C.A.

Beach, L. R. and Connolly, T. (2005), *The Psychology of Decision Making,* Second edition ed, Sage, Thousand Oaks, California.

Black, T. and Westwood, M.,J. (2004), "Evaluating the development of a multidisciplinary leadership team in a cancer-center", *Leadership & Organization Development Journal,* vol. 25, no. 7/8, pp. 577-591.

Blaikie, N. (2007), *Approaches to Social Enquiry,* 2nd edition ed, Polity Press, Cambridge.

Boaden, R.,J. (2006), "Leadership development: does it make a difference?", *Leadership & Organization Development Journal,* vol. 27, no. 1/2, pp. 5-27.

Bowman, G. and Thompson, D. (1995), *A Classification System for Nurses; Work Methods: The Bowman Classification,* National Institute for Nursing, Oxford.

Briner, R., Denyer, D. and Rousseau, D. (2009), "Evidence-Based Management: Concept Cleanup Time?", *Academy of Management Perspectives,* vol. 23, no. 4, pp. 19--32.

Brundrett, M. (2006), "Evaluating the Individual and Combined Impact of National Leadership Programmes in England: Perceptions and Practices", *School Leadership & Management,* vol. 26, no. 5, pp. 473-488.

Burgoyne, J. (1973), *The evaluation of managerial development programmes with special reference to the Manchester Business School* (unpublished PhD thesis), Manchester Business School, .

Burgoyne, J., Hirsh, W. and Williams, S. (2002), *The Value of Business and Management Education,* , The Association of Business Schools, London.

Burgoyne, J. (1973), "An Action Research Experiment in the Evaluation of a Management Development Course", *Journal of Management Studies,* vol. 10, no. 1, pp. 8-14.

Burgoyne, J. and Cooper, C. (1975), "Evaluation methodology", *Journal of Occupational Psychology,* vol. 48, no. 1, pp. 53-62.

Burgoyne, J., Hirsh, W. and Williams, S. (2003), "The Value of Business and Management Education", *Lancaster University Management School Working Paper, .*

Burgoyne, J. and Singh, R. (1977), "Evaluation of Training and Education", *Journal of European Industrial Training,* vol. 1, no. 1, pp. 17-21.

Burke, M. J. and Day, R. R. (1986), "A Cumulative Study of the Effectiveness of Managerial Training", *Journal of Applied Psychology,* vol. 71, no. 2, pp. 232-245.

Bush, T. and Glover, D. (2005), "Leadership development for early headship: the New Visions experience", *School Leadership & Management,* vol. 25, no. 3, pp. 217-239.

Campbell, D. T. and Stanley, J. C. (1963), *Experimental and quasi-experimental designs for research,* Rand McNally, Chicago.

Carlson, K. D. and Schmidt, F. L. (1999), "Impact of experimental design on effect size: Findings from the research literature on training", *Journal of Applied Psychology,* vol. 84, no. 6, pp. 851-862.

Carr, S., Lhussier, M., Reynolds, J., Hunter, D. and Hannaway, C. (2009), "Leadership for health improvement - implementation and evaluation", *Journal of Health Organization and Management,* vol. 23, no. 2, pp. 200-215.

Catanello, R. and Kirkpatrick, D. L. (1968), "Evaluating training programs - the state of the art", *Training & Development,* vol. 22, no. 5, pp. 2-9.

Collins, D. (2002), "Performance-Level Evaluation Methods Used in Management Development Studies From 1986 to 2000", *Human Resource Development Review,* vol. 1, no. 1, pp. 91-110.

Collins, D. (2001), "Organizational performance: the future focus of leadership development programs", *The Journal of Leadership Studies,* vol. 7, no. 4, pp. 43-54.

Collins, D. B. and Holton, E. F. (2004), "The Effectiveness of Managerial Leadership Development Programs: A Meta-Analysis of Studies from 1982 to 2001", *Human Resource Development Quarterly,* vol. 15, no. 2, pp. 217-248.

Collins, J. and Denyer, D. (2008), "Leadership, learning and development: A framework for evaluation", in Turnbull James, K. and Collins, J. (eds.) *Leadership Learning: Knowledge into Action,* Palgrave, London.

Cook, T. and Campbell, D. (1979), *Quasi-experimentation: Design and analysis for field settings,* Rand McNally, Chicago.

Cunningham, G. and Kitson, A. (2000), "An evaluation of the RCN clinical leadership developmpent programme: part 2", *Nursing Standard,* vol. 15, no. 13, pp. 34-40.

De Casterlé, B. D., Willemse, A., Verschueren, M. and Milisen, K. (2008), "Impact of clinical leadership development on the clinical leader, nursing team and care-giving process: A case study", *Journal of Nursing Management,* vol. 16, pp. 753-763.

Denyer, D., Tranfield, D. and Ernst van Aken, J. (2008), "Developing Design Propositions through Research Synthesis", *Organization Studies,* vol. 29, pp. 393-413.

Denzin, N. and Lincoln, Y. (eds.) (2000), *Handbook of Qualitative Research,* 2nd edition ed, Sage publications, Thousand Oaks, California.

Dexter, B. and Prince, C. (2007a), "Evaluating the impact of leadership development: a case study", *Journal of European Industrial Training,* vol. 31, no. 8, pp. 609-625.

Dexter, B. and Prince, C. (2007b), "Turning managers into leaders: assessing the organizational impact of leadership development", *Strategic Change,* vol. 16, no. 5, pp. 217-226.

D'Netto, B., Bakas, F. and Bordia, P. (2008), "Predictors of management development effectiveness: an Australian perspective", *International Journal of Training & Development,* vol. 12, no. 1, pp. 2-23.

Duda, R.,B. (2008), "The impact of a leadership development course for junior academic medical faculty", *Leadership in Health Services,* vol. 21, no. 3, pp. 200-207.

Duke Corporate Education (2010), *Learning and Development in 2011: A Focus on the Future; Duke CE's Client Study,* , Duke Corporate Education.

Dvir, T., Eden, D., Avolio, B. J. and Shamir, B. (2002), "Impact of transformational leadership on follower development and performance: a field experiment", *Academy of Management Journal,* vol. 45, no. 4, pp. 735-744.

Easterby-Smith, M. (1994), *Evaluating Management Development Training and Education,* 2nd edition ed, Gower Press, Farnborough.

Easterby-Smith, M. (1986), *Evaluation of Management Education, Training and Development,* Gower, Aldershot.

Easterby-Smith, M., Thorpe, R. and Lowe, A. (2002), *Management Research: An Introduction,* 2nd edition ed, Sage publications, London.

Easterby-Smith, M. (1981), "The Evaluation of Management Education and Development: An Overview", *Personnel Review,* vol. 10, no. 2, pp. 28-36.

Edmonstone, J. and Western, J. (2002), "Leadership development in health care: What do we know?", *Journal of Management in Medicine,* vol. 16, no. 1, pp. 34-47.

Frese, M., Beimel, S. and Schoenborn, S. (2003), "Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision", *Personnel Psychology,* vol. 56, no. 3, pp. 671-698.

Goldstein, I. (1986), *Training in organizations: Needs assessment, development and evaluation,* Brooks/Cole, Monterey, CA.

Greene, J. (1997), "Evaluation as advocacy", *Evaluation Review,* , pp. 25-35.

Greenhalgh, T. and Peacock, R. (2005), "Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources", *British Medical Journal,* vol. 331, pp. 1064-1065.

Hackman, J. and Oldham, G. (1980), "Motivation through the design of work", in *Work redesign,* Addison Wesley Publishing Company, Massachusetts, pp. 71-99.

Hamblin, A. (1974), *Evaluation and control of training,* McGraw-Hill, New York.

Hamilton, T.,A. and Cooper, C. (2001), "The impact of outdoor management development (OMD) programmes", *Leadership & Organization Development Journal,* vol. 22, no. 7/8, pp. 330-340.

Hannum, K. (2004), "Best practices: Choosing the right methods for evaluation", *Leadership in Action,* vol. 23, no. 6, pp. 15-20.

Hannum, K. and Martineau, J. (2008), *Evaluating the impact of leadership development,* Pfeiffer, San Francisco.

Hannum, K., Martineau, J. and Reinelt, C. (2007), *The Handbook of Leadership Development Evaluation,* Jossey-Bass, San Francisco.

Harper, J. and Beacham, T. (1991), *Training Performance Measurement: A guide to Some Practical Approaches,* NHS Training Authority, Bristol.

Hayes, J. (2007), "Evaluating a Leadership Development Program", *Organization Development Journal,* vol. 25, no. 4, pp. 89-95.

Hayward, I. (2009), *Evaluation of a Leadership Programme, PhD thesis (unpublished)* University of Bradford, .

Healey, M. J. and Rawlinson, M. B. (2004), "Interviewing techniques in business and management research", in Wass, V. J. and Wells, P. E. (eds.) *Principles and PRactice in Business and Management Research,* Dartmouth, Aldershot, pp. 123-146.

Hesseling, P. (1966), *Strategy of Evaluation Research in the Field of Supervisory and Management Training,* Van Gorcum, Amsterdam.

Hirsh, W., Burgoyne, J. and Williams, S. (2002), "The Value of Business and Management Education", *Report for the Association of Business Schools,* .

Hirsh, W. and Burgoyne, J. (2009), "Clarifying your learning and development strategy", *Training Journal,* , no. June, pp. 31-34.

Holton, E. (1996), "The flawed four-level evaluation model", *Human Resource Development Quarterly,* vol. 7, no. 1, pp. 5-21.

Holton, E. and Naquin, S. (2005), "A critical analysis of HRD evaluation models from a decision-making perspective", *Human Resource Development Quarterly,* vol. 16, no. 2, pp. 257-280.

Kaufman, R. and Keller, J. (1994), "Levels of Evaluation: Beyond Kirkpatrick", *Human Resource Development Quarterly,* vol. 5, no. 4, pp. 371-380.

Kearns, P. (2005), *Evaluating the ROI from Learning,* Chartered Institute of Personnel and Development, London.

Kim, H. and Cervero, R. M. (2007), "How power relations structure the evaluation process for HRD Programmes", *Human Resource Development International,* vol. 10, no. 1, pp. 5-20.

Kirkpatrick, D. L. (1998), *Evaluating training programs: The four levels,* Berret-Koehler, San Fransisco.

Kirkpatrick, D. L. (1978), "Evaluating in-house training programs", *Training & Development,* vol. 32, no. 9, pp. 6-9.

Kirkpatrick, D. L. (1960a), "Techniques for evaluating training programs: Part 3 - Behavior", *Journal of ASTD,* vol. 14, no. 1, pp. 13-18.

Kirkpatrick, D. L. (1960b), "Techniques for evaluating training programs: Part 4 - Results", *Journal of ASTD,* vol. 14, no. 2, pp. 28-32.

Kirkpatrick, D. L. (1959a), "Techniques for evaluating training programs", *Journal of ASTD,* vol. 13, no. 11, pp. 3-9.

Kirkpatrick, D. L. (1959b), "Techniques for evaluating training programs: Part 2 - Learning", *Journal of ASTD,* vol. 13, no. 12, pp. 21-26.

Kouzes, J. M. and Posner, B. Z. (1997), *Leadership practices inventory,* Jossey-Bass Pfeiffer, San Francisco.

Kraiger, K., Ford, J. K. and Salas, E. (1993), "Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation", *Journal of Applied Psychology,* vol. 78, no. 2, pp. 311-328.

Lyles, M. A. and Thomas, H. (1988), "Strategic Problem Formulation: Biases and Assumptions Embedded in Alternative Decision-Making Models", *The Journal of Management Studies,* vol. 25, no. 2, pp. 131.

Lynch, R., Leo, S. and Downing, K. (2006), "Context dependent learning: its value and impact for workplace education", *Education & Training,* vol. 48, no. 1, pp. 15-24.

Marshall, C. and Rossman, G. (1995), *Designing Qualitative Research,* 2nd edition ed, Sage publications, Thousand Oaks, California.

Martineau, J. (2004), "Laying the groundwork: First steps in evaluating leadership development", *Leadership in Action,* vol. 23, no. 6, pp. 3-8.

McLean, S. and Moss, G. (2003), "They're happy, but did they make a difference? Applying Kirkpatrick's framework to the evaluation of a national leadership program", *The Canadian Journal of Program Evaluation,* vol. 18, no. 1, pp. 1-23.

Michalski, G. and Cousins, B. (2001), "Multiple perspectives on training evaluation: probing stakeholder perceptions in a global network development firm", *American Journal of Evaluation,* vol. 22, no. 1, pp. 37-53.

Miller, D.,L., Umble, K.,E., Frederick, S.,L. and Dinkin, D.,R. (2007), "Linking learning methods to outcomes in public health leadership development", *Leadership in Health Services,* vol. 20, no. 2, pp. 97-123.

Newstrom, J. (1984), "A role-taker/ time-differentiated integration of transfer strategies", Vol. August, 1984, Toronto, Ontario, Toronto, Ontario, .

Orr, M. (2007), "Learning Advanced Leadership: Findings from a Leadership Development Programme for New Superintendents", *Educational Management Administration & Leadership,* vol. 35, no. 3, pp. 327-347.

Parry, K. W. and Sinha, P. N. (2005), "Researching the Trainability of Transformational Organizational Leadership", *Human Resource Development International,* vol. 8, no. 2, pp. 165-183.

Pawson, R. and Tilley, N. (1994), "'What works in evaluation research?'", *British Journal of Criminology,* vol. 34, pp. 291-306.

Phillips, J. (1995), "Return on Investment - Beyond the four levels", Holton, E. (ed.), in: *Academy of HRD,* .

Preskill, H. and Torres, R. (1999), *Evaluative Inquiry for Learning in Organizations,* Sage, Thousand Oaks, California.

Reade, Q. and Thomas, D., (2004), *Critics question value of leadership training.*

Robson, C. (2002), *Real World Research: A Resource Book for Social-Science Practitioners-researchers,* Blackwell, Oxford.

Rosenthal, R. (1979), "The "File Drawer Problem" and Tolerance for Null Results", *Psychological Bulletin,* vol. 86, no. 3, pp. 638-641.

Rosti, R.,T. and Shipper, F. (1998), "A study of the impact of training in a management development program based on 360 feedback", *Journal of Managerial Psychology,* vol. 13, no. 1/2, pp. 77-89.

Russ-Eft, D. and Preskill, H. (2001), *Evaluation in organizations: A systematic approach to enhancing learning, performance and change,* Sage, Thousand Oaks, C.A.

Russon, C. and Reinelt, C. (2004), "The Results of an Evaluation Scan of 55 Leadership Development Programs", *Journal of Leadership & Organizational Studies,* vol. 10, no. 3, pp. 104-107.

Sackett, P. R. and Mullen, E. J. (1993), "Beyond formal experimental design: Towards an expanded view", *Personnel Psychology,* vol. 46, no. 3, pp. 613-628.

Sadler-Smith, E. and Shefy, E. (2007), "Developing Intuitive Awareness in Management Education", *Academy of Management Learning & Education,* vol. 6, no. 2, pp. 186-205.

Samuelson, W. and Zeckhauser, R. (1988), "Status quo bias in decision making", *Journal of Risk and Uncertainty,* vol. 1, no. 1, pp. 7-59.

Saunders, M., Lewis, P. and Thornhill, A. (2007), *Research Methods for Business Students: Fourth Edition,* Fourth edition ed, FT Prenctice Hall, Harlow.

Sirianni, P. M. and Frey, B. A. (2001), "Changing a Culture: Evaluation of a Leadership Development Program at Mellon Financial Services", *International Journal of Training & Development,* vol. 5, no. 4, pp. 290-301.

Skipton, L.,H. and Goff, M. (2003), "Leadership Development as an Intervention for Organizational Transformation A Case Study", *Consulting Psychology Journal: Practice & Research,* vol. 55, no. 1, pp. 58-67.

Stufflebeam, D. (1989), "'The CIPP model for program evaluation'", in Striven, M. and Stufflebeam, D. (eds.) *Evaluation models,* Kluwer-Nijhoff, Boston.

Swanson, R. and Holton III, E. (1999), *Results: how to assess performance, learning, and perceptions in organizations,* Berrett-Koehler.

Tagg, C., (2010), *NVivo 8 workshop*, Cranfield University.

Talbot, C. (1992), "Evaluation and Validation: A Mixed Approach", *Journal of European Industrial Training,* vol. 16, no. 5, pp. 26-32.

Tannenbaum, S. I. and Woods, S. B. (1992), "Determining a Strategy for Evaluating Training: Operating Within Organizational Constraints", *HR. Human Resource Planning,* vol. 15, no. 2, pp. 63-81.

Tanton, M. and Fox, S. (1987), "The Evaluation of Management Education and Development: Participant Satisfaction and Ethnographic Methodology", *Personnel Review,* vol. 16, no. 4, pp. 33-40.

Thurley, K., Graves, D. and Hult, M. *An evaluation strategy for management development,* , Unpublished report.

Tranfield, D., Denyer, D. and Smart, P. (2003), "Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review", *British Journal of Management,* vol. 14, no. 3, pp. 207-222.

Tushman, M.,L., O'Reilly, C.,A., Fenollosa, A., Kleinbaum, A.,M. and McGrath, D. (2007), "Relevance and Rigor: Executive Education as a Lever in Shaping Practice and Research", *Academy of Management Learning & Education,* vol. 6, no. 3, pp. 345-362.

Tyson, S. and Ward, P. (2004), "The Use of 360 Degree Feedback Technique in the Evaluation of Management Development", *Management Learning,* vol. 35, no. 2, pp. 205-223.

Wagner, J. A.,III (1995), "Studies of individualism-collectivism: Effects on cooperation in groups", *Academy of Management Journal,* vol. 38, no. 1, pp. 152-172.

Warr, P., Bird, M. and Rackham (1970), *Evaluation of Management Training,* Gower, Aldershot.

Weiss, C. (1990), "Evaluation for decisions", in Alkin, M. (ed.) *Debates on Evaluation,* Sage, Newbury Park, C.A., pp. 171-184.

Wilson, K. and Corrall, S. (2008), "Developing public library managers as leaders", *Library Management,* vol. 29, no. 6/7, pp. 473-488.

Woltring, C., Constantine, W. and Schwarte, L. (2003), "Does leadership training make a difference? The CDC/UC public health leadership institute: 1991-1999", *Journal of Public Health Management and Practice,* vol. 9, no. 2, pp. 103-122.

Yin, R. K. (2009), *Case study research: Fourth edition,* Fourth edition ed, Sage publications, Thousand Oaks, CA.

Yorks, L., Beechler, S. and Ciporen, R. (2007), "Enhancing the Impact of an Open-Enrollment Executive Program Through Assessment", *Academy of Management Learning & Education,* vol. 6, no. 3, pp. 310-320.

Zelditch, M. (1962), "Some methdolological problems of field studies", *The American Journal of Sociology,* vol. 67, pp. 566-576.

Appendix A: Consultation panel

Table 8. Consultation panel

Person	Job title	Organisation	Role on consultation panel
Ms Anita Beal	Students Off-Site Service (S.O.S.) Liaison Officer	Cranfield School of Management	Support in sourcing articles
Dr. Eddie Blass	Associate Dean	University of Hertfordshire School of Education	Advisor on learning and education, sounding board for ideas
Dr. Cliff Bowman	Professor of Strategic Management	Cranfield School of Management	Informal mentor
Dr. David Denyer	Senior Lecturer in Organization Studies, Director of MSc by Research in Leading Learning and Change	Centre for Executive Learning and Leadership, Cranfield School of Management	Supervisor and Specialist in systematic review
Dr. Narendra Laljani	Director of Qualification Programmes	Ashridge	Advisor in relation to executive education
Ms Lorraine Oliver	Library Manager	Ashridge	Literature search advisor
Dr. Colin Pilbeam	Senior Research Fellow and Director PhD Programme	Cranfield School of Management	Member of review panel
Dr. Ellen Pruyne	Research Fellow	Centre for Research in Executive Development, Ashridge	Advisor on learning and education
Professor Kim Turnbull James	Professor of Executive Learning, Director of the Centre for Executive Learning and Leadership, Director of Faculty Development	Centre for Executive Learning and Leadership, Cranfield School of Management	Member of review panel
Ms Heather Woodfield	Information specialist	Cranfield School of Management	Literature search advisor

Appendix B: Databases accessed for systematic review

Table 9. Databases accessed

Database	Brief description
ABI Trade & Industry & Global/ Proquest	ProQuest provides access to multiple fields of research, from arts, literature, and social science to general reference, business, science, technology, and medicine. ProQuest includes the world's largest digital newspaper archive, periodical databases comprising the output of more than 9,000 titles and spanning more than 500 years, the pre-eminent dissertation collection, and various other scholarly collections.
Business Source Complete (EBSCO)	EBSCO features the full text of over 3,600 publications including scholarly journals, trade publications and popular business magazines.
ERIC	The ERIC (Education Resources Information Center) database is sponsored by the U.S. Department of Education to provide extensive access to educational-related literature. Of the major areas of coverage, the ones relevant to this study include: adult, career, and vocational education. Dates of coverage: 1966-current.*
PsycINFO	PsycINFO provides abstracts and citations to the scholarly literature in the psychological, social, behavioural, and health sciences. In October 2009 there were 2.8 million records available.*

Sources: Database publishers' promotional material/ websites, accessed 31 May 2009 and 5 October 2009.

* ERIC and PsycINFO are accessed through CSA Social Sciences which incorporates over 20 individual databases covering topics in education, policy development, psychology and sociology.

Appendix C: Search filters

<u>Peer reviewed/scholarly articles</u>

To apply a consistent initial quality screen, only articles from peer-reviewed/scholarly journals were included in the search. It is assumed that peer review process is robust and provides an independent quality check of articles that cannot be assured with non peer-reviewed publications. However, this was only an initial screening, and quality was reviewed again once full papers had been selected for inclusion.

<u>Published 2000 and later</u>

As discussed, the review followed on from earlier meta-analyses that covered a similar topic area. As such, only articles published in 2000 or later were included in the search.

<u>Abstracts</u>

The search was limited to abstracts. On the various databases, this filter is described as "Citation and abstract" (Proquest), "Abstract or author-supplied abstract" (EBSCO) and "Abstract" (CSA Social Sciences).

<u>Not publication title</u>

The "and not" filter was used to remove hits where the keywords come up in the publication title, but not in the article title or abstract.

Appendix D: Search strings

A variety of terms are used to describe management and leadership development. These, along with the term 'evaluation' were incorporated into a single search string for the purposes of the review, shown in table 10 below along with the returns (prior to selection) generated by running this search through EBSCO, Proquest and CSA Illumina.

Table 10. Search string 1

Search string 1, using term 'evaluation'	ProQuest search returns	EBSCO search returns	ERIC/ Psyc INFO returns
(management development OR executive development OR executive education OR management education OR leadership development) AND (evaluation) AND NOT (management development OR executive development OR executive education OR management education OR leadership development) in publication title	172	63	112

Following advice that the search string was too narrow, a second search was run with the terms 'impact' and 'effectiveness' included, but excluding 'evaluation' as that had already been covered. This search elicited the returns shown in table 11 below, again prior to selection. Some of these are duplicates of the articles identified through search string 1.

Table 11. Search string 2

Search string 2, using terms 'impact OR effectiveness'	ProQuest search returns	EBSCO search returns	ERIC/ Psyc INFO returns
(management development OR executive development OR executive education OR management education OR leadership development) AND (impact OR effectiveness) AND NOT (management development OR executive development OR executive education OR management education OR leadership development) in publication title	405	246	222

Appendix E: Selection criteria

Articles identified through the search strings shown in tables 10 and 11 above were first scanned by title to eliminate erroneous articles which had somehow made their way into the list of hits identified. After this initial selection process, abstracts were reviewed for relevance. A set of criteria was applied to determine whether or not the full paper would be considered for review. These inclusion criteria are listed in table 12 below. In some cases, the full article had to be considered, not just the abstract, to elicit the necessary information on whether the article met the inclusion criteria. Table 13 makes explicit the criteria used to exclude articles.

Table 12. Inclusion criteria for abstracts

Criteria	Decision	Rationale
Subject matter	Study must relate to a management or leadership development programme	The search strings are intended to only bring up material that covers the relevant areas, but in case of erroneous articles being identified in the search, a check will be made to ensure that the subject matter is relevant.
Type of article	Only empirical studies will be considered	The scoping study deals with meta-analyses and theoretical articles. This tightly focused review is concerned only with empirical studies.
Research methodology	Evaluation methodology must be clearly described	The review questions concern not only the effectiveness of management and leadership development, but also how effectiveness is assessed. As such, it is important that the evaluation methodology is clearly described.
Research methods	All research methods will be embraced, unlike the previous two meta-analyses (Burke and Day, 1986; Collins and Holton, 2004) this study which put restrictions on the methods considered for analysis	Restricting the methods by which an evaluation is conducted reduces the range of studies which might be considered. Given the relatively short time frame (2000 and later) of studies selected, it is felt to be important that all evaluation methods are considered. No comparative statistical analyses are proposed, and qualitative research methods can therefore be considered alongside quantitative and mixed method approaches.
Location	Studies located in/derived from all regions of the world will be considered	Whilst most of the research is anticipated to derive from the US and Europe, location is not considered to be an important factor in the selection of articles.
Language	English only	The majority of readily-accessible databases are in English, English is a widely used language for research publication, my language skills mean that English is the only realistic option for reading and fully understanding academic papers and translation would be too resource and time-intensive.

Table 13. Exclusion criteria for abstracts

Criteria	Decision	Rationale
Type of article	Theoretical papers, meta-analyses and summaries of other articles excluded	The review looks specifically at empirical studies, i.e. where an evaluation of a programme has been conducted and reported. The scoping study provides rich background information about theoretical models of evaluation and effectiveness of management and leadership development. Whilst searching only in scholarly/peer reviewed journals, these sometimes include short introductions to the issue or reviews of other articles. These are not considered to be of sufficient relevance for inclusion.
Focus of article	Exclude articles that are concerned with the evolution and/or planned development of organisations over time rather than with a specific intervention designed to address a specific organisational need.	The focus of the search is on the evaluation of management and leadership development programmes, not on the more general development of organisations, or on the effectiveness of management and leadership development with no evidence given of evaluation methodology.

Appendix F: Quality appraisal

Four aspects of each article were graded for quality on a 1-3 scale: theory and context, methodology, results and analysis and contribution to knowledge. Only those articles which scored 2 or 3 on at least two of the four criteria were included in the systematic review, i.e. articles which scored 1 on more than two criteria were rejected as they were not considered of adequate quality to merit inclusion in the review.

Table 14. Quality criteria

Criteria	Absent/Low - 1	Medium – 2	High – 3
Theoretical background & context	Little or no description of theoretical background and context in which paper is set	Reasonable description of theoretical background, link with the context in which paper is set	Well-argued theoretical background, key concepts clearly defined, explicit and obvious relationship between context for paper and theoretical underpinning.
Methodology	Methodology poorly described, insufficient detail to enable study to be replicated, inappropriate methodology for addressing research questions	Methodology outlined with some degree of detail, methods appear appropriate for research questions.	Clear explanation of why research methods were selected and are appropriate to address research questions. Detailed and clear description of how study was conducted.
Results and analysis	Results not presented and/or findings and analysis do not appear congruent	Results presented and explained at least in part by analysis	Excellent presentation of results and a full and well-argued analysis
Contribution to knowledge	Little or no theoretical or empirical contribution	A modest contribution made to theoretical or empirical knowledge made	A significant contribution made to either theoretical or empirical knowledge

The header of the table spans "Level achieved" over the three level columns.

Appendix G: Data extraction

The fields for data extraction were developed using a combination of standard citation information, quality appraisal criteria and codes used in earlier studies by Burke and Day (1986), Collins (2002) and Collins and Holton (2004). These earlier studies sought to categorise articles to facilitate the process of meta-analysis. Whilst this review stops short of meta-analysis, it was considered valuable to build on the coding categories of these earlier studies as a way of structuring data extraction and subsequent synthesis and to facilitate potential future comparisons of data sets.

A slightly modified version of Collins and Holton's (2004) outcome categories was used as the framework for classifying the level at which evaluations sought to capture data. These categories were themselves derived from Burke and Day's (1986) criterion measures and Swanson and Holton's (1999) *Results Assessment System*. They classified outcomes along two dimensions: criterion level and subjectivity-objectivity, yielding six outcome categories:

Table 15. Collins and Holton's (2004) Outcome categories, adapted

#	Outcome category	Description of category
R x n	**Reactions**	Satisfaction with how programme was run, including quality of faculty, venue, sessions, materials and design, generally assessed immediately after a session or programme.
1	**Knowledge – subjective** (K-S)	Principles, facts, attitudes, and skills learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the participant or trainer
2	**Knowledge – objective** (K-O)	Principles, facts, attitudes and skills learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardised test
3	**Behaviour/expertise – subjective** (BE-S)	Measures that evaluate changes in on-the-job behaviour perceived by participants, or global perceptions by peers or a supervisor
4	**Behaviour/expertise – objective** (BE-O)	Tangible results that evaluate changes in on-the-job behaviour or supervisor ratings of specific observable behaviours
5	**System results/performance – subjective** (SP-S)	Organisation results perceived by respondents, not reported by company records (for example, subordinates' job satisfaction or commitment to the organization), and group effectiveness perceived by subordinates.
6	**System results/performance – objective** (SP-O)	Tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings

(Source: Collins and Holton, 2004, p.225, adapted with inclusion of Reaction)

A further category, "reactions" was included to distinguish between participants' comments on the quality of the programme's delivery, design, the venue and so on and the "knowledge-subjective" outcome category. Reactions are not considered to be an outcome, but it was felt to be useful to include this category for the purposes of differentiation.

Table 16. Data extraction form

Citation information	
Author	
Year	
Title	
Journal	
Volume	
Issue	
Pages	
Date	
Keywords	
Abstract	
Name of database	
URL	
Quality Assessment (1-3 rating)	
Quality of theory & context	
Quality of methodology	
Quality of results and analysis	
Contribution to Knowledge	
Descriptive information	
Study site and location	
Programme type (including format, content, teaching & learning methods)	
Group composition	
Customised/ open, individually/ organisationally focused	
Evaluation and findings	
Evaluation data collection methods	
Evaluation data analysis methods	
Outcome categories 1-6: knowledge/learning, behaviour/expertise, system results/performance. Each category either subjective, objective or both	
Comments on evaluation methodology, e.g. limitations	
Key findings	
Implications/contribution	
Summary	
Research notes	

Appendix H: Shortlisting and quality appraisal

Tables 17 and 18 below provide a summary of the numbers of articles under consideration at each stage in the systematic review selection process. Two search strings were used following advice that the first was insufficient to capture all relevant material. Ideally, the search would have been conducted only once, using all three terms, 'evaluation', 'impact' and 'effectiveness', and duplicate articles from each database would have been recorded so as to increase traceability. These are learning points for future literature searches but are not felt to have compromised the quality of the database search, they merely made the process more laborious for the researcher.

Any article identified through the search of one database that was also found in one of the other databases was reviewed only once. At the stage of searching abstracts, references from the various databases were exported to the bibliographic management software, Endnote, into a combined list. The number of duplicate articles has not been recorded since if the article was already in Endnote, it was not exported again from the next database to be searched. The databases were searched in the order: Proquest, EBSCO, CSA Social Sciences. So if, for example, an article appeared in Proquest and the reference details were exported into Endnote they were *not* exported again even if it appeared in EBSCO or CSA Social Sciences. Similarly, when the second search string was used, using the terms 'impact' and 'effectiveness' instead of 'evaluation', reference details of articles with relevant abstracts were only exported into Endnote if they were unique, ie they had not already been exported from search string 1.

As such, in tables 17 and 18 below, the number of total hits and hits considered relevant from title for search string 1 and search string 2 include duplicates, whilst the number of abstracts comprises only unique articles.

Table 17. Number of articles at each stage of selection process for search string 1

Database	Total hits	Relevant from title	Relevant from abstract
Proquest	172	46	
EBSCO	63	16	23
CSA Social Sciences (ERIC & PsycINFO)	112	24	

Table 18. Number of articles at each stage of selection process for search string

Database	Total hits	Relevant from title	Relevant from abstract
Proquest	405	56	
EBSCO	246	38	32
CSA Social Sciences (ERIC & PsycINFO)	222	22	

The total number of unique articles selected to go through to the quality appraisal stage was 55. Of these, 25 met the required quality standard of no more than two scores of 1. The remainder either fell below the quality standard, or failed to meet the inclusion criteria.

Cross-referencing of the 25 articles which met the requisite quality standard identified a further 12 articles which were sourced and subjected to the same quality appraisal. This resulted in a further seven articles being selected for inclusion in the systematic review. Given the importance placed on cross-referencing as a source of identifying relevant material (Greenhalgh and Peacock, 2005), this is a relatively low number, but is in keeping with the review timeframe of articles published in the last ten years: particularly for those articles published in or shortly after 2000, there is little opportunity for them to have referenced other articles published in or after 2000.

At this stage it was decided to remove one article as it was a replica of another article (Dexter and Prince, 2007a; Dexter and Prince, 2007b), i.e. whilst published in a different journal, it was written by the same authors and described exactly the same programme and evaluation process and as such was felt not to have added to the review. The article excluded received a lower score than the one which was kept in the review.

A total of 31 articles from the combined search list and including cross-referenced articles made it through the quality grading process and into the systematic review. Table 19 below summarises the quality appraisal process described above.

Table 19. Number of articles included in and passing quality appraisal stage

Stage in review process	Number of articles
Shortlisted articles prior to cross-referencing	55
Articles meeting quality standard	25
Articles shortlisted through cross-referencing	12
Cross-referenced articles meeting quality standard	7
Replica articles	1
Total number of articles for inclusion in systematic review	**31**

The articles shortlisted for systematic review are listed in table 20 below, by author, showing date published and title, which gives an indication of the scope of the article.

Table 20. List of review articles by author

Author(s)	Date	Title
Alampay, & Morgan	2000	Evaluating external executive education at Dow Chemical: its impact and the Pygmalion effect
Black & Westwood	2004	Evaluating the development of a multidisciplinary leadership team in a cancer-center
Blackler	2004	The design and evaluation of a leadership programme for experienced chief executives from the public sector
Block & Manning	2007	A systemic approach to developing frontline leaders in healthcare
Boaden	2006	Leadership development: does it make a difference?
Brundrett	2006	Evaluating the Individual and Combined Impact of National Leadership Programmes in England: Perceptions and Practices
Bush & Glover	2005	Leadership development for early headship: the New Visions experience
Carr et al	2009	Leadership for health improvement - implementation and evaluation

Cunningham & Kitson	2000	An evaluation of the RCN clinical leadership development programme: part 2
De Casterlé et al	2008	Impact of clinical leadership development on the clinical leader, nursing team and care-giving process: A case study
Dexter & Prince	2007	Evaluating the impact of leadership development: a case study
Duda	2008	The impact of a leadership development course for junior academic medical faculty
Dvir et al	2002	Impact of transformational leadership on follower development and performance: a field experiment
Edmonstone & Western	2002	Leadership development in health care: What do we know?
Ford et al	2008	A partnership approach to leadership development for directors of nursing in older people's services in Ireland--Articulating the impact
Frese et al	2003	Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision
Hamilton & Cooper	2001	The impact of outdoor management development (OMD) programmes
Hancock et al	2005	The impact of Leading Empowered Organisations (LEO) on leadership development in nursing
Lynch, Leo & Downing	2006	Context dependent learning: its value and impact for workplace education
McClean & Moss	2003	They're happy, but did they make a difference? Applying Kirkpatrick's framework to the evaluation of a national leadership program
Miller et al	2007	Linking learning methods to outcomes in public health leadership development
Orr	2007	Learning advanced leadership: findings from a leadership development programme for new superintendents
Parry & Sinha	2005	Researching the Trainability of Transformational Organizational Leadership
Sadler-Smith & Shefy	2007	Developing Intuitive Awareness in Management Education
Skipton Leonard & Maynard	2003	Leadership Development as an Intervention for Organizational Transformation A Case Study
Tushman et al	2007	Relevance and Rigor: Executive Education as a Lever in Shaping Practice and Research
Tyson & Ward	2004	The Use of 360 Degree Feedback Technique in the Evaluation of Management Development
Umble et al	2006	Evaluating the imact of the Management Academy for Public Health: developing entrepreneurial managers and organisations
Wilson & Corrall	2008	Developing public library managers as leaders
Woltring et al	2003	Does leadership training make a difference? The CDC/UC public health leadership institute: 1991-1999
Yorks, Beechler, & Ciporen	2007	Enhancing the Impact of an Open-Enrollment Executive Program Through Assessment

Journals providing review articles

The spread of journals from which these 31 articles came is summarised in table 21 below. As can be seen, the majority of journals included provided a single article, whilst three journals, the *Academy of Management Journal*, *Leadership & Organization Development Journal* and *Leadership in Health Services* each provided three articles, and five journals provided two articles. There is an emphasis on journals from the health sector, with seven journals reflecting healthcare. The subject focus of the chosen articles will be discussed in more detail in the next section.

Table 21. Journals providing articles for systematic review

Journal Title	Number of articles
Academy of Management Journal	3
Leadership & Organization Development Journal	3
Leadership in Health Services	3
Human Resource Development International	2
Journal of Nursing Management	2
Journal of Public Health Management and Practice	2
Management Learning	2
School Leadership & Management	2
Academy of Management Learning & Education	1
Consulting Psychology Journal	1
Education & Training	1
Educational Management Administration & Leadership	1
International Journal of Health Care Quality Assurance	1
Journal of European Industrial Training	1
Journal of Health Organization and Management	1
Journal of Management in Medicine	1
Library Management	1
Nursing Standard	1
Personnel Psychology	1
The Canadian Journal of Program Evaluation	1

Year of publication of review articles

Table 22 below shows how many articles were published in each of the 10 years covered by the systematic review, from 2000 to 2009. 2007 yielded a particularly large number of articles relative to other years, for which there is no obvious explanation. The other years each provided between one and four articles.

Table 22.Year of publication of articles in systematic review

Year published	Number of articles
2009	1
2008	4
2007	7
2006	4
2005	3
2004	3
2003	4
2002	2
2001	1
2000	2

Appendix I: Detailed systematic review findings

Sector

The healthcare sector dominated amongst the articles reviewed. Whilst management and leadership development is an area of investment for this sector, this skew does not represent the proportional spend by this sector relative to the commercial sector or other parts of the public sector. Much more likely is a combination of the emphasis in the health sector on evidence-based decision-making, the requirement for externally-commissioned evaluation to demonstrate value for public spend, access to the sector for researchers, and the likelihood of the investigators coming from an academic background and hence being more likely to write up the results into publishable articles.

Even with the health sector excluded, the proportion of articles from the public sector is greater than from the private sector, probably for similar reasons to those articulated above.

Table 23. Sector of review articles

Sector	Number of articles
Health	14
Local government	2
Education	2
Manufacturing	2
Public service	2
Agriculture	1
Human resources	1
National government	1
Non-profit research organisation	1
Technology & Communications	1
Varied	4

Programme type

Most of the programmes in the review articles were customised programmes, i.e. commissioned by and designed for a specific organisation or sector. This definition of customised is further split between those programmes designed at sector level for individuals in particular types of roles where individuals apply or are invited onto the programme, e.g. school heads (Bush and Glover, 2005; Brundrett, 2006), NHS HR Directors (Boaden, 2006), Public health managers (Carr et al., 2009), and programmes designed for specific organisations, e.g. Derby City Council (Dexter and Prince, 2007a) to support organisational development.

A small number of the programmes under consideration were open programmes, i.e. those run by a provider for which managers from a range of sectors and roles apply, e.g.

a senior executive general management programme run by a US University (Yorks et al., 2007) and a programme on developing intuitive awareness (Sadler-Smith and Shefy, 2007).

The type of programme will have consequences on what sort of evaluation is desirable and possible, a matter for discussion in later sections. For example, the extent to which it is realistic to assess organisational change from a programme attended by only one individual from an organisation is likely to be much more limited than a programme designed with the specific intent of bringing about cultural change within the organisation.

Content focus

The overwhelming focus of the programmes featured in the review articles was leadership. A small number (n=3) were described as management development, rather than leadership development programmes, but even these showed areas of overlap, e.g. two of the three programmes that were billed as 'management' programmes had leadership development as an explicit component (Yorks et al., 2007; Lynch et al., 2006). A couple of programmes focussed on a very specific topic, such as developing intuitive awareness (Sadler-Smith and Shefy, 2007).

Whilst the headline topic area was predominantly leadership or management, most articles were explicit about a number of sub-categories within the programme. Personal development, communication skills, change, culture, conflict resolution, problem solving and teamwork all featured in multiple articles, and several concepts of leadership were made explicit, including collaborative leadership, transformational leadership and distributed leadership. Context also featured strongly, for example, an article on leadership development for junior medical faculty (Duda, 2008) describes one programme objective as being about learning how to identify and secure leadership positions in academic medicine, whilst programmes focused on school leadership (Bush and Glover, 2005; Brundrett, 2006) talk about the context for school improvement and purpose of schools, and the *Leadership for Health Improvement Programme* (Carr et al., 2009) relates all programme objectives back to the context of public health.

All articles described the objectives set for the programme and desired outcomes, which framed the choice of content and teaching and learning methods used.

Programme design and teaching & learning methods

The programmes described in the review articles ranged from, at the extreme shortest, a one-day course (Duda, 2008) and two-day transformational leadership programme (Parry and Sinha, 2005) through to a 14-month leadership programme divided into six 'learning events' (Carr et al., 2009). Whilst the shortest of the programmes described were run with straightforward face-to-face delivery, most of the other programmes comprised a variety of learning media, and teaching and learning methods. Typically, there were blocks of residential workshops interspersed with activities designed to embed and enhance the learning, such as online seminars and support, action learning

groups, work-based/action research projects, coaching and mentoring. Within the workshops, the pattern was for active participation of learners and group work as well as personal development-focused activities and some more formal teaching.

For example, the *Leading Managers Programme* (Dexter and Prince, 2007a), designed for Derby City Council, comprised six modules over 12 months, with 360 feedback, personal action planning and learning logs, whilst the *Trent Leadership Development Programme* (Edmonstone and Western, 2002) split participants into Development Groups in which they worked in action learning sets and received personal mentoring. Participants also created a learning network with the whole-group, attended conferences and had single-issue workshops. The *Canadian Agriculture Lifetime Leadership* (McLean and Moss, 2003) revolved around a computer-mediated conference and six face-to-face seminars and an open programme run by a US university for senior executives was an intensive four-week residential programme with content sessions and an action learning component (Yorks et al., 2007).

This range of designs and teaching and learning methodologies reflects the trend in management and leadership development toward action-orientation and application to embed learning, and demonstrates the use of multiple mechanisms by which to bring about learning and change.

Outcome categories

Collins and Holton's (2004) list of six outcome categories (itself derived from Burke and Day's (1986) criterion measures and Swanson and Holton's *Results Assessment System* (1999) was used as a means of categorising the level at which outcomes were evaluated in each of the programmes under review. Collins and Holton took their analysis further by using meta-analytical techniques to manipulate quantitative data in each of their review articles in order to produce an effect size for each of their six categories.

This review seeks only to classify the level at which outcomes were measured and, provide an indication of whether the result was found to be positive, neutral/inconclusive or negative in each case. A meta-analysis was not attempted since a number of the articles under review used only qualitative data collection and analysis techniques, which are not conducive to inclusion within the methodology of meta-analysis.

Table 24 shows the outcome categories used along with a concise description of each category.

Table 24. Collins and Holton's (2004) Outcome categories, adapted

#	Outcome category	Description of category
R x n	**Reactions**	Satisfaction with how programme was run, including quality of faculty, venue, sessions, materials and design, generally assessed immediately after a session or programme.
1	**Knowledge – subjective** **(K-S)**	Principles, facts, attitudes, and skills learned during or by the end of training as communicated in statements of opinion, belief, or judgment completed by the participant or trainer
2	**Knowledge – objective** **(K-O)**	Principles, facts, attitudes and skills learned during or by the end of training by objective means, such as number of errors made or number of solutions reached, or by standardised test
3	**Behaviour/expertise – subjective** **(BE-S)**	Measures that evaluate changes in on-the-job behaviour perceived by participants, or global perceptions by peers or a supervisor
4	**Behaviour/expertise – objective** **(BE-O)**	Tangible results that evaluate changes in on-the-job behaviour or supervisor ratings of specific observable behaviours
5	**System results/performance – subjective** **(SP-S)**	Organisation results perceived by respondents, not reported by company records (for example, subordinates' job satisfaction or commitment to the organization), and group effectiveness perceived by subordinates.
6	**System results/performance – objective** **(SP-O)**	Tangible results, such as reduced costs, improved quality or quantity, promotions, and reduced number of errors in making performance ratings

(Source: Collins and Holton, 2004, p.225, adapted with inclusion of Reactions)

For the purposes of differentiation, "reactions" was included as a category so as to ensure that participants' comments on the programme's design, quality of delivery and so on were not confused with the first outcome category, "knowledge-subjective". Reactions to the programme will not be discussed in terms of programme outcomes. That is not to say they are not valuable, and attention should be paid to how stakeholders respond to a programme since negative feedback can indicate reasons why learning and performance outcomes are not achieved.

A synthesis of the levels of outcome assessed in each of the articles under review is shown in table 25 below. The table also shows how many outcome levels were assessed in each programme evaluation, and the total number of programmes assessing outcomes at a particular level.

Table 25. Outcomes assessed in each review article, using Collins and Holton's (2004) Outcome categories

Author(s)	Outcome category							No. Levels assessed
	Rxn	1 K-S	2 K-O	3 BE-S	4 BE-O	5 SP-S	6 SP-O	
Alampay, & Morgan	Y	Y			Y			2
Black & Westwood				Y				1
Blackler	Y	Y		Y				2
Block & Manning	Y		Y	Y	Y			3
Boaden	Y		Y	Y		Y		3
Brundrett		Y		Y		Y		3
Bush & Glover	Y	Y		Y		Y		3
Carr et al	Y	Y	Y	Y				3
Cunningham & Kitson			Y	Y	Y	Y	Y	5
De Casterlé et al				Y		Y	Y	3
Dexter & Prince		Y		Y		Y		3
Duda		Y		Y				2
Dvir et al		Y	Y	Y	Y			4
Edmonstone & Western				Y	Y	Y		3
Ford et al	Y	Y		Y				2
Frese et al			Y	Y				2
Hamilton & Cooper			Y		Y			2
Hancock et al	Y			Y	Y			3
Lynch, Leo & Downing		Y		Y		Y		3
McClean & Moss	Y	Y		Y	Y	Y		4
Miller et al		Y		Y	Y	Y		4
Orr	Y	Y		Y				2
Parry & Sinha					Y			1
Sadler-Smith & Shefy				Y				1
Skipton Leonard & Maynard	Y				Y	Y		2
Tushman et al		Y		Y		Y		3
Tyson & Ward	Y	Y		Y	Y		Y	4
Umble et al		Y	Y	Y		Y		4
Wilson & Corrall	Y		Y	Y				2
Woltring et al				Y				1
Yorks, Beechler, & Ciporen			Y	Y	Y			3
Total number of articles assessing each outcome level		**17**	**10**	**27**	**13**	**13**	**3**	

It can be seen that subjective measures of assessment (n=57) were far more prevalent than objective measures (n=26) across the board, whether in the context of knowledge, behaviour/expertise or systems results/performance. The single category which was

assessed more than any other was behaviour/expertise-subjective. Both of these patterns are likely to be due at least in part to practical factors: it is easier to access programme participants - who can provide subjective assessments of their own behaviour/expertise - than other stakeholders, and subjective measures are easier to develop than objective measures. There also appears to be widespread recognition of the fact that learning in itself is necessary but not sufficient, and it is the relation of what is learned to changes in practice that is important and should be measured.

A substantial number of studies sought to identify systems outcomes (n=16), particularly subjective outcomes, with half of studies attempting to collect data at this level. This is perhaps surprising given that many of the programmes focused on individuals from disparate organisations yet it is encouraging that attempts are being made at trying to correlate programme attendance with not just individual but organisational improvement. However, only a small number of studies attempted objective system results/performance-level evaluation. This, I would suggest, is indicative of the challenges in designing measures that are sufficiently high level to provide an assessment of organisational outcomes, yet can be attributed specifically to a leadership or management development programme.

Most of the evaluations described attempted to assess either two or three categories of outcome (see table 26 below). A handful focused on only one or on four categories, whilst none attempted to assess all six outcome categories. This is likely to be partly down to resource constraints and partly due to other variables, such as who the evaluators were, what the underlying purpose of conducting the evaluation was and the difficulties discussed above, and in earlier literature, of relating programme impact to system outcomes. None of the programmes assessing a single level of outcome used just knowledge, whether subjective or objective, as the unit of analysis: knowledge was only ever used alongside a measure of behaviour/expertise and/or system results/performance.

Table 26. Number of outcome categories assessed in each review article

Number of categories assessed	Number of articles
One category	4
Two categories	9
Three categories	12
Four categories	5
Five categories	1
Six categories	0
Total	**31**

It should be noted that in some of the articles it was quite difficult to pin down the category of outcome being assessed as this approach to separating out outcomes at different levels was not congruent with the evaluation design. For example, several of the articles were case studies that attempted a holistic portrayal of impact rather than to subdivide impact into different categories.

Evaluation designs, data collection and analysis methods

The design classifications included by Collins and Holton (2004) in their meta-analysis of the impact of managerial leadership development programmes were: Pre-test Post-test With Control (PPWC), Post-test Only With Control (POWC) and Single Group Pre-test Post-test (SGPP). This expanded the design selection from that used in Burke and Day's (1986) earlier meta-analysis, which included only studies with a control or comparison group. This review has not made research design an inclusion/exclusion criterion, i.e. all research designs were considered. The review studies were categorised as falling into one of four categories: the three used in Collins and Holton's study, plus a fourth: Single Group Post-test Only (SGPO).

Table 27. Evaluation design

Number of categories assessed	Number of articles
Pre-test Post-test With Control (PPWC)	2
Post-test Only With Control (POWC)	2
Single Group Pre-test Post-test (SGPP)	8
Single Group Post-test Only (SGPO)	19

As table 27 above shows, the predominant evaluation design (n=19) was Single Group Post-test Only (SGPO), whilst a quarter of the studies employed a Single Group Pre-test Post-test (SGPP) design. Typically, the SGPO designs were multi-method, including interviews, surveys, focus groups and, less frequently, participant observation (McLean and Moss, 2003; Orr, 2007; De Casterlé et al., 2008; Wilson and Corrall, 2008). Several were described explicitly as case studies (Brundrett, 2006; Orr, 2007) and specifically of an instrinsic (Black and Westwood, 2004) or instrumental (De Casterlé et al., 2008) type, again implying a multi-method approach.

SGPP designs were facilitated through the use of repeat 360s (Yorks et al., 2007; Parry and Sinha, 2005; Skipton and Goff, 2003; Tyson and Ward, 2004) and feedback against capabilities defined for the programme (Boaden, 2006), pre-programme surveys to determine development needs and programme expectations compared with similar surveys applied post-programme (Bush and Glover, 2005; Wilson and Corrall, 2008) and essays on leadership perspectives collected prior to programme commencement (Orr, 2007) compared with attitudes to leadership post-programme. A couple of the SGPO designs used a retrospective pre-test, whereby questions were posed after the programme about behaviours exhibited before as well as after the programme (Duda, 2008; Miller et al., 2007). Miller et al., for example, used a 1-7 rating scale to assess the extent to which leadership behaviours taught on the programme were practised before and six months after the programme. This approach, whilst pragmatic, is not without the potential methodological flaw of recall bias.

Of the two Pre-test Post-test With Control (PPWC) studies, Frese et al (2003) used a non-equivalent dependent variable design (Cook and Campbell, 1979, p.118) where the control was a set of variables, rather than a control group, to compare changes in

inspirational communication skills, which were taught, against other communication skills that had not been taught. The pre-post aspect of the design was implemented by independent raters assessing skill levels before training and again after training. To strengthen the experimental nature of the design and reduce bias, the raters were shown video recordings so that they were not privy to whether an individual's performance was before or after their training. Hamilton and Cooper's (2001) study of the impact of outdoor management development programmes assessed the experimental group one week before and one week after the programme using three pre-existing questionnaires: the Team climate inventory (TCI), Occupational Motivation Questionnaire (OMQ) and Pressure Management Indicator (PMI) and assessed a control group one month prior to the experimental group's training.

Of the POWC studies, Dvir et al (2002) employed a longitudinal, randomised field experiment in their study of the impact of transformational leadership on follower development and performance in a military setting. They compared the performance of experimental and control groups using a multitude of pre-existing instruments, including Hackman and Oldham's (1980) Growth Needs Index, a Collectivistic Orientation Scale based on Wagner (1995), a measure of critical-independent thinking and five standard military measures of performance: written and practical light weapons tests, a physical fitness test, obstacle course and marksmanship. It is difficult to imagine such an intensive experimental design being feasible, or even desirable, in less disciplined and structured organisational environments than the military. Alampay and Morgan (2000) applied a statistical regression model controlling for education, tenure and job level to compare performance outcomes of the experimental group with a control group several years after an executive education programme in the Dow Chemical Company.

The PPWC and POWC designs described above are in a clear minority relative to SGPP and SGPO designs. They are derived from an experimental mindset which contrasts with the realist philosophy (Pawson and Tilley, 1994; Robson, 2002) and explanatory intent, ie desire to explain and not just report results, reflected in many of the review studies. It is interesting to note that had the same evaluation design restrictions been applied as in Burke and Day's (1986) meta-analysis, only four of the 31 studies in the review would have been considered, or using Collins and Holton's (2004) selection criteria, only 12. This suggests a trend away from experimental and quasi-experimental evaluation designs in recent evaluation literature and toward more explanatory designs. It is possible that had Burke and Day and Collins and Holton searched for such designs, they would have found them, but it is certainly the case that in this review there were simply not the numbers of experimental designs as in the earlier meta-analyses.

A number of the survey instruments used were pre-existing instruments considered appropriate to evaluate the programme in question. Several of these have already been mentioned in the descriptions of the PPWC and POWC evaluation studies above. Others include Cunningham and Kitson's (2000) use of the Multi-factor Leadership Questionnaire (MLQ) (Parry and Sinha, 2005; Dvir et al., 2002; Bass and Avolio, 1990) and Organisation of Care instrument (Bowman and Thompson, 1995) in a study of the RCN *Clinical Leadership Development Programme*; the use of the Leadership Practices

Inventory (Kouzes and Posner, 1997) in the *Canadian Agriculture Lifetime Leadership* programme (McLean and Moss, 2003) and the PDI Profilor 360 instrument in a non-profit research organisation (Skipton and Goff, 2003). Other instruments were designed specifically for the programmes in question. The benefit of using a pre-existing tool is availability of documented evidence about its reliability and potential for benchmarking, whilst the strength of a bespoke survey or interview protocol is the specificity with which it can be designed to relate to the programme in question.

A further point to make about the evaluation designs is the longitudinal dimension incorporated into a number of studies. This is clearly an integral part of pre-test-post-test designs, but a number of post-test only designs also involved collecting data at several points during and/or after a programme and, for longstanding programmes, the collection of data over a number of years, e.g. Woltring et al's (2003) eight-year time horizon for evaluating the *Public Health Leadership Institute's* programmes and Wilson and Corrall's (2008) participant observation during the programme, review event six to nine months after programme completion, pre- and post-programme questionnaires and mid-programme and post-programme interviews.

Data analysis was less well described and reported in some of the review articles than other aspects of the studies. It is not possible to say conclusively in such cases whether the analysis process itself was weak, or whether it was just poorly reported. It can be said across the board, however, that the methods of data analysis used were congruent with the evaluation design and methods of data collection. For studies involving qualitative data, largely interviews and focus groups, but also documents, e.g. learning logs, thematic analysis was used extensively and described with varying degrees of detail. Black and Westwood (2004), for example, provide a detailed description of how their interviews were audio-taped, transcribed and analysed using multiple readings, with transcript comments and constructs categorised and then distilled down to eight overarching categories reported by all participants. In contrast, Brundrett (2006) simply states that MaxQDA software was used to facilitate qualitative analysis, whilst De Casterlé et al (2008) used NudIST to assist their thematic analysis. Carr et al (2009) took a collaborative approach to data analysis within the research team, between the research team and programme leaders, and between the research team and programme participants which was congruent with their guiding principles of appreciative inquiry, soft systems methodology and illuminative inquiry.

Whilst some of the thematic analyses appeared to be entirely open to the interpretation of the data emerging, others used pre-determined categories, such as Dexter and Prince (2007a) who classified comments into knowledge, skills and attributes, and then created sub-sections within each 'bucket', and Tushman et al (2007) who coded interview comments on scales of 1-5, from "no evidence of learning/change" to "great evidence", where each scale was tied to one of the four levels of the Kirkpatrick framework.

In addition to categorising interview comments, Dexter and Prince (2007a) also used vignettes to tell stories. Whilst this approach is useful for illustrative purposes, when quotes and stories are used exclusively without backup of other analysis methods, such as by Wilson and Corrall (2008), it implies a less than rigorous analytical process.

Survey data was analysed using a range of statistical tests, most commonly paired *t*-tests (Miller et al., 2007; Hamilton and Cooper, 2001; Cunningham and Kitson, 2000) but also a range of other tests depending on the particulars of what was being analysed and how. Parry and Sinha (2005), for example, used a Wilcoxon matched pair signed rank test to compare pre- and post-training scores for each of five transformational leadership factors, a Kruskal Wallis ANOVA (*H*-test) for differences between cohorts and a Mann Whitney *U*-test to compare the public and private sectors, whilst ANOVA and Repeat Measure ANOVA (RMANOVA) was used by Dvir et al. (2002) and Frese et al. (2003) for comparing pre- and post-training results.

In studies using both qualitative and quantitative methods, conclusions were drawn based on the story reflected by the combined data.

Limitations and compromises to the evaluation designs explicitly mentioned by authors included: the issue of attribution, lack of clear measures of programme success (Dexter and Prince, 2007a), limits on generalisability and transferability (Dexter and Prince, 2007a; Sadler-Smith and Shefy, 2007; Miller et al., 2007; Dvir et al., 2002), lack of long-term measurement of impact (Wilson and Corrall, 2008; Frese et al., 2003), over-reliance on self-report data (Sadler-Smith and Shefy, 2007; Frese et al., 2003), positive orientation towards development leading to positive respondent bias, both in participants opting to go on programmes where choice is voluntary, and in responding to evaluation surveys, interviews etc. (Orr, 2007) and influence of learning transfer factors outside of the remit of the evaluation (Tushman et al., 2007).

Effectiveness of management and leadership development

Each of the review articles was considered for its claims about the effectiveness of the management/leadership development programme under consideration. Articles were given a score of 4 to 1, where 4 = highly successful/programme objectives met in full, 3 = moderate evidence of success, 2= neutral/inconclusive and 1 = negative impact. Where the results and discussion allowed, such a rating was applied to each of the outcome categories assessed in the articles, e.g. an article assessing three outcome categories was considered for effectiveness in each of these three categories. In a minority of articles, results could not be clearly linked to outcome categories, and an overall assessment of effectiveness across outcome categories was made. Whilst informed by the results and discussion scoring an article, the difference between a score of 3 and 4 involved an element of subjectivity in the decision making process. Due to the range of evaluation methodologies used, many of which were not quantitative, it was not possible to calculate effect sizes based on statistical tests and synthesise these into a meta-analysis by as had been done in earlier studies.

It was found that no programmes reported overall negative effects. Having said this, there were instances of negative elements within an overall neutral or positive programme score, e.g. Yorks et al. (2007) where direct reports' scores of their line manager's behaviour were lower after the programme than before, and De Casterlé et al.

(2008) where nurses' direct reports didn't always consider changes in their line managers' behaviour after training to be positive.

A couple of studies resulted in ambivalent/inconclusive findings, namely Alampay and Morgan (2000) where the Pygmalion effect seems to have been in evidence: in studying the job performance of participants relative to a control group, they found that participants' performance improved, but that this began in the year that participants were selected, rather than following the programme, i.e. the very process of being chosen to take part in a development experience seems to have boosted participants' confidence, morale and/or other factors that contribute to performance. In Hamilton and Cooper's (2001) study of the impact of outdoor management development, three of the four main assessment scales showed no improvement following training. This was explained through the high levels of pressure and low levels of mental well-being, as identified through the Pressure Management Indicator (PMI), being experienced by 50% of participants. Finally, Skipton Leonard and Maynard (2003) obtained no measureable changes in individual skills or behaviours, but found there was strong evidence for significant improvement in organisational functioning as a result of leadership development.

Notwithstanding these few instances of neutral/inconclusive effect, all other studies showed at least some degree of positive impact, with some studies showing high levels of effectiveness, where for example, all programme objectives were achieved (Boaden, 2006; Black and Westwood, 2004) or all assessment measures gave positive results (Parry and Sinha, 2005; Frese et al., 2003).

There was some evidence to suggest that lower levels of outcome, i.e. those related to participant knowledge and behaviour/expertise as opposed to systemic outcomes, elicited greater success than higher levels. For example, McClean and Moss (2003) found the evidence for the *Canadian Agriculture Lifetime Leadership* (CALL) programme to be more compelling at lower than higher levels of the Kirkpatrick framework, and Cunningham and Kitson (2000) measured significant improvements in leadership performance of nurses across a number of dimensions, but only *indications* of patient improvement. Similarly, de Casterlé et al. (2008) reported increased effectiveness of a head nurse who had undergone leadership training in the areas of self-awareness, communication and vision, improvements across the team in terms of communication, responsibility, empowerment and job clarity, and an apparent, but anecdotal, improvement in patient-centred communication, continuity of care and interdisciplinary collaboration. Stronger evidence of effectiveness at lower levels of outcome is not surprising given the challenges of attribution and appropriate measurement at higher levels, but the value of attempting this type of evaluation should not be underplayed. It is noteworthy, for example, that of the four articles that attempted evaluation at only one outcome level, none of these relied on knowledge-outcomes as a means of assessing programme effectiveness. In all cases where knowledge outcomes were assessed, at least one behaviour/expertise or systems-level outcome category was also used to add strength to the findings.

Part of the challenge in attempting to distil the results down to a single measure of effectiveness is that a number of the evaluations were richly descriptive and simple reporting of an effect size is incompatible with the evaluation intent and paradigm. Further to this, many of the studies took a relativist stance, comparing which outcomes from a programme were achieved to greatest and least effect, and/or an explanatory stance, for example trying to identify which teaching and learning approaches were most impactful in achieving outcomes, or what factors helped and hindered the outcomes being achieved. For example, Brundrett's (2006) sectoral study of educational leadership programmes found that the *HEADLAMP* programme for headteachers produced a diffuse set of experiences and results, whilst the *National Professional Qualification for Headship* (NPQH) and *Leadership Programme for Service Headteachers* (LPSH) were both found to be consistently valuable – the design features which enabled this value were then discussed in detail. Orr's (2007) study of a leadership programme for superintendents reported how the programme led to improvement in three areas: 1. replicating ideas and practices from the seminar series, 2. establishing goal direction for participants' district and board work, and 3. strengthening participants' general leadership capability. In addition to mapping these areas of effect, the study also found that the collaborative enquiry process was the most valuable aspect of the programme design. Tushman et al. (2007) conducted a comparative analysis of the *Leading Change and Organisational Renewal* (LCOR) open programme and looked at its impact according to three design variables: degree of customisation, target audience, i.e. team or individual and seniority of participants. Whilst all programmes were found to have some degree of impact at individual and/or organisational level, the programme design with the most substantial impact was the customised senior team design.

It should be noted that the particular intent of a study affected how the results were written up and displayed. Returning to the notion of formative versus summative evaluation, where formative is about improvement and summative about demonstrating value added, a couple of studies described how assessment and feedback resulted in ongoing design improvements, e.g. Bush and Glover's (2005) evaluation of the *New Visions* programme for headteachers, and Orr's (2007) collaborative action research approach. These improvements are not documented with any degree of detail. It is highly likely that other studies used evaluation data for improvement even when the explicit purpose for publication is to demonstrate programme effectiveness.

Given the challenges of producing a synthesis of management and leadership development effectiveness across a diverse range of evaluation designs, using studies with great variation in their intent and approach, it can be well understood, in addition to the methodological requirements of meta-analysis, why Burke and Day (1986) and Collins and Holton (2004) limited the design types included in their field syntheses to those with quantifiable effect sizes. However, there is undoubtedly something lost when high quality studies are excluded because they do not conform to an experimental or quasi-experimental design and the picture is incomplete when complex evaluation studies are boiled down to numbers and statistics.

To fall foul of precisely this criticism, however, and make a sweeping overall summation based on the 31 articles in this review, the conclusion is that substantial evidence exists for the effectiveness of management and leadership development programmes in a range of contexts across a range of programme types and delivery mechanisms. Effectiveness varies according to a number of variables to do with the quality of the programme's design and delivery, the measures applied and external factors associated with learning transfer.

Appendix J: Client interview protocol

Introduction: Thank you for your time. This interview should take no more than 45 minutes. I would like to record and transcribe it, with your permission? The data will be stored confidentially. When writing up, I will use job titles and intend to name the programme, but individuals and companies will be anonymised.

To re-cap, I am researching how organisations make decisions about management/ leadership development programmes and the role that evaluation plays. Any questions before we start?

Background data *(gathered from document analysis; recap/check accuracy of data in interviews with lead organisational contact. For other interviewees, just provide brief summary of programme under discussion) Type of programme, programme design, programme content, scale i.e. number of iterations, number of participants, timeframe.*

What is your role in relation to the programme? *How involved are you? What type of involvement do you have? What is your job title?*

Why and how did the programme start? *Where did the idea come from? What was the business need? Who identified this need? Why was a training programme decided on as the way of addressing the need? Who decided? How did you identify and select the training provider? Who was the target group of participants – has this changed at all?*

Has the programme been a success? *How do you know? What evidence do you have? Who has benefited? What impact has the programme had? Who is interested? What information do they want to know?*

What has happened to the programme, and why? *Has it continued, or was it changed, or stopped? What prompted the decision? What evidence was used to reach this decision? Who was involved in deciding? Would you do anything differently next time?*

Have you done any evaluation? *Why did you decide to conduct/commission an evaluation? What was the main purpose? What were you seeking to achieve?*

How was the evaluation done? *Who did it? What data collection and analysis methods were used? When? Why was this approach taken? Was a report written?*

How were the results used? *Who wanted to know what/What information did different stakeholders want? How were the findings shared? Who saw the results? What happened as a result? How did the evaluation findings influence the decision about what happened to the programme? What other information was used when making decisions about the programme?*

How did you develop an evaluation design? Where does your expertise come from? Are you familiar with any models of evaluation?

Close: That is all I wanted to ask about the programme. Do you have any questions or comments? Is there anyone else I could speak to about the programme? Thank you for your time.

Appendix K: Client director interview protocol

Introduction: Thanks for your time. Interview should take approx. 30 minutes. Would like to record and transcribe, with your permission? Data will be stored confidentially. Would like to use names, job titles and name organisations when writing up.

 Re-cap on purpose of research: how organisations make decisions about investment in management and leadership development programmes and role of evaluation.

Background data: *confirm programme design, target participant group, no. interations run, longevity of programme.*

How did the programme start? *Tender? Direct approach? What was the business need the programme sought to address? Who was involved? What was the process? What was the client most interested in? How did they make the decision?*

Ongoing design & support: *Who is involved from the client side? What sort of involvement do they have? What information do they have and use to inform their view? What kind of feedback to they want from Ashridge?*

Is the programme working?: *How do you know? What sort of information do you use?*

Evaluation?: *How do you use the end-of-programme Happy sheets? Any other evaluation?*

Close: Thanks, that's all. Intend to collect data during March and April and write up in May/June. Will circulate summary results to all involved, having checked each company vignette with relevant clients and client directors first.

Appendix L: Nvivo coding tree

Parent node	Child node	Grandchild node	Refs	Refs	Refs	Refs
Background data		1	1			
	Company information		6	7		
	Interviewee role in company		8	10		
	Interviewee role in programme		11	12		
	Payment for programme		1	1		
		Central budget		3	6	
		Paid for by participants' division		7	8	
	Programme design		14	29		
	Programme history		8	13		
	Target population		13	22		
Decision-making & management structures		1	1			
	Ongoing dialogue		5	6		
	Programme steering group		7	16		
	Report to Board, board decision		8	11		
	Review meeting		9	13		
	Steering group for all L&D programmes		9	18		
	Tender & selection panel		9	10		
Decision-points		1	1			
	Change provider		1	1		
	Choose participants		12	29		
	Choose provider		15	29		
	Conceive of programme		14	31		
	Design & content		12	18		
	Modify programme		13	47		
	Continue programme					
	Postpone programme		5	7		
	Stop programme		13	21		
Evaluation - attitudes to		1	1			
	Ambivalent, uninterested		4	5		
	Critical, skeptical of		1	1		
	Difficult		1	2		
	Evaluation has limitations		6	6		

	Expertise		5	6		
	Important		7	9		
	Intensive evaluation not necessary		2	2		
	Interesting		1	1		
	Less attention paid when high scores or client happy		5	7		
	Need to know or do more		7	11		
	Systematic		2	3		
Evaluation level - Kirkpatrick		1	1			
	Behaviour		11	24		
	Learning		4	4		
	Reaction		15	32		
	Results		12	27		
Evaluation purpose		1	1			
	Control		2	3		
	Get buy-in from budget holders		1	1		
	Improve programme		16	31		
	Justify role or evaluate self		3	5		
	Keep board informed		3	3		
	Learn		1	1		
	Other		1	1		
	Prove value		9	23		
	Select future participants		2	3		
Evaluation time horizon		1	1			
	Long term		3	4		
	Short term		3	6		
Feedback collection methods		1	1			
	Formal, intentional		6	8		
		180 or repeat 360		5	8	
		Appraisal process		5	8	
		Control group comparison		2	2	
		Focus groups		2	2	
		Follow-up participant survey		4	5	
		Happy sheets or feedback form		19	51	

		HR participants report back		2	3	
		Interviews with line managers		3	3	
		Interviews with participants		8	17	
		Participants send in feedback		1	1	
		Profiling past participants		3	5	
			Participant promotion	6	8	
			Participant retention v. turnover	2	2	
		Project report-backs, live case		3	5	
		Social network assessment tool		1	1	
		Staff survey tools		2	3	
	Informal, incidental			6	13	

		Chance conversations after programme	12	28	
		Conversations during programme	6	8	
		First hand experience of programme	14	34	
		Following participant progress	5	7	
		Gut feel	2	2	
		Number of nominations for next iteration	1	1	
		Others reporting change in participant behaviour	3	4	
		Participant feedback filtered back via senior management	6	11	

Rationale, decision drivers			1	1	
	Business & strategy drivers		1	1	

		Budget, cost	8	10	
		Company growth	1	1	
		Create aligned strategic & leadership thinking	2	3	
		Internationalisation of company	5	6	
		New organisational strategy	2	2	
		Organisational re-structuring, change	2	5	
		Participant networking	5	6	
		Topical business issues	9	16	

	HR-drivers					
		Changes to L&D strategy		7	10	
		Develop specific management population		8	15	
		Fit niche in HR L&D Leadership strategy		9	12	
		Identified competence gap		10	15	
		Retain talent		3	9	
	Individual-drivers		2	2		
		CEO's opinion		3	4	
		Key stakeholders leaving role		4	6	
		Lack of HR advocacy		1	2	
		New in post		5	5	
		Own ideas		5	5	
		Senior management presence		1	1	
	Programme-drivers					
		Good content and design		1	1	
		Insufficient participant nominations		4	4	
		It's new		1	1	
		Negative evidence from programme		6	7	
		Population exhausted		10	11	
		Programme feedback		5	6	
		Review of pilot		2	2	
		What's worked on other programmes		3	6	
	Provider-drivers					
		Advice from provider		1	3	
		Chemistry with faculty		2	5	
		Direct prior experience of provider		6	9	
		Impression during visit or tender		5	7	
		Provider reputation		6	7	
		Tailored approach		3	3	
Speculative, hypothetical		13	39			
Stakeholders, people		1	1			
	Ashridge					
		Ashridge administrator		2	2	
		Ashridge faculty		10	16	

		Client director		17	40	
	Client					
		Business sponsor		4	8	
		Chief Executive		10	24	
		Executive board or team		15	31	
		HR		14	33	
		HR Director		13	27	
		L&D		6	10	
		Line manager		6	9	
		Managing director		2	4	
		Participants		6	15	
		Programme manager		16	59	
		Programme owner		7	20	
		Senior management		18	48	

Appendix M: Sample coding

The example below shows codes assigned to one of the eight programme vignettes. The vignettes were generated by integrating material from the two or three interviews relating to a particular programme, supported by documentary secondary data. Each of the interview transcripts was coded in NVivo qualitative analysis software using the coding tree shown in Appendix L above.

Leading as a [Company A] Manager, Company A

Vignette text	Coding
The CEO of Company A chairs a Capability Council, of which one of the four capabilities in focus is leadership. Beneath the Capability Council is the Leadership Forum, comprised of senior HR and business leaders, which sets the overall agenda for leadership development and has been the instigator of a new leadership curriculum. Individual programmes are overseen by a reference or steering group which reports to the Leadership Forum and, ultimately, the Capability Council. The programme manager for *Leading as a [Company A] Manager* has the wider role of Global Leadership Capability Manager and is responsible for all programmes in the leadership curriculum.	Stakeholders/Client/Chief executive Decision-making & management structures Stakeholders/Client/HR, Stakeholders/Client/Senior management Decision-making & management structures/Programme steering group Stakeholders/Client/Programme manager Background data/interviewee role in company
These changes are relatively recent, and at the time that *Leading as a [Company A] Manager* was instigated, company A's Leadership Academy would commission programmes on request where a gap or business need had been identified. In the case of *Leading as a [Company A] Manager* a senior HR manager *"did the rounds"* [client director] with a colleague to identify potential providers. This was followed by a formal tendering process and a presentation to HR and line managers, from which a preferred provider was selected. The same senior HR manager who had led the provider identification process was influential in this decision.	Decision-making & management structures Decision-point/Conceive of programme Decision-point/Choose provider Stakeholders/Programme manager, Senior HR Decision-making & management structures/ Tender & selection panel Stakeholders/HR manager, line managers
The current curriculum of leadership	Decision-making & management structures

programmes is designed around leadership 'elements' (competencies) for leaders at specific transition points in the organisation. *Leading as a [Company A] Manager* was identified as a programme suitable for supporting leaders as they transition into roles where they will be leading other managers and/or global functions. As a consequence, the specificity with which candidates are selected has increased, although the population from which programme participants are drawn is largely the same as before. Another reason for greater emphasis on the selection process is that the programme had become highly popular, and a new HR Vice President recognised that requests to participate were driving selection, rather than it being a strategic choice. The selection policy was changed when the original programme sponsor retired. *"Individually, these people got a lot out of it, but they were perhaps not the right people any more...They are not part of a group that has a lot of future. They are good performers but they don't have necessarily the potential to go on."[Client director]*	Rationale, decision drivers/HR drivers/develop specific management population, fit niche in HR L&D Leadership strategy Background data/target population Decision-points/choose participants Decision-points/choose participants Stakeholders/client/senior management Rationale, decision-drivers/individual drivers/key stakeholders leaving role Rationale, decision-drivers/fit niche in HR L&D Leadership strategy Decision-points/choose participants
The programme manager, together with a steering group, decides what the programme should focus on and any substantive re-design issues. Minor alterations are dealt with by the programme manager and client director. The steering group comprises HR and business managers, including a senior business sponsor, and has become less formal than when it was first set up. The senior HR team is kept informed of any changes.	Decision-point/design & content, modify programme Decision-making & management structures/steering group Stakeholders/client/programme manager, senior HR, business sponsor
The first iteration of the programme was run as a pilot. After the pilot, and subsequent iterations of the programme, changes were made based on participant feedback from the programme review	Decision-point/modify programme Evaluation purpose/Improve programme Feedback collection methods/formal, intentional/feedback form

forms, conversations with participants, and the personal experience of the programme manager and faculty team who were present throughout the programme. More feedback was solicited from participants of the pilot than of later programmes, though the same range of data sources were used. The programme manager or another member of the HR team always attends the programme. In the early iterations, a representative of the Global Leadership team and an executive board member were present throughout. Now, senior executives attend and take a role in delivering the programme. The CEO, amongst others, has commented that this is the *"best programme ever" [Client director]* and past participants who have been promoted to more senior roles in the company have also been strong advocates.	Stakeholders/client/programme manager Stakeholders/Ashridge/client director, faculty Feedback collection methods/informal, incidental/conversations during programme, first-hand experience of programme. Stakeholders/client/programme manager, HR Stakeholders/client/senior management Background data/programme design Stakeholders/client/senior management Stakeholders/feedback collection methods/informal, incidental/participant feedback filtered back via senior management
Evolving business needs are anticipated to be an ongoing reason for adapting the programme. Two reasons were given for why the programme might be terminated in the future: the participant population has been exhausted, or an alternative training programme or resource requirement has taken priority. In terms of deciding to keep running the programme, it appears that the visibility of and to senior managers is very positive, and important: *"It makes a big impact. It's a very natural flow....That is part of the success: the fact that we have senior management and they open up." [Client director]*	Decision-point/modify programme Rationale, decision-drivers/Business & strategy drivers/changing business needs Decision-point/stop programme Rationale, decision-drivers/programme-drivers/population exhausted Rationale, decision-drivers/Business & strategy drivers/changing business needs Decision-point/continue programme Rationale, decision-drivers/individual drivers/senior management presence Evaluation level – Kirkpatrick/results
Evaluation *"Happy sheets" [Programme manager]* are distributed at the end of each programme. The provider's standard programme review form is used, which is similar to the form company A uses for other training programmes. Participants	Feedback collection methods/formal, intentional/feedback form Evaluation-level – Kirkpatrick/reactions

rate how well the programme met its stated objectives, the value to them as an individual and the extent to which they would recommend the provider. They are also asked to rate individual sessions and they have an opportunity to write comments. Participants' scores are aggregated and a summary produced which is distributed by the programme administrator to the faculty team and the programme manager.	Stakeholders/Ashridge/Ashridge administrator, Ashridge faculty/Client director
The results from the *"happy sheets"* *[Programme manager]* give the programme manager confidence that the programme is going well. When the scores are high, as they have typically been, little is done with the results. It appears that closer attention is paid when other factors, e.g. the recent focus on transitions as part of the new leadership strategy, prompt a need to redesign the programme: *"When we went to redesign I went into the survey results to see where the highlights are and where we can see that it's not that good, the results are a little bit lower and we see what we can do, see where we can change it." [Programme manager]*	Feedback collection methods/formal, intentional/feedback form Evaluation purpose/Control Evaluation – attitudes to/Less attention paid when scores high or client happy Evaluation purpose/Improve Rationale, decision-drivers/Business & Strategy drivers/new organisational strategy Decision-points/modify programme Decision-points/modify programme Evaluation purpose/Improve
The programme manager also has informal conversations with participants during and after the programme about what they thought about the topics covered and how they might take them further.	Feedback collection methods/informal, incidental/conversations during programme, conversations after programme
A survey was conducted in 2007 of participants on all five 2007 iterations of the leadership programme. This was initiated by the programme manager at the time. Participants were asked questions about the impact of the programme on their self-perception, their behaviour in specific areas and on the company; and specific questions about	Feedback collection methods/formal, intentional/follow-up participant survey Stakeholders/client/programme manager Stakeholders/client/participants Evaluation level – Kirkpatrick/behaviour, results. Evaluation purpose/Prove value

post-programme contact with line managers, colleagues and other course participants. It was not clear how the results of this evaluation were used and what, if any, influence they had.	
There is a desire from the HR side to pay more attention to demonstrating value: *"What we do need to do, I think, is to have a stronger relation about how there is value for the business, the return on investment for programmes like this...As a whole we need to look at the effect that the [leadership] curriculum has on the whole organisation and the development of the organisation over a longer time frame".* [Programme manager]	Stakeholders/HR, programme manager Evaluation purpose/prove value Evaluation level - Kirkpatrick/results Evaluation time horizon/long term
From other quarters there is scepticism about the purpose and value of evaluation: *"I also think...these people have to give themselves a reason for existence. And they think that having a lot of papers, a lot of evaluation helps. But I think it's true: it's when the CEO meets someone who's been on the programme and says, 'so how was it?' I think that's the more valuable feedback."* [Client director]	Evaluation – attitudes to/critical, sceptical of Evaluation purpose/justify role or evaluate self Stakeholders/client/Chief Executive, Participants Feedback collection methods/informal, incidental/chance conversations after programme
Summary Evaluation of *Leading as a [Company A] Manager* comprises feedback forms distributed at the end of the programme seeking reaction-level feedback and a survey focusing on impact was conducted in 2007, though it is not clear how the data from this was used or how influential it was. The feedback forms are used to check that participants are satisfied with the programme, and to support design modifications. Other contributors to design	Feedback collection methods /formal, intentional/feedback form Evaluation level- Kirkpatrick/reaction Feedback collection methods /formal, intentional/follow-up participant survey Evaluation level – Kirkpatrick/results Evaluation purpose/control, improve Decision-point/modify programme Rationale, decision-drivers/Business & strategy drivers/new organisational strategy Rationale, decision-drivers/programme

modifications are changes in the leadership strategy, informal participant feedback and firsthand experience of the programme. The programme has high visibility with senior managers in company A and one of its features of perceived success is the involvement of these senior figures in delivering the programme. The programme sits within a well integrated support and management structure.	drivers/programme feedback Feedback collection methods/firsthand experience of programme, conversations during and after programme Stakeholders/senior management Rationale, decision-drivers/individual drivers/senior management presence Decision-making & management structures