ANNA-MAIJA LIUHANEN

How are University Evaluations used?
– The Perspectives of two Finnish Universities

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Economics and Administration of the University of Tampere, for public discussion in Paavo Koli Auditorium (Kanslerinrinne 1, Tampere), on January 11th, 2008, at 12 o’clock.

University of Tampere
How are University Evaluations used?

The Perspectives of two Finnish Universities
# Contents

Acknowledgements ........................................................................................................... 7  
Abstract ......................................................................................................................... 9  

## INTRODUCTION

1.1 Research on the Utilisation of Evaluation ............................................................... 12  
1.2 The Purpose of this Study and the Research Question .............................................. 15  
1.3 Evaluation of Twenty Finnish Universities ............................................................. 16  
1.4 Research Design ....................................................................................................... 19  

## 2. THE CONTEXT OF FINNISH UNIVERSITY EVALUATIONS

2.1 Changes in Finland and in the International Environment ...................................... 26  
2.2 The Internal Context of University Evaluations ...................................................... 35  
2.3 Concluding Remarks ............................................................................................... 37  

## 3. EVALUATION AND QUALITY ASSURANCE – POLITICAL JACKKNIVES AND TOOLS FOR IMPROVEMENT AND ACCOUNTABILITY

3.1 From Peer Review to Quality Management .............................................................. 39  
3.2 Terminology in Evaluation of Higher Education ..................................................... 42  
3.3 Purposes of Evaluation .......................................................................................... 46  
3.4 External Evaluation of Higher Education ............................................................... 48  
    3.4.1 The Appearance of Quality Assurance Agencies ........................................ 48  
    3.4.2 The Four-step Model .................................................................................. 50  
3.5 Approaches in General Evaluation ......................................................................... 56  
    3.5.1 Methodology-based Approaches ................................................................... 59  
    3.5.2 Criteria-based Approaches ......................................................................... 62  
3.6 Opposite Trends ...................................................................................................... 63  

## 4. UTILISATION OF EVALUATION – THE NAME OF THE GAME

4.1 The Concept of Utilisation (Use) ............................................................................. 66  
4.2 Utilisation – the Primary Purpose and the Main Problem of Evaluation ................. 70  
4.3 Evaluation is used for Different Purposes .............................................................. 71  
4.4 Various Users .......................................................................................................... 77  
4.5 Process Use and the Utilisation of Results ............................................................. 78
Acknowledgements

The first time I decided I would like to learn more about evaluation was in summer 1993, straight after the institutional evaluation of the University of Oulu. We were waiting for the report of the peer review team at the time. However, some years passed before the idea of writing a doctoral thesis on university evaluations occurred to me. Those who encouraged me were Professors Pentti Meklin and Kauko Hämäläinen – Pentti by sending a handsome list of evaluation literature, and Kauko, a colleague for several years at the Finnish Higher Education Evaluation Council, by urging me to get started. Sincere thanks to both.

In 2000 I was accepted into the second doctoral programme arranged by the Ministry of Education for its staff in cooperation with the Department of Political Science at the University of Helsinki. Participating colleagues and programme leaders, Professors Markku Temmes and Turo Virtanen, provided the first arena for discussion. It was clear to me that my thesis would be about university evaluations, but I had not quite decided on the approach. The decisive suggestion came from Professor Evert Vedung, who recommended the utilisation perspective. My incomplete thesis “found a home” when Professor Seppo Höltä took up his post with the Higher Education Group at the University of Tampere. Since then Seppo and Pentti have acted as my supervisors. My warmest thanks for their comments, advice, criticism and support.

Special thank should also be given to the rectors of the two case study universities, who willingly opened the doors of their universities to me, and to all the other interviewees who – like their rectors – generously found the time to be interviewed. Further, warm thanks go to the rectors of three other Finnish universities, who took time to consider the credibility and transferability of my draft description of the utilisation of university evaluation.

Further, I am grateful to the dissertation reviewers – Professors David Dill and Jarmo Vakkuri – for their criticism and suggestions.

I also wish to thank those staff members of the Higher Education Group who word-processed all the interviews.

Dr. Ian Dobson is a person who knows more about the literal detail of my thesis than anyone else. I have learned a lot during the process of language checking, through his very professional approach to my use of the English language. Without his expertise, the English in which this thesis has been written might have been harder to understand. I hope that I have not added too many words to Ian’s dictionary of “Finglish”.
Warm thanks belong also to all the dear friends who kept asking how I was doing, and encouraged me in so many different ways.

Last but not least I thank my sons Sasu and Vesa, who in the middle of their own studies and at the early stage of their own professional careers have patiently listened to mother talking about her dissertation, and – above all – solved a variety of technical problems.

Oulu, 21 November 2007
Anna-Maija Liuhanen
Abstract

The external evaluation of higher education has been a growing business in Europe since the 1980s. The first countries to undertake evaluations were the United Kingdom, the Netherlands, Denmark and France. Others followed, and the number of evaluation or quality assessment agencies of various types grew quickly. Their approaches, however, varied. In some countries the focus was on degree programmes, while in others – such as Finland – the focus was on higher education institutions. There were also differences in the terminology used. While quality assessment and quality assurance were the prevalent terms, Finland – like France – chose to speak of evaluation.

The Finnish government decided that all universities should be evaluated. External evaluation was not warmly welcomed by higher education institutions; rather it was considered to be interference and a control mechanism. Even though the emphasis in Finland from the very beginning was on improvement, evaluation raised doubts and questions within universities. Perhaps this was to be expected.

One of the purposes of university evaluations was to improve Finnish universities’ capacity for change in a situation where the old central model of governance was giving way to a newer, more devolved approach. The new approach allowed the universities to exercise more power over their internal affairs.

The aim of this study was to establish how universities utilised their evaluations. The question was considered from the perspective of two Finnish universities. The overarching research question “how” included the purposes of evaluation use, the users of the evaluation, and the different areas of use such as teaching, research, the service function or regional role of universities (the so-called third task), and university management. Further, it included the question why universities might or might not use the results of their evaluations.

The study indicates that participation in the process prompted a range of ways of utilising evaluation. Second, it shows that people in different positions used the evaluation for purposes typical of those in their own group. Third, it suggests that in addition to the institution and its internal culture, several factors played an important role in the utilisation of evaluation: individuals, interaction and the capacity for evaluation to be used.

The utilisation of evaluation was also considered from the perspectives of a learning organisation and of organisational learning, assuming that there could be a connection between them and the utilisation of evaluation. Signs of organisational learning as it related to the evaluation could be found at both case-study universities.
Key words:
Evaluation, quality assessment, utilisation, higher education, university, learning organisation, organisational learning
Introduction

“Luck is where preparedness meets opportunity.”
Text on the back of a t-shirt, summer 2003.

Over the past 20 years external evaluation of higher education – or quality assurance, as it is often called – and evaluation in general have been growth industries. Major national and international programmes and individual organisations have all been subjected to evaluation. In Europe, the introduction of various European Union programmes in particular has brought about an increase in the number of evaluations. Vedung (2003) refers to a ‘wave of evaluation’, and Albaek (1997) asks why everyone is so keen to undertake evaluations. Higher education has not escaped the boom. In Finland, the external evaluation of higher education was first introduced in the mid 1980s, in research. Soon thereafter it was extended to higher education institutions and degree programmes. In this respect, Finland followed international developments relating to changes in the relationship between governments and higher education. For Finnish and other continental European universities, the changes have meant increasing autonomy. Evaluation was introduced as a counter-balancing element. So far, the external evaluation of higher education has been conducted mainly at the national level, but in line with the Bologna process, its focus has increasingly become international. Due to a range of national and political expectations, the external evaluation of higher education has not always been conducted with both the Government and the universities having the same purpose in mind. Thus, evaluation has become a sort of multi-purpose tool – a jackknife with different blades for different purposes.

In 1995 Finland followed other European countries and established the Finnish Higher Education Evaluation Council (FINHEEC), which is an independent expert body. The duties of FINHEEC are stipulated by a decree (1320/1995), according to which the Council is

- to assist higher education institutions and the Ministry of Education in issues relating to evaluation,
- to conduct evaluations for the accreditation of polytechnics,
- to organise evaluations on the operations of higher education institutions and on higher education policy,
- to take initiatives to develop higher education and its evaluation,
- to engage in international cooperation in evaluation, and to promote research on evaluation of higher education.
As evaluation is practical by nature, the use to which it is put is considered to be the main criterion of a successful evaluation. There is a strong emphasis on improvement in evaluations of higher education institutions and programmes in Finland. This can be seen, for example, in the FINHEEC principles that include utility (actually ‘impact’ which is said to mean utility; Action Plan 2000–2003), and in the national policy document in which are guidelines laid out for the evaluation and quality assurance of higher education (Ministry of Education 6:2004).

However, little research has been undertaken on how higher education institutions use external (or university) evaluations or how useful these evaluations are considered to be. The studies by Moitus and Seppälä (2004), and Hämäläinen and Kantola (2002), and various follow-up evaluations have focused on the dissemination and implementation of recommendations, that is, the (so called) rational use to which these evaluations can be put. However, these studies do not discuss the (so called) political and cultural use of evaluation (Albaek 1997) or the ‘enlightenment’ which might come from them.

Understanding the utilisation of institutional evaluation is of special interest now, as Finland is implementing a second round of institutional evaluations, namely an audit of the internal quality assurance systems used by each Finnish higher education institution. This study could also be of interest to decision-makers both in Finnish higher education institutions and at the Ministry of Education as well as to other quality assurance agencies and various student organisations.

1.1 Research on the Utilisation of Evaluation

Practical consequences are expected of different types of evaluation. In the literature, the consequences are considered from two perspectives – utilisation and impact. These two perspectives are close to each other in meaning, but are not identical. While utilisation can be considered to be a process, impact refers to the consequences of evaluation and / or its utilisation. The literature on utilisation relates mostly to program evaluation, i.e. the evaluation of major policy programmes. Research on external evaluation of higher education, in turn, is focused mainly on the impact of evaluations. For the purposes of this study, both lines of research are of interest.

In what follows, the so-called programme evaluation is referred to as general evaluation, and this term is used to cover all forms of evaluation other than the evaluation of higher education.

The three purposes most usually attributed to evaluation are accountability, improvement, and enlightenment. In the external evaluation of higher education, accountability and improvement are the most common, but there are other purposes as well, such as informing funding decisions, and assigning institutional status (Brennan
and Shah 2000, 31–32). According to Chelimsky, the purpose of an evaluation conditions the use that can be expected of it (Chelimsky 1997, 18). However, an evaluator can never know the purposes for which an evaluation will be used.

Research on the utilisation of evaluation focuses mainly on the utilisation of results (e.g. Shadish et al. 1991; Vedung 1997; Pawson and Tilley 2000). Focusing on the use of evaluation process is more recent (Forss et al. 2002, Patton 1998, Segerholm 2001, Valovirta 2002a). Other issues of interest in utilisation studies are the types or purposes of evaluation use, and the ways and means of facilitating evaluation use. There is no generally accepted theory on the utilisation of evaluations, but rather a range of lists of helpful means or strategies for enhancing the utilisation of evaluation has been drawn up, e.g. by Shadish et al. (1991, 55), and Vedung (1997, 279–287). Patton’s utilisation-focused evaluation is all about how to enhance the utilisation of evaluation (1997). Shadish et al., however, suggest an “ideal (never achievable) theory of evaluation” that would consist of five components: social programming, knowledge construction, valuing, knowledge use, and evaluation practice (Shadish et al. 1991, 36–64). In the context of this study the fourth component, knowledge use, is the most important of the five.

The development of evaluation approaches in general evaluation reflects changes in social science methodology, where different phases are described as evaluation generations (Shadish et al. 1991; Pawson and Tilley 2000, Heinonen 2001). The development from positivist towards various participatory evaluation approaches can be seen as an attempt to widen the involvement of those evaluated and thus to improve the utilisation of evaluation by offering a voice to those who will be or have been evaluated (e.g. Pohjola 1997; Patton 1997; Fetterman 2001). Evaluation approaches are discussed in Chapter 3.5 below.

The impacts of the external evaluation of higher education have been considered for instance, by Brennan and Shah (2000), and Dill (1997, 1998, 2000). These authors have taken an international perspective. Brennan and Shah focus on the impacts of external evaluation on higher education institutions in general, while Dill (2000) compares the characteristics and impacts of academic audits in the United Kingdom, Hong Kong, New Zealand and Sweden. Studies by Nilsson and Wahlén (1999), and Stensaker (1997 and 2000) have a national perspective. They all agree that the institutional approach, which considers universities as organisational entities (or focusing at the institutional level of universities), does not seem to filter down to the basic unit level of the institutions well enough. In contrast with other higher education researchers, Segerholm (2001) adopts the concept of utilisation when studying evaluations of Swedish higher education programmes. Research on the impact of audits and other institutional evaluations is mainly concerned with the impact in higher education generally. Even though the impacts of evaluation are obviously expected to be in line with the purposes of evaluation, Leeuw (2001a) points out that there can also be side-effects.
Evaluation method is an issue for evaluation both generally and specifically in higher education. In higher education, the word method is used mainly to refer to the so-called four-step model consisting of a coordinating body (mostly an agency), self-evaluation, external evaluation and site visit by a peer review team, and a published report. The model is discussed in Chapter 3.4.2.

There is considerable variation between countries in each of the steps. For example, Brennan prefers the term framework instead of model (1997, 11). In higher education evaluation, issues of interest in the discussion have been the legitimacy of evaluation (e.g. Brennan 1997; Brennan and Shah 2000; Vartiainen 2004), and different sanctions connected to evaluation (e.g. Franke 2002). A related perspective is that of power, discussed for example by Harvey (2004), Barnett (1994), and Brennan (1999, 225–227). The survey carried out by the European Network of Quality Assurance Agencies, ENQA, offers an overview of the approaches and methods used by different agencies (Quality Procedures in European Higher Education, 2003). Woodhouse (1998), in turn, discusses the future of the agencies and the methods used by them.

Context plays an important role in the utilisation of evaluation. As organisations, higher education institutions have some features typical among them, such as loose coupling (Weick 1976) and dual structure, consisting of academic and administrative (or enterprise) dimensions (Clark 1983). Becher and Kogan (1992) recognise four different levels in higher education, namely central authority, institution, basic unit and individual, and two different modes – normative and operational. Further, universities are bottom-heavy, in that academic expertise is located within the basic units (e.g. Birnbaum 1989; Clark 1983; Hölttä 1995). Disciplinary cultures also play a role (Becher 1989; Kekälä and Lehikoinen 2000). In addition to these common features, each higher education institution has its own culture related to its age, size, and disciplinary profile (Clark 1983; Välimaa 1995), which may be of importance when considering the utilisation of evaluation.

As to the external context, the most important differences relate to national traditions and regulations, or to put it in Clark’s words, to the relationships between state, market and academia (Clark 1983). In Finland and in other Nordic countries, the government–academia relationship is characterised by trust and dialogue (Hölttä 1995, 30–35; Smby and Stensaker 1999). The changes in the relationship between universities and the state, which to a great extent are behind the growth in the external evaluation of higher education, have been discussed for instance, by Bauer et al. (1999), Bleiklie et al. (2000), and Välimaa (1999). Hölttä (1993, 1995) and Rekilä (1998, 2003) have considered these changes from the Finnish perspective.

Research on the utilisation of evaluation relates mainly to methods and contexts that differ from those of higher education evaluation. In impact studies, which are more common in higher education, higher education institutions are mainly considered in general. Rarely is the focus on single institutions. However, research by Välimaa et al. (1998) focused on a single institution. Further, while impact studies focus on
change, utilisation studies are also interested in users and the purposes of and utilisation of evaluation. Thus, to date there is little empirical knowledge on the utilisation of evaluation in universities.

When considering the utilisation of research, Lampinen emphasises that it is important to study utilisation in real circumstances, in organisations and as a part of decision-making processes (Lampinen 1992, 9–10). This also holds for evaluation, especially now that the external evaluation of higher education is expanding.

1.2 The Purpose of this Study and the Research Question

The purpose of this study is to add to the existing knowledge about the utilisation of evaluations in higher education. The study focuses on the utilisation of institutional evaluations in Finnish universities, also referred to as “total evaluations”. In Finnish these are referred to as korkeakouluarviointi, yliopistoarviointi or kokonaisarviointi (Stenqvist 1993, 45; Välimaa 1994). The main attention is on evaluation utilisation in universities. In addition, the perspectives of the Ministry of Education and of the members of external evaluation panels established to assist in the evaluations (also referred to as peer review teams) are briefly considered. Thus, the focus is on the most important users – the universities and the Ministry of Education. Other users, such as external stakeholders and the media, have not been considered here.

Higher education institutions, their research activities, and the teaching they offer are evaluated in different ways and from various perspectives. What makes the utilisation of university evaluations especially interesting? First, the perspective of university evaluations with their organisational focus differs from the more traditional discipline-based evaluations of teaching and research. Second, in university evaluations, the FINHEEC ideas of tailoring the evaluations to the needs of each university and of ownership of the evaluation process were taken further than in other types of evaluations. (These ideas are discussed below, under Evaluation of Twenty Finnish Universities). The assumption at the agency was that ownership of the process would increase universities’ chance of using evaluations for improvement.

This study is not about how well the expectations of university evaluations were met, or whether universities followed the recommendations of external panels or not. Nor is it about the “amount” of utilisation – that is the “successfulness” of the evaluations, although these are discussed briefly.

Bearing these issues in mind, the research question for this thesis is how are Finnish university evaluations used? The answer to this question has been sought by constructing a conceptual framework of the utilisation of evaluation and its dimensions based on earlier research, and by applying that framework to the utilisation of university evaluations at two Finnish universities. The outcomes of this study are

How are University Evaluations used? – 15
comprised within an analytical description of the utilisation of institutional evaluation at the two universities. In addition to users within the case study universities, a brief description of how the Ministry of Education and the members of the peer review teams used the evaluation has been provided.

The main contribution of this study is that it describes the utilisation of evaluation in a university context. It does so by combining research on evaluation utilisation with an empirical study on the utilisation of Finnish university evaluations.

1.3 Evaluation of Twenty Finnish Universities

In 1986 Finnish Government decided that all universities should establish internal evaluation systems (Government decision of 25 September). Although the decision related to internal evaluation, it can be considered to be the first step towards the establishment of external evaluation. Five years later the Ministry of Education invited two universities to be pilot institutions for an institutional evaluation project. The aims of the project were

1) to collect information relevant to the improvement of the quality and performance of the universities,
2) to analyse the universities’ organisation and administration, changes, and capacity for change, and
3) to gain experience for establishing regular evaluation of Finnish universities (Stenqvist 1993, 45).

The purpose of the pilot project, expressed somewhat differently in separate documents, was twofold. On the one hand, it was hoped that an evaluation procedure that would strengthen the self-regulation of Finnish universities could be established. On the other hand, information for strategic planning within the Ministry and the universities would to be produced (Stenqvist, 29 March, 1992). A need for universities to become self-regulating and for them to take responsibility for maintaining and improving quality were emphasised (KOTA-työryhmän muistio 1985; Stenqvist 25 Nov, 1992; Rekilä 1996, 84). At this stage the evaluation scheme was not national but rather, the Ministry wanted to identify good examples (Director, Ministry of Education, 2000). Later, the main aim of the university evaluations was to support the improvement of Finnish universities by strengthening their institutional capacity for change and for self-regulation.

In 1995 the Government decided that all higher education institutions were to be evaluated by the year 2000 (Education and Research 2000). This meant that there would need to be university evaluations, also referred to as institutional evaluations.
By that year, the first six university evaluations had been coordinated by the Ministry of Education, with evaluations of the remaining 14 universities to be conducted by the Finnish Higher Education Evaluation Council FINHEEC.

University evaluations had been carried out in other countries including France, the United Kingdom, Sweden, Norway and Spain. (For reports on evaluations in France, the United Kingdom, and Sweden, see Hämäläinen et al. 2001a; for reports on evaluations in Norway and Spain, see Institusjonsevaluering av Universitetet i Bergen, 2001, and Garcia et al. 1995, respectively). In addition, the European University Association (EUA, previously CRE) had implemented an Institutional Evaluation Programme and thus conducted a number of institutional evaluations (Barblan 1996).

The Finnish Approach

Finnish university evaluations followed the four-step model for higher education outlined on page 5. The Ministry of Education provided the pilot universities with a list of themes to be covered in the self-evaluation and supplied evaluation reports from other countries to provide an indication of how to conduct and report on the self-evaluation. Despite the guidance offered by the Ministry, the pilot universities carried out their self-evaluations quite differently from each other. At one university the approach was based on surveys. At the other, each department was asked to contribute, based on a check-list of guidelines. (Liuhanen 1993, Sallinen et al. 1994, respectively). Subsequently, the different approaches were considered to have been beneficial, as together they offered a better basis for learning (Stenqvist 1993, 46).

Later, when FINHEEC coordinated the evaluations, the universities were free to choose the timing of their evaluation within a broad time frame, to decide on the detailed foci of evaluation, and to propose and veto members of peer review teams. The initial responsibility for the appointment of peer review teams lay with FINHEEC. This was part of ‘process ownership’, meaning that the universities undertake the evaluation for themselves rather than for FINHEEC or for the Ministry. Further, FINHEEC offered the universities the option of having an international quality assurance agency to conduct the evaluation. The aim was that each university should undertake an evaluation that could serve the university’s needs, and thus make evaluation an instrument for institutional improvement. Three Finnish universities were evaluated by the European University Association (EUA) as a part of the Finnish national programme (Three Finnish Universities in the International Perspective), and one Finnish university was evaluated by the European Foundation for Management Development (EFMD) (Foppen et al. 1998).

All 20 Finnish universities were evaluated. One evaluation focused on teaching (von Wright et al. 1995), one on administration (Virtanen and Mertano 1999), and four on universities’ regional role or external impact (Dahllöf et al. 1998, Goddard et
The evaluations of the remaining 14 universities focused on institutional structures and functions.

University evaluations can focus on inputs, processes, and/or outputs. The main focus of Finnish university evaluations was on management, decision-making and quality (Evaluation of Higher Education – the first four years, 28). Quantitative data were collected, but in the evaluations, performance indicators had a minor role only. The quality of teaching and research were not usually dealt with, but the teaching and research infrastructure was. As the terms ‘university evaluation’ and ‘institutional evaluation’ suggest, the evaluations were targeted at the organisation and its functioning. When introduced in the early 1990s, the organisational approach to university evaluations was novel one in Finnish higher education, compared with traditional disciplinary evaluations of research or teaching, targeted at the levels of basic units and individual academics. As to method, the four-step model outlined on page 5, typical of higher education evaluation was applied. The two main purposes usually attributed to higher education evaluations are improvement and accountability (Brennan 1999, 223, Höltä 1988; Vroijenstijn 1995). Barnett, however, speaks about enlightenment and surveillance (1994, 83). In Finnish university evaluations, improvement and change were emphasised, which however does not indicate the absence of accountability, but rather a combination of internal (improvement) and external (accountability) elements.

Each university organised the self-evaluation process according to its own needs. FINHEEC followed the policy established by the Ministry of Education, and did not provide institutions with a handbook for self-evaluation. Instead, both FINHEEC and the Ministry relied on the good examples which could be drawn from other countries, and later from other Finnish universities. A representative of FINHEEC (or the Ministry in the early cases) was available to consult with universities during their self-evaluations. Almost all self-evaluation reports were published. Peer review teams were appointed by the Ministry of Education, later by FINHEEC, after consultation with the university in question. The members of peer review teams were usually academics with expertise in evaluation, academic leadership and management. A match between the disciplinary profile of the university, and the peer review team members’ profile was sought. In addition to academic members, some teams also had external stakeholders, but there were no student members. Most peer review teams were comprised predominantly of international members, but each had at least one Finnish member. The need for this international scope is a consequence of the small size of the Finnish higher education system, in which people tend to be known to each other, with the possible consequence of weakening the credibility of Finnish evaluations. In addition, international peer review teams were expected to provide universities with new perspectives and ideas. However, with the exception of the evaluations conducted by the CRE (later the EUA) and the EFMD (European Foundation for Management Development), and additional role of the Finnish member on the peer review teams was
to ensure that the national circumstances were fully understood by the other members. In the more recent evaluations, the participation of peer review teams was increased so that the team was also able to participate in the planning phase of the evaluation. Following the completion of each evaluation, the university organised a seminar at which the evaluation report was made public and its contents discussed.

Reports
The reports of the external panels were published with the names of the members of the peer review teams being identified on the cover of the report. This included those evaluations carried out by the EUA and the EFMD. By doing this FINHEEC sought to emphasise the independence of the team, and to demonstrate that the peer review teams were considered to be responsible for the actual judgment and the recommendations. The reports were distributed to all Finnish higher education institutions, to the Finnish Parliament and Ministry of Education, to national and local authorities, to student organisations, and to other quality assurance agencies internationally.

No additional funding was provided to universities for the completion of university evaluations – not to carry out the evaluation or for e.g. the good quality of their report. The universities report annually to the Ministry of Education about how they have utilised the information drawn from evaluations.

For a more detailed description of the Finnish university evaluations, see Liuhanen 2001, 12–16.

1.4 Research Design
The aim of this study is to describe the utilisation of university evaluation at two Finnish universities. The formulation of the research question is based on earlier research on the utilisation of evaluation and on the experience I have drawn from my involvement in university evaluations as a FINHEEC member of staff. In order for an in-depth description to be provided, an analysis, interpretation, and understanding of the use of evaluation in a university context are needed. This calls for a hermeneutic approach that emphasises understanding instead of explanation, an internal perspective and empathy instead of external review (Niiniluoto 1999, 56). To obtain the internal perspective and to understand evaluation use within universities, case study and thematic interviews have been used. The internal perspective can not be reached through an analysis of documents, as self-evaluation reports and the reports of the peer review teams are usually a compromise of the views of those responsible for those reports. Reports represent the “official” truth, or management perspective, and do not report the different opinions or feelings of individual team members.
Case study has been described as ‘an exploration of a bounded system or a case or multiple cases over time through detailed, in-depth data collection involving multiple sources of information rich in context’ (1994, 61). It has also been characterised as a study of the particularity and complexity of a single case, coming to understand its activity within important circumstances (Stake 1995, xi), and as an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between the phenomenon under examination and the context are unclear (Yin 1994, 13–14). Out of these three definitions, Yin’s is closest to the methodology adopted for this study, as it emphasises the unclear boundaries between the phenomenon under examination, i.e. the utilisation of evaluation, and its context.

The use of case study has made it possible to reach a deeper understanding of certain universities and the manner in which they utilise evaluation than would have been possible through a survey of several or all twenty universities. Another reason why case studies and interviews are beneficial is that in their official reports to the Ministry of Education, universities report only a part of the actual / real utilisation of an evaluation. Thus, answers to questions such as who uses evaluation and for what purposes will not usually be found in the universities’ reports to the Ministry of Education, and neither in follow-up reports nor in other documents (Yin 1994, 3–9). Earlier research has offered explanations for this, but mainly in non-university contexts (e.g. Feinstein 2002; Hämäläinen and Kantola 2002; Vedung 1997). In official documents universities are mainly considered as whole entities, with no attention being paid to individuals or organisational units within universities which might hold different attitudes about the use of evaluation. Further, documents seldom register personal feelings or opinions. Their focus is on how evaluation is used by a university overall, while the focus of this study is on how evaluation was used at different organisational levels within a university.

I chose two universities for case study. In Yin’s terms this is a multiple case study, while Stake would call it a collective case study. Both emphasise the need for careful selection of the cases (Yin 1994, 38–53; Stake 1995, 3–4). Further, this is an instrumental case study to illustrate an issue, i.e. the utilisation of university evaluation (Stake 1995, 3). The study begins as a within-case analysis, and continues as cross-case analysis (Creswell 1994, 63). Stake emphasises the unique nature of each case: “We do not study a case primarily to understand other cases. Our first obligation is to understand this one case.” (Stake 1995, 4). Yin represents the opposite view and emphasises that case studies, like experiments, are generalisable to theoretical propositions, not to populations or universes (Yin 1994, 9–11, 30–32, 36). Yin calls it analytical generalisation. This study does not aim at generalisation, but the results offer a basis for analytical generalisation in the sense of propositions.

Interviews were the main source of information for the study. Interviews conducted at the two case universities represent the core of the research material. In addition, representatives of the Ministry of Education were interviewed, to complete the de-
scription from the perspective of the other intended user. Further, to strengthen the credibility and transferability of the analysis and interpretation, members of the peer review teams and the rectors of three other universities were interviewed, the latter after case descriptions had been drafted. Finally, both self-evaluation reports and the external reports have been used (Yin 1994, 13).

In this study, interviews provided the means of reaching the users of evaluation, and therefore the internal context of the utilisation of evaluation. Interviewing enables the researcher to get behind official documents. However, the description of evaluation use based on interviews is not perfect, but rather provides a glimpse of the experiences and thoughts of those interviewed (Hirsjärvi and Hurme 2001, 41). The interviews conducted for this study can be considered to have been thematic interviews. Thematic interviews are semi-structured, falling between structured and open-ended interviews. Instead of detailed questions, they rely on certain predetermined themes, and thus emphasise the voice of the interviewees rather than that of the researcher (Hirsjärvi and Hurme 2001, 47–48). According to Hirsjärvi and Hurme, thematic interviews are the most relevant for a study of this type because the interviewees can be active subjects in the research situation —"creating meanings", as they put it. In order to construct a description of evaluation use in a university, the views and interpretations of different actors are essential. A related point is that there is no clear picture of the utilisation of university evaluations in universities in various units at various levels, or by individual stakeholders. (Hirsjärvi and Hurme 2001, 35). The interview themes are based on the conceptual framework of the study presented in Chapter 4.8 below, and on the literature on evaluation use. For the themes, see Appendix 1. The starting point is that each interviewee was asked to consider evaluation use from his / her own perspective – the position held and / or personal interests, based on his/her experience. The picture that can be built up of evaluation use depends on the perspective of the interviewees.

Selection of the Case Study Universities
The twenty Finnish university-level higher education institutions differ considerably in their history, size, disciplinary profile, location, and internal culture. The university evaluations were, by definition, focused at the institutional level, but the special theme in five of them brought an additional element to those evaluations. As to the evaluation method, one part of it, namely how the self-evaluation was carried out, depended heavily on the universities themselves, i.e. on their internal context. As the purpose of the study is to increase understanding rather than to provide an explanation, the rationale for selecting cases is closer to Stake’s criterion of maximising learning (1995, 4), than to the one suggested by Yin which is based on either literal or theoretical replication (1994, 46).
To make sure that people still remembered the evaluation under examination, it was important to use universities that had been evaluated recently, thus excluding the pilot universities and others evaluated in the early 1990s. When looking for universities that had just had an evaluation, or were preparing for a follow-up evaluation, it seemed reasonable to assume that the follow-up evaluation would have refreshed people’s memories of the actual evaluation, and therefore to consider the measures and discussions based on those evaluations. The next question was whether the case study universities should be similar or different as to their approach, and/or as to their internal context. Approach refers both to evaluations that were focused generally at the institutional level, and to those with a special theme. In the end, it was decided to emphasise a similar approach at two different universities, assuming that a wider variety of evaluation use would be found, because of the context dependency of evaluation (e.g. Sinkkonen and Kinnunen 1994, 21; Valovirta 2000, 85–93; Feinstein 2002; Weiss 1995).

Why two? Why not a single case or several cases? The reason for having more than one was to seek more variation in evaluation use. The benefit of focusing on a single case would have been a deeper understanding of that university. For example, it would have been possible to extend the interviews down to the department level. Why not include all the twenty universities in the study, and get a more inclusive picture and deeper understanding? For reasons of research economy, the only possible way of reaching all 20 universities would have been via a survey. However, considering the research questions, a survey would not have been the best method, for reasons discussed above. Another alternative would have been to interview the rectors of all twenty universities. In that case the result would have offered a description of how the rectors or the universities use evaluation, or how each rector saw evaluation use within their own university. The point of this research is to examine evaluation use within universities, not only by universities, and to trace the actual users of the evaluations. Descriptions given by rectors would have been relatively brief.

The case universities are different from each other in many respects. One has a traditional disciplinary profile and in Finnish terms, it is old and rather large. The other is a good deal younger and smaller, with a more focused disciplinary profile. However, both universities share a strong research orientation. The theme of their evaluations was also similar: their external impact and regional role. At the time of the interviews one was planning for a follow-up evaluation and the other had just carried out a follow-up and was waiting for the report. The time span from the actual evaluation to this study and the interviews was three years in the first university, and five years in the other. Due to the follow-up evaluation, it was possible to seek additional evidence from the external evaluators of both universities.

The study does not aim a comparison between the two case study universities. However, comparison can not be completely avoided.
Qualitative Approach
On the continuum between qualitative and quantitative research, this study is closer to the qualitative end. The study relies on empirical analysis based on text, and it can be seen as solving a puzzle (Alasuutari 1999, 32–33; Töttö 2004, 10). Further, the perspective is that of those studied, the cases were selected to fit the purpose, and no hypotheses were set. (Creswell 1994, 16; Eskola and Suoranta 2000, 13–24). However, the analysis does not rely on empirical material alone, and the framework for analysis was based on earlier research. Even though the study (in Stake’s words), “shoots happenings rather than causes”, causes are not entirely excluded as the question ‘why’ indicates. Hence, the analysis has been characterised as a dialogue between the literature referred to and the study’s empirical material, with the researcher moving between the two and acting as an interpreter (Stake 1995, 8–9, 37–39).

My Position as Researcher
I have been involved in university evaluations in three different roles: first as an internal coordinator of one of the pilot evaluations, second as the senior FINHEEC advisor responsible for the coordination of 14 university evaluations, and third, as a researcher into university evaluations. The first two roles provided the pre-understanding of the university context and of university evaluations needed for the third, that of researcher, studying the utilisation of university evaluations (Niiniluoto 1999, 32).

During twelve years on the payroll of a Finnish university I worked at the institutional level and within a faculty. However, I learned most about how universities work during the years spent as the internal coordinator of the pilot evaluation, and as the coordinator of the consequent strategic process of the university.

As the FINHEEC coordinator of university evaluations there was involvement in the planning and coordination of the evaluations at the national level. In practice this meant advising and consulting with universities. After the appointment of the peer review teams by the Council, it meant recruiting the peers, and organising an introduction to the Finnish higher education system for them. In one of the follow-up evaluations I also joined the peer review team on their site visit, but not as an observer, not as a team member (Goddard, Teichler et al. 2003). Thus, I have a comprehensive overview of Finnish university evaluations.

Further, as national coordinator I was a member of the steering group for the institutional evaluation in both case study universities. However, I only attended the earliest meetings when most advice was needed. Thus, I was involved in both evaluation projects studied here, but not in the actual evaluation conducted by the peer review teams. This meant that to some extent I am studying my own work, which might be considered either an advantage or a handicap. The advantage is that in the steering group meetings I had an opportunity to learn about the universities, as an extension to my knowledge of FINHEEC. The handicap has been my closeness to
the agency, and the consequent danger of over-emphasising the agency perspective, and legitimating FINHEEC’s own work. However, rather than coming from a need to provide legitimacy, my interest in the utilisation of evaluations stems from questions and moments of doubt, and from a consequent need to gain a better and deeper understanding.

Many of the interviewees were known to me. It was not possible to exclude them from the interviews as in many cases they were the very persons who knew most about the evaluations and how they were used. Even though the themes considered in the interviews were not particularly sensitive, it is possible that the interviewees would want to give as good a picture as possible of the utilisation of evaluation in their university, knowing that they were “speaking to the national agency”. The interviews of several different persons in each university, representatives of the Ministry of Education, members of the peer review teams, and rectors of other universities, offer different perspectives on evaluation use. Together with evaluation reports this can be considered triangulation in the meaning of using various data sources (Eskola and Suoranta 2000, 68–74).

Limitations of the Study
As the interviewees were mainly rectors, deans, and institutional and faculty administrators, the study does not directly relate to the experience at the department level. However, among the interviewees there were both academics and administrators who had moved from the department level to new positions after the evaluation, and deans represent both the institutional and basic unit levels. Still, the voice of the basic units is weak, and the perspective is mainly that of key actors at institutional and faculty levels.

The study describes the utilisation of evaluation within two different universities. It does not offer a basis for generalisation to other universities or other organisations, but the results are generalisable to theoretical propositions which means analytical generalisation (Yin 1994, 9–11). Stake emphasises that the real business of case study is particularisation, not generalisation. However, even though the two case study descriptions cannot as such be transferred to other universities, they can offer the basis for what Stake calls ‘petite generalisations’. ‘Petite generalisation’ refers to a situation where refinement of understanding is reached through the case study, while in ‘grande generalisation’ the case study invites a modification of the generalisation by offering a counter-example (Stake 1995, 7–8).

Next Chapters
The study is divided into six chapters. This introduction is followed by a description of the context of Finnish university evaluations, consisting of both external and internal
contexts (Chapter 2). Under the title ‘External Context’ the national and international developments that influenced Finnish universities in the 1980s and 1990s are discussed. Internal context, in turn, refers to the character of universities as organisations.

In Chapter 3, evaluation and quality assurance are discussed, starting with an explanation of the terminology used. Then external evaluation of higher education is discussed, followed by the development of evaluation approaches in general evaluation.

Chapter 4 focuses on the concept of utilisation and on the various purposes of evaluation use. Further, different users and the utilisation of the evaluation process and its results are discussed.

Chapter 5, the empirical part of the study, begins with a description of the utilisation of evaluation in the case study universities. Then the two descriptions are combined to present the common features of utilisation in the case study universities.

Finally, in Chapter 6, the reliability of the study is considered.
2.1 Changes in Finland and in the International Environment

University evaluations were initially carried out during a time of change. First the policy context of Finnish university evaluations has been considered, starting with changes in the national system for governance of universities, hereafter described as the national steering system and certain other changes. Second, international developments that were behind the Finnish developments have been discussed. Third, the internal context of university evaluations has been described.

From the old Model...

Up until the early 1990s Finnish public sector administration was subjected to strong and direct central control. As state-run enterprises, universities are part of the public sector. The detailed government control mechanism covered organisational structures, recruitment of the major human resources, detailed allocation of financial resources, and the disciplines and degree programmes to be taught. The regulations covered, and still cover, the degree programmes to be taught, and the aims and structures of academic degrees, but not their content. (For a detailed description of the Finnish higher education system and how it is managed, see Hölttä 1995, 21–37; Hölttä and Rekilä 2003; and Ministry of Education 2004:20).

...through a Period of Changes

In addition to the changes in the national steering system, the policy context of Finnish universities also changed in other ways in the late 1980s and early 1990s (Hölttä 1988, Hölttä and Pulliainen 1991). At the beginning of the 1990s the Finnish economy went through a deep recession, which among other things resulted in a cut of 16 per cent in direct state funding to the universities (Rekil and Saarinen 1996). At the same time the establishment of the polytechnic (non-university) sector created a dual higher education system in Finland. Consequently, the universities came to face growing competition for funding, staff, and students, and a diversification of higher education institutions.
Finally, in 1995 Finland joined the European Union, which provided access to the Union’s research and structural funds. From the perspective of the university evaluations, the change in the government – university relationship can be considered to be the most important, as it emphasised universities at the institutional level.

... to Management by Results and Balancing Factors
The new system for university governance, labelled by the Ministry of Education as ‘steering-by-results’, reflects a shift from ‘steering’ through inputs and regulation, to ‘steering’ through outputs and information provision (Temmes et al. 2002, 9–14). The budget system was changed from line item to lump sum budgeting, and extensive deregulation and decentralisation occurred. The Finnish universities, which had traditionally been managerially weak at the institutional level, were expected to make decisions that had previously been made either by the Ministry of Education, the President of the Republic, or the Parliament. Among these decisions were the internal allocation of funding, and the appointment of professors. Thus universities were expected to take more responsibility for their activities, and to cope with a growing external complexity (Hölttä 1988 and 1995; Välimaa 1999). When the decision-making powers of the universities were increased, three balancing elements were introduced into the governance system. First, the Ministry of Education was to hold annual negotiations with each university to reach mutual agreement on objectives, results and appropriations. Second, small but important performance-based elements were included in the university appropriations. The third element was evaluation, which was required by the State to supply qualitative information about the universities’ performance. The Ministry of Education emphasised the need to give more decision-making powers to the universities on the one hand, and the importance of evaluation and reporting on their performance on the other (Jäppinen 1989; Rekilä 1996, 84). Lampinen characterised the change as a move from planning to evaluation and ‘steering’ (Lampinen 2003, 25). As a consequence of the policy changes, legislation concerning universities was amended, and in 1997 a new Universities’ Act replaced the former university-specific laws. The universities were given more autonomy to decide about their structure and administration, the authority of individual leaders increased compared with that of multi-member bodies, and the term of Rectors was extended from three to five years. The Act further included an obligation for universities to evaluate their activities and effectiveness, to be subject to external evaluations, and to publish the results of these evaluations (Universities’ Act § 5). The change in the policy for governance covered all education sectors. Accountability and transparency were a part of a major change in the whole education policy in 1991, when the Government accepted the first development plan for education and research (Laukkanen 1998, 140).
New Catchwords

At the time of these changes, a number of new catchwords, such as efficiency, effectiveness, accountability and responsiveness became a part of Finnish universities' vocabulary. An example is the title of the self-evaluation report of the University of Art and Design: “Quality, Efficiency and Effectiveness” (see also Kinnunen et al. 1998a, 16–17). Also, concepts such as management and leadership, and a culture of evaluation appeared in discussions about higher education (e.g. Rekilä 1996, 85; Välimaa 1999, 32–37). Quality, one of the main issues in many European countries, was less discussed in Finland. However, it does appear in some documents produced by the Ministry of Education prior to 1995. One of them states that even though the quality of higher education is not a problem in Finland, it is worthy of attention as a results-based steering mechanism emphasises quantity (Stenqvist March 29, 1992). The sudden economic depression of the early 1990s raised questions about how the quality of higher education could be maintained, and revealed a need for strategic management (Stenqvist, Keinonen, and Kells, 1993). (For more about the changes see Hölttä 1995; Kinnunen et al. 1998b, 13–14; Rekilä 2003; Pollit and Bouckaert 2000; Public Management Reforms: Five Country Studies 1997).

Evaluation comes...

The idea of systematic evaluation of Finnish universities was first presented in 1985 by the Ministry of Education’s KOTA working group, which considered the evaluation of the performance of higher education institutions (KOTA is a Finnish acronym for Korkeakoulujen toiminnan arviointi, which in English means ‘evaluation of the performance of higher education institutions’). The report stated that evaluation was needed in higher education, and when carried out professionally, could help universities to react to existing and anticipated needs, to improve the quality of their activities and to tackle their weaknesses. Thus the early ideas of the 1990s – responsiveness, quality, and capacity for change – were already there. A mere change in atmosphere towards self reflection would, according to the report, bring about positive changes in universities. The report recommended that two kinds of evaluations be conducted, those targeted to higher education institutions and their departments, and others to research and teaching and programmes in each discipline nationwide (KOTA 1985, 72). The KOTA report can be considered to be the first step taken by the Ministry of Education to move universities towards external evaluation and transparency, along with the first national research evaluation carried out by the Academy of Finland in 1984. A national university data base, also called KOTA, was established. As a consequence, quantitative data on the resources and performance of universities became publicly available (http://kotaplus.csc.fi:7777/online/Etusivu.do). In addition, the 1986 government decision insisted that universities build their own evaluation systems. Even though the 1985 KOTA report seemed to refer to external evaluation,
How are University Evaluations used?

evaluation was considered to be the responsibility of the universities (Rekilä 1996, 86). However, external discipline reviews commenced in 1990 (Humanistisen koulutusalan...1993; Alanen et al. 1992), and institutional evaluations followed in 1992 (Kogan et al. 1993; Davies et al. 1993). As recommended by the KOTA report, the combination of evaluations of research, institutions and disciplines thus became part of the Finnish higher education evaluation strategy. Also, the principle of evaluation as a tool for universities (as presented in the KOTA report) was adopted into Finnish higher education policy, and into the decree concerning FINHEEC, the Finnish Higher Education Evaluation Council (1320/1995).

... and raises Doubts in the Universities

Within universities the introduction of external evaluations led to heated discussion and debate. Their response can be seen as a defensive reaction of the academic community to external intervention, but obviously the economic recession of the early 1990s also had an influence on the discussion. Under the circumstances it was assumed that evaluations could lead to funding cuts, and that quantitative performance measurement would rule over qualitative evaluation. Result-based steering and management were criticised by university staff for not being appropriate to universities. Mälkiä and Vakkuri (1996, 89–121) provided a description of the critique and fears of that time. They also discussed the lack of trust that, according to them, would be a consequence of performance measurement and external evaluation. They considered two different models of external evaluation of higher education – evaluation for allocating resources, and evaluation for long-term development. However, the first evaluations of degree programmes and higher education institutions were planned in an atmosphere of mutual understanding between universities and Ministry of Education. The Ministry did not seek to establish a national evaluation scheme that led to resistance from universities, nor did they wish to destroy the positive atmosphere of cooperation which existed between the Ministry and the universities. Instead, evaluation was commenced on a voluntary basis with pilot evaluations in order to gain experience and to find examples of how to conduct university evaluations (Director, Ministry of Education, 2000). According to Hölttä, there was a wide degree of trust between the central and institutional levels of the higher education system (Hölttä 1995, 34). As indicated by Mälkiä and Vakkuri, the trust did not extend down to the level of academic departments.

Steering by Results is not enough

In spite of their preferred nomenclature with the steering by results system, the State was not interested only in results, but also in processes. Laukkanen, who considers the primary and secondary levels of education, points out that the central administration
did not seek to distance itself from local activities, but rather approached them by introducing new tools such as information-based management, and evaluation of effectiveness (Laukkanen 1998, 138). Developments in the higher education sector can be considered similar. As the quantitative results of universities were available in the KOTA database, evaluation of higher education programmes and institutions focused on processes. Only the evaluation of research was focused on results. Interpreting from the example below, one Finnish university’s experiences with the new steering system demonstrates that the State did not leave universities alone, but wanted to influence their operations – not only based on evaluation, but also through evaluation.

“Changes in the steering methods used by the Ministry … have conflicting effects on the autonomy of universities. On the one hand, the internal autonomy of the University … has increased with the use of framework budgeting. On the other hand, the use of performance criteria linked with allocation of funds, the structural reform through short-term development projects, the relatively strict definition of performance objectives, and the new personnel management policy, will diminish or at least restrict the University’s actual independence.” (Kinnunen et al. 1998b, 68)

**Major Change in nine Years**

By the end of the 1990s evaluation had become a key developmental tool not only for higher education, but more generally for Finnish public administration. There is, however, a difference between the evaluation of higher education institutions, and the evaluation of other public administration organisations. In the former, improvement is emphasised, while in the latter the emphasis is mainly on accountability and effectiveness (Harrinvirta et al 1998, 3; Valtionhallinnon arviointityöryhmän loppuraportti 1999, 1–5). The nine years between the launching of the launching of the pilot university evaluations to reporting the last ones (1991–2000) were a time of renewal for Finnish higher education policy and direction. By the time of the last evaluations, result-based steering had become routine, the economic situation was very good, the polytechnic sector was “ready” in the sense that no new institutions were to be established, Finland was already a member of the European Union, the Bologna process had been launched, and the regional or service role of universities had reappeared on the Finnish higher education policy agenda (Ministry of Education 2001). Further, external evaluation, including university evaluations, had become routine. Thus, the policy contexts of the first and last university evaluations were quite different. Over time, universities came to know what to expect from a university evaluation, and it was possible for them to learn from the evaluations of other universities, and about evaluation more generally.

The changes in Finland were a reflection of international developments. Universities throughout Western Europe faced major changes in late 1980s and early 1990s. Student numbers were generally on the increase, which meant more and different
types of students, and higher education had to cope with new expectations, and an increasing need for public funding (Clark 1997; Hölttä 1995; 9–14; Välimaa 1999). Another line of change was the growing dependence of enterprises on new information technologies, and consequently on universities. Further, the traditional mode of research was questioned by what was called a new form of knowledge production characterised by transdisciplinary, applied and useful research (“mode 2 knowledge production”) (Gibbons et al. 1994). The growth of national higher education systems, and their increasing economic importance added to the interest of governments in higher education (Välimaa 1999, 23–29).

…from central Regulation towards Self-regulation

Due to the growth and the increasing complexity of higher education systems and the changed expectations towards them, state regulation of higher education and implementation of changes over the system, became increasingly difficult (Brennan 1999, 223-224; Välimaa 1999). There was a change from central control to self-regulation. Regulation of inputs was replaced by control of outputs, and self-regulation by universities was the mechanism they used to meet output requirements.

According to Maassen and Stensaker, self-regulation refers to the capacity of an organisation to obtain, receive, and process information about itself, and to act on the basis of that information (Maassen and Stensaker 2003, 85–95). Hölttä uses “self-regulation” to refer to a new national strategy that transfers initiative and autonomy to universities, and control from inputs to outputs (Hölttä 1995, 15). State regulation is replaced by higher education institutions’ self-regulation. Kells and Stenqvist consider self-regulation to be more or less as synonymous with an evaluation culture. Self-regulation is something that is required from a university. It is not only the authority to regulate oneself, but also provides the responsibility and capacity to do so (Kells and Stenqvist 1994, 14–17). According to Rekilä the concept of self-regulation has not been widely or consistently used in Finnish higher education policy simply because there is no shared understanding about self-regulation in Finnish universities. She further argues that as a consequence, the role of evaluation in the government regulation of higher education has remained unclear (Rekilä 1996, 85–86).

The changes brought about unpredictability in the institutional environment, which in turn provided a challenge to institutional leadership. What was needed was institutional management – the capacity to change and respond to changes in the environment (Hölttä 1995; Välimaa 1999; Yorke 1999).

Evaluation as an Instrument for Steering and Learning

The introduction of processes of evaluation is closely linked to steering through information, which is one of the key elements in the theory of open systems (Birnbaum
Sinkkonen and Kinnunen describe information-based steering as feedback that enables comparison, and which might lead to change (Sinkkonen and Kinnunen 1994, 49–54). Because information-based steering is limited in its capacity to pre-determine the ways in which people’s behaviour might change, it is usually only accepted as a steering mechanism after considerable deliberation. Instead of coercive means, information-based steering calls for persuasive skills, expertise, reputation, and credibility (Temmes et al. 2002, 11–13). This is illustrated by House, who claims that evaluation persuades rather than convinces, argues rather than demonstrates, is credible rather than certain, and is variably accepted rather than compelling (1980, 73).

When considering different policy tools, Bleiklie et al. state that evaluation has gained acceptance as a learning tool. Learning tools are related to New Public Management ideology: “The idea that target groups can learn from experience is part and parcel of this ideology.” However, evaluation may also be used as an authority tool in order to create a basis for giving adequate orders or for distributing incentives in an efficient manner. (Bleiklie et al. 2000, 110–116). Considering this, surprisingly little discussion on learning has been connected to the various evaluation approaches.

**New Public Management**

The changes in governance relate to policy premises known as New Public Management (NPM), and to the rise of what is called the evaluative state (Neave 1998). The roots of NPM are to be found in the neo-liberal policy trend of the early 1980s that aimed at minimising state interference in market activities, and downsizing the public sector and public expenditure. Later, the policy aims shifted towards controlling the growth of the public sector, increasing the effectiveness and efficiency of public services, and improving the quality of public services. These were the means to the end of decentralisation, deregulation, and result-based steering (Harrinvirta et al 1998, 25). (For more about NPM, see Kogan and Hanney 2000, Haveri 2002, Laukkanen 1998, and Pollitt 1993).

The new governance models based on NPM were created for public administration in general. Thus, they do not pay attention to the special characteristics of higher education institutions, but rather consider them as being similar to other organisations (e.g. Bleiklie 1996, Kinnunen et al. 1998a, 16–17). An interpretation of the NPM policy at the university level provided in the self-evaluation report of a Finnish university, and points to the need for evaluation:

“One main objective in using evaluation as a steering instrument for higher education policy has been the improvement of the efficiency and performance of universities. The same development demand has been emphasised in the State budget since 1990. The prevailing neoliberal policies, emphasis on utilitarian and efficiency thinking, the principles of management by results, the reduction in resources, and the prioritisation of instruction and research have together led to the need for increasing and diversifying the evaluation..."
The increasing and diversifying evaluation of universities (Yorke 1999; Kinnunen et al. 1998a, 13–14) refers to an increasing state interest, which can be seen as conflicting with the decentralisation ideas of NPM and steering by results. Even though decentralisation was on the agenda, governments were interested in controlling a wider array of affairs. In this sense power was centralised rather than decentralised (Bleiklie et al. 2000, 51–55; see also Pollit 1993, 111). Thus the principle of decentralisation of the NPM ideology is not “pure”, but simultaneously there are opposite tendencies.

Why Evaluation?
In addition to the changes in state regulation, the increasing student numbers, and the interest in, and the increasing demand for external evaluation can be interpreted as a loss of consensus about purposes and standards in diverse higher education systems (Brennan 1999, 223–224). In Finland the loss of consensus is evident in the discussions about university and polytechnic education, especially the degrees which could be awarded by the two sectors (Review of Education Policy in Finland 2002; Pratt et al. 2004). The loss of consensus about the purpose and standards of higher education, and the new governance arrangements reflect a change in the relationship between governments and higher education. In addition to teaching and research, universities are expected to support social and economic developments more directly than before (Hölttä 1995, 15; Yorke 1999, 1). They are now considered a part of the national innovation system more clearly than before (Saari 2002, 145–147). As a consequence, the service function of universities, also called the third task or the regional role, was to be included in the new Universities’ Act and in the forthcoming funding formula. When the basic question is, whether research is useful for society, the old Humboldtian dichotomies of basic and applied, ‘pure’ and ‘dirty’ (sic) research are, according to Välimaa, no longer adequate. The Humboldtian university is being replaced by what Välimaa calls the pragmatic university, which should be responsive to the needs of wider and more varied stakeholder groups (Välimaa 1999, 24–25). The responsiveness of individual departments is not enough, but universities should be able to respond on an institution-wide basis to opportunities in, and threats from, the external environment (Yorke 1999, 7). This emphasises the role of institutional leadership and management (Kogan et al. 2000, 202). Bleiklie et al. refer to the above developments as redefinition of the main functions of higher education institutions (Bleiklie et al. 2000, 37, 40–46).

The quest for quality, and the establishment of quality assurance agencies was a reaction to the changed expectations. The quality discussion refers to a doubt that universities have the capacity to be able to react to the challenges (Yorke 1999, 2). Yet, referring to Trow, Brennan and Shah argue that there was no evidence of seri-
ous quality problems in higher education (Brennan and Shah 2000, 30). Rather, the problem was political (Harvey and Askling 2003, 70–71). In a situation of increasing student numbers and decreasing funding the question was whether higher education institutions and programmes could “deliver quality”. In spite of the fact that neither the increase in student numbers nor the decrease in government funding was equally radical in all OECD countries, the quest for quality spread to all of them. Westerheijden et al. (1994, 21) refer to policy borrowing. In spite of the opposite trends of student numbers and government funding, and of the efficiency requirements, the cost aspect is seldom considered in the evaluation of higher education. This may relate to the “automatic” increase of efficiency caused by the above mentioned opposite trends of funding and student numbers.

Bauer et al. (1999, 28) consider quality as an umbrella concept, which in many countries has been used to cover various reform intentions, such as an increasing need for effectiveness, innovation, and accountability. Askling and Stensaker argue that quality is a disguise of the NPM rhetoric, and a consequence of ongoing changes in state-university relationships, and increasing external influence on higher education (Askling and Stensaker 2002, 114–115; see also Lemaitre 2001). The utilitarian purposes mentioned above give an additional dimension to the discussion concerning quality (Bauer et al. 1999, 67–102). Quality no longer refers only to academic quality. In Clark’s terms, the locus of power between the three corners of the triangle in defining quality -academia, state, and market – is changing (Clark 1983, 143). The emphasis on customer perspective and competition refers to market, the utilitarian third task to both market and state.

The above developments provide the background of external evaluation in higher education, and the establishment of quality assurance agencies. The establishment of the Finnish Higher Education Evaluation Council did not bring about major changes in the national evaluation approach. Rather, evaluations conducted through FINHEEC continued very much along the lines of the earlier evaluations.

The assisting role given to the Council refers to improvement rather than to control, and corresponds with what Brennan and Shah assume: In cases where the level of state authority is high, the purpose of the agency is change (Brennan and Shah 2000, 28). Even though the word ‘change’ is not used in the decree, it is obvious that the Council was established to bring about and support change in higher education. According to the background memorandum for the decree, one of the Council’s tasks would be “... to make evaluation an integral part of the normal activities of the higher education institutions” (Ministry of Education, 1995). Thus the agency was not only for conducting external evaluations, but it was also expected to take the role of an educator or a change agent. The change to be implemented was the introduction of evaluation in higher education institutions.
2.2 The Internal Context of University Evaluations

In the literature universities are mainly considered as cultural institutions producing and disseminating knowledge (e.g. Becher 1989, Clark 1983, Birnbaum 1989). Becher and Kogan (1992) point out that through the training it provides, higher education also helps in the production of national wealth. Further, both Bleiklie et al. (2000, 46-51) and Bauer et al (1999, 78) discuss the public agency ideal, while Clark (1998) considers the characteristics of an entrepreneurial university that resemble those of the corporate enterprise ideal of Bleiklie et al. (2000, 50–51).

Changed expectations by governments and labour markets influenced perceptions of what type of organisation universities should be – autonomous cultural institutions, public agencies, or corporate enterprises (Bleiklie et al. 2000, 46–51). In Finland, universities have traditionally been combinations of public agency and cultural institution. As public agencies they are responsible for the education of future civil servants and professionals. The ideal of universities as cultural institutions relates to their task in knowledge production and to their autonomy as institutions, while the ideal of universities as corporate enterprises reflects NPM ideas – universities are expected to be producers of educational and research services.

In their self-evaluation reports for the university evaluations, Finnish universities emphasise their role as public agencies. The two other ideals get less attention. As public agencies they not only consider the general task of educating professionals, but also their role in regional development. The more recent role as part of the national innovation system is also taken into account, side by side with the role of autonomous cultural institution.

“Lappeenranta University of Technology is a product of the regional development policy of the 1960s. --- The scope of teaching for the university was based on national surveys of future needs of engineers in industry. The aim of regional development policy was to distribute study possibilities in different parts of Finland.” (Kyläheiko 1998, 5–6) (My emphasis)

“According to the programme of the Science and Technology Policy Council (Finland: Knowledge and Knowhow, VTN Helsinki 1996), the universities play a crucial role in the national innovation process. --- Our duty is to promote the industry and economy of Finland through research and education in close cooperation with them. --- The general mission of HUT, as that of any other top university, is to assist mankind in solving many growing global issues, to explore the unknown, i.e. to do leading edge research and to promote basic and engineering sciences and technology as intellectual assets to be shared by all people.” (Helsinki University of Technology, 1–2) (My emphasis)

The ideal of the university as a cultural institution is little emphasised. However, the description of the University of Turku as “characteristically an autonomous national institution of higher education with only limited regional goals and accountability”
better relates to a cultural institution rather than to a public agency or corporate enterprise (Puukka 2000, 1). It should be noted that the self-evaluation reports for university evaluations were written at the institutional level, where people know the trends and phrases of the time. At the department level the ideal of the university as a cultural institution is probably more alive and perhaps so self-evident that there is no need to mention it.

The corporate enterprise ideal is a product of the 1990s. In Finland, it was first evident in the discussion about management by results (Jäppinen 1989). Starting from the pilot evaluations, the role of management and a need for university strategies have been emphasised in university evaluations (e.g. Kogan et al. 1993; Davies et al. 1993; Dahllöf et al. 1998; Goddard et al. 2000). Also, the Ministry of Education has required the universities to prepare a range of strategies (for example, for internationalisation, for their regional role, student selection, guidance and tutoring, etc).

As characteristics of an entrepreneurial university Clark lists strengthened steering core, expanded developmental periphery, diversified funding base, stimulated academic heartland, and integrated entrepreneurial culture (Clark 1998, 5–8). Out of these, strengthened steering core and diversified funding base are often discussed in the university evaluations (Davies et al. 1993, Junge-Jensen et al. 1995, Three Finnish Universities…1998, Foppen et al. 1998, Junge-Jensen et al. 1998). Expanded developmental periphery is mainly considered in evaluations in which the regional role or external impact of universities is discussed (e.g. Davies et al. 1993, Dahllöf et al. 1998, Goddard et al. 2000, Goddard, Asheim et al. 2003, and Kells et al. 2000).

Universities are loosely coupled organisations with a dual structure that consists of institutional (enterprise) and discipline dimensions, and their decision-making has been described, among alternatives, as ‘organised anarchy’ (Weick 1976; Clark 1983; Birnbaum 1989; Hölttä 1995, 158). These characteristics are often referred to when complaining about the difficulty of managing higher education institutions (e.g. Kinunen et al. 1998b, 68–71). However, there are integrating elements or mechanisms as well, such as common norms and values, institutional sagas, and organisational cultures (Clark 1972 and 1983; Välimaa 1995). In addition, institutional management itself, and different internal quality assurance procedures, work as integrating elements (Hölttä 1993).

Organisation culture or other integrating elements are little discussed in the self-evaluation reports of Finnish universities. However, the statement that “TSEBA-spirit should be increased....” (Strategy 2000, 28) indicates an attempt in the direction of an integrating element. (TSEBA is the abbreviation of the name of the university, Turku School of Economics and Business Administration). At the Åbo Akademi University the “special responsibility to provide for the university-level research and education needs of the country’s Swedish language population” (Three Finnish Universities…., 64) is the kind of element to make a strong organisation culture.
In an article about leadership in universities the rector of a Finnish university considered the challenges posed to university leadership by a loosely coupled organisation and how to cope with them. As an answer, he suggested leadership at all of institutional, faculty and department levels, and “a common understanding of direction in the university, and enthusiasm to progress towards it” (Reponen 1999). “The common understanding of direction”, or vision, can be considered an integrating element. In addition, the internal performance negotiations carried out annually in Finnish universities offer an effective integrating mechanism.

When first launching university evaluations, the Finnish Ministry of Education used the concepts “self-regulation” and “capacity for change” to describe the expected direction of change (Stenqvist June 10, 1992). The point was not quality, but rather institutional capacity building. The concepts of self-regulation and capacity for change came close to those used by Bauer et al., namely a capacity for action, a space of action, and realised autonomy. The realised autonomy of a university depends on its space of and capacity for action. A university’s formal space of action depends on the authority granted to it by the State (through legislation and other regulations). A university’s capacity for action, in addition to ability, includes its knowledge and skill, and its awareness and motivation to act (Bauer et al. 1999, 35). In a situation where a university’s space of action suddenly increases, for example through deregulation or decentralisation, its capacity for action may lag behind. This is clearly indicated in the self-evaluation report of the University of Turku: “The University has not fully exploited the new opportunities”. The report also states that the university is still on its way from administration to management (Puukka 2000, 58; see also Kinnunen et al. 1998a, 95–98).

2.3 Concluding Remarks

University evaluations started at a time of major change. The devolvement of decision-making powers and the consequent strengthening of institution-level management, as well as external evaluation, came as a consequence of the new models for governance and self-regulation. In everyday life, however, both the concept of self-regulation and the relationship between state regulation and evaluation have remained unclear.

Based on the self-evaluation reports for the institutional evaluations, most Finnish universities can be described as being on the path between administration and management. As to the three ideal types, it is difficult to find the corporate enterprise ideal in the self-evaluation reports. The public agency role, in turn, is not disappearing, but rather increasing.

Although the term quality appeared in some early documents about university evaluations, it was never widely discussed or defined. Rather than quality, the topic
in the Finnish discussion was evaluation. University evaluations were not about quality improvement in the sense that there would have been doubts about the students’ learning experience. Rather they were about improvement and strengthening of institutional management and leadership in a situation of increasing institutional autonomy and responsibility. Focused at the institutional level and emphasising the universities’ responsibility for quality, evaluations had the role of a change agent.
3

Evaluation and Quality Assurance
– Political Jackknives and Tools
for Improvement and Accountability

In this chapter the focus is on evaluation. First the history of external evaluation or quality assurance of higher education, and the concept of quality from the perspective of higher education are considered. Then, the terminology related to evaluation and quality assurance is presented. The third part considers the purposes of evaluation, and the fourth, external evaluation of higher education. Finally, different approaches in general evaluation are discussed.

3.1 From Peer Review to Quality Management

Traditionally higher education institutions have various internal and external mechanisms in place to assure academic quality (e.g. Henkel 1998, 291–295). In Continental Europe universities have strong state regulation, British universities use external examiners, and US universities have an accreditation system. Traditional internal quality assurance mechanisms include student selection and assessment, and peer review for filling posts and for evaluating research. In addition, US universities utilise internal reviews arranged by universities themselves (El-Khawas and Shah 1998). When external evaluation of higher education was introduced in Europe in the 1980s, various quality-based evaluation concepts and approaches, originally developed for use in the private sector under the title Total Quality Management (TQM) along with various New Public Management ideas, were imported into higher education. (For material about TQM, see e.g. T. Kekäle 1998). Consequently, in addition to the traditional forms of quality assurance, the apparatus for evaluating higher education came to include new tools and concepts, such as quality assurance, quality audit and academic audit. (These concepts are discussed below in Chapter 3.2) Universities faced major changes: adoption of a managerial approach, external evaluation, and the new quality terminology. Barnett suggests that ‘total quality care’ would suit higher education better than ‘total quality management’. Management is needed in higher education too, but rather management for quality than management of quality (Barnett 1992,
65). In Finland the discussion was focused on evaluation rather than quality in the 1990s. Quality gained a footing in Finnish universities only with the introduction of the Bologna process.

For European higher education the quest for quality and other changes discussed above meant the appearance of external quality assessment and assurance. External evaluation did not, however, replace traditional (mainly internal) academic quality assurance procedures. Rather, the two coexist. Even though the evaluation of higher education has its own special features, it can also be considered a part of a wider development, or as suggested in the Introduction, it can be seen as being at the crossroads of different evaluation traditions. First, there is peer review, the traditional method of evaluating research (but also teaching and higher education institutions) with a history dating back centuries (Niiniluoto 1987, 11). Second, there is performance audit, which is an extension of financial audit, and shares its origins with quality audit and academic audit (Wisler 1996; Brooks 1996; Chelimsky 1996). Third, there is general evaluation or evaluation of social programs, developed in the USA from the 1960s (Sinkkonen and Kinnunen 1994, 62; Albaek 1997, 371–372). Finally, there is quality assessment, a newcomer that was imported into universities in the 1980s along with other NPM ideas (Brennan and Shah 2000, 1–8). Although the focus of this study is higher education evaluation, other types of evaluation are also considered, as they enable higher education evaluation to be seen in a wider perspective. Further, the literature on the utilisation of evaluation relates mainly to general evaluation.

The figure below gives an idea of the age of the different traditions, and shows how recent has been the arrival of the external evaluation of higher education in Europe. It also shows the sources from which higher education evaluation has borrowed various elements.

**Figure 1. The age of various evaluation approaches**

* ‘audit’ meaning checking accounts
Traditional and Modern Quality

The concepts of traditional and modern quality in higher education, introduced by Becher (1997), illustrate the change in expectations about higher education discussion above. Traditional academic quality, defined internally by the guild, has a centuries-old history. Also, the first form of external evaluation of universities, the US accreditation system that Kells calls a collaborative self-regulation system (Kells 1992, 22), was created by universities themselves, that is, by the guild. Within the guild there are shared academic and disciplinary values, mainly implicit, on which to base the quality criteria used in peer review (e.g. Kekäle and Lehikoinen 2000, 121–139). The modern quality concept assumes that external stakeholders also have a say when quality of higher education is defined and assessed. Consequently, parents, employers, local authorities, governments, and other funding bodies should be involved in defining the quality of higher education as well as students. As there is no common value base or a shared understanding about quality among various stakeholders, implicit quality criteria are no longer sufficient. The modern quality concept calls for formal, explicit criteria (Bleiklie et al. 2000, 193–197). As Vroijenstijn puts it, quality has to be negotiated (Vroijenstijn 1995, 14–15). Part of the negotiations take place at a political level – nationally, and increasingly also supranationally (Bologna Process and the Communiqués of Prague, Berlin, and Bergen, Joint Quality Initiative, and the Dublin Descriptors). Other negotiations take place at the levels of university, faculty, department, or individually, with teachers and students.

The five different types of quality recognised by Harvey and Green (1993) illustrate the modern quality concept and different stakeholder interests:

- exceptional
- perfection or consistency
- fitness for purpose
- value for money, and
- transformation.

Exceptional quality reflects a pursuit of excellence, and represents the traditional academic quality concept, especially in research. The pursuit of ‘better’ also characterises the ever-improvable quality considered by Kekäle (T. Kekäle 1998, 15). Perfection and consistency seem to refer to production or administration rather than to teaching or research. However, there are usually university-specific and national regulations concerned with higher education. In addition, as part of the Bologna process, several projects that aim to establish explicit criteria for European higher education degrees are under way. For example, the Joint Quality Initiative is focused on generic descriptors (http://www.jointquality.org/), and the Tuning Project is aimed at subject-specific descriptors (http://tuning.unideusto.org/tuningeu/). Further, in the process of accreditation, higher education institutions or programmes are considered against certain threshold
criteria, although sometimes these are implicit. As the service function (third task) and customer perspective have received more attention in higher education in recent years, universities have also come to know quality as fitness for use (often fitness for purpose), and as value for money. Fitness for purpose is the kind of quality that was evaluated in the case study universities described in this dissertation. The fifth type, quality as transformation, characterises teaching and learning. The different types of quality reflect the values and expectations of various stakeholders. Van Vught made a telling statement when he argued that besides being a political concept, quality is also a multi-dimensional and subjective concept (van Vught 1994, 36).

Combining the Newcomer with Tradition
The introduction of Total Quality Management in higher education meant combining traditional internal quality assurance of higher education with a newcomer. Among other things, this led to an extensive use of peer review in the new forms of evaluation. However, the change in evaluation approaches was also part of a major change in expectations and ideals concerning higher education. The modern quality concept, and the consequent introduction of external evaluation have not made internal evaluation unnecessary; quite the contrary. External quality assessments make use of the results of internal evaluations. Higher education institutions are expected to conduct internal evaluations, and to have internal quality assurance procedures in place (Ministry of Education 2004:6). In Finland, under the University Act, universities are obliged to evaluate their activities and to publish the results of those evaluations (Universities Act 1997, §5).

3.2 Terminology in Evaluation of Higher Education
The terminology in evaluation of higher education can be confusing; different terms are used to refer to more or less similar phenomena. The key concepts are evaluation and quality assessment or, more recently, quality assurance. Hämäläinen et al. have chosen ‘evaluation’ as the superordinate term (Hämäläinen et al. 2001a, 7). Brennan, like most higher education researchers, uses the quality-related terminology (Brennan 1999, 220; Brennan and Shah 2000, 5). In 1997 and 1999 Brennan used the term ‘quality assessment’ as a generic term, in 2000 ‘quality management’. In this study, evaluation is the main concept. However, when discussing higher education evaluation in this dissertation, the quality-related terms as defined below have been used in instances in which they are more precise. The terms that are relevant here are quality assessment, quality assurance, accreditation, and quality audit or academic audit.
Quality management has been considered here as being synonymous with quality assurance. These concepts are considered briefly below. First, however, the concept of evaluation is discussed.

**Evaluation is about Value giving**

“If you evaluate something or someone, you consider them in order to make a judgment about them, for example how good or bad they are.” This is a dictionary definition given to the word ‘evaluation’ (Collins 1995, 565). As to the concept of evaluation, evaluators and scholars have found it difficult to define (Sinkkonen and Kinnunen 1994, 21; Vakkuri and Meklin 2003). This is illustrated by Chelimsky and Shadish (1997, xii): “The now-classic answer is that evaluation is about determining merit or worth - although it is far from clear that we have consensus on that answer across the diversity of the evaluation profession.” In addition to merit and worth, significance is usually mentioned when defining evaluation (e.g. Heinonen 2001, 21; Sinkkonen and Kinnunen 1994, 25). According to Heinonen, merit, worth, and significance are internal, while judgments such as ‘important’, ‘legitimate’, and ‘motivated’ refer to external value (Heinonen 2001, 21; see also Raivola 1995, 34). Thus, the core of evaluation is value giving. Different definitions reflect the diversity mentioned by Chelimsky and Shadish, and the practical nature and context dependency of evaluation (Sinkkonen and Kinnunen 1994, 21; Harrinvirta et. al 1998, 5).

In the Finnish Ministry of Education’s ‘Strategy for the Evaluation of Education’, evaluation is defined as an interpretative analysis of the object or activity under scrutiny with a view to determining the benefit or value it produces (Strategy... 1997, 2). This means that in addition to value giving, evaluation includes interpretation. Laukkanen adds a third element to the definition, namely information gathering. According to him, evaluation means the definition of the merit and worth of a phenomenon or a subject, based on an analysis of the available information on that phenomenon. Evaluation research, in turn, is based on systematic information gathering (Laukkanen 1998, 32–34). Thus evaluation consists of information gathering, interpretative analysis, and value judgment. In evaluation of higher education there is usually one more phase, namely making recommendations.

Laukkanen points out that evaluation is retrospective by nature, which according to him is essential for understanding the concept of evaluation. Evaluation looks forward only by offering information for the planning of actions for improvement (Laukkanen 1998, 33-34; Vedung 2001, 139). There are, however, also ex ante evaluations, for example the evaluation of research proposals for funding decisions, and ex nunc or intermediate evaluations that are used to check whether an evaluation is proceeding as expected, and to take corrective measures where needed (Vakkuria and Meklin 2003). The future-oriented forms of evaluation are also based on an analysis of the past.
Criteria are Part of Evaluation

Evaluative judgment calls for criteria, either explicit or implicit, and applying criteria, in turn, includes comparison. Thus, there is also a comparative element in evaluation. According to Vuorela the essence of evaluation is to be found in the comparison of facts against criteria or objectives – in other words “what is” is compared with “what should be” (Vuorela 1997, 38). As will be seen below, the comparison of what is and what should be refers to audit rather than to evaluation.

Criteria are based on values (Sinkkonen and Kinnunen 1994, 37, 74–75). For example, an evaluation focused on economy, efficiency, and effectiveness is based on different values than and evaluation focused on equity and democracy, or another based on academic values. Brennan and Shah recognise four types of values in external evaluation of higher education, namely academic, managerial, pedagogic, and employment values. Each type leads to different evaluation criteria, and consequently to different judgments (2000, 14–15). They also relate to different stakeholders. Thus a major question in evaluation is whose values it is based on. The criteria used in an evaluation reflect the locus of power in evaluation (Barnett 1994; Haverinen 2002).

In evaluation of higher education which is based on peer review, criteria have traditionally been implicit and the process opaque (Barnett 1994, 173; Niiniluoto 1987, 12–13). Along with the European development, however, a need for transparency and more explicit criteria in evaluation of higher education has emerged. For example, the standards and guidelines for quality assurance in higher education called for in the Berlin Communiqué in 2003, were accepted by the Ministers at the Bergen meeting in 2005 (Standards…).

Vague Terms

The word ‘assessment’, which is used interchangeably with ‘evaluation’ in higher education, does not have the value-giving element of evaluation, but refers rather to determining the level of performance, or to measuring (Heinonen 2001, 22; Vuorela 1997, 38). However, the criteria used in assessment are also based on values (Sinkkonen and Kinnunen 1994, 108). In higher education, ‘quality assessment’ usually refers to actual assessment of academic quality, i.e. quality of research and teaching (Dill 1998, 70). Assessment as such is a one-off activity, but quality assessment is also used more generally to cover both actual quality assessment and the structures and processes involved in assuring quality in higher education (Brennan and Shah 2000, 5), i.e. quality assurance.

Audit, Performance audit, Quality audit, Academic audit, and Accreditation

When used without prefixes, audit is an examination of an organisation’s accounts in order to make sure that they have been prepared correctly (Collins 1995, 98). Thus, audit as such is not evaluation. However, there is performance audit – an extension of financial audit towards evaluation (Pollitt and Summa 1996; Power 1999). According
How are University Evaluations used?

To Wisler, performance audit is traditionally focused on normative questions for which the answer compares “what is” with “what should be” (Wisler 1996, 2; cf. Vuorela above). In higher education, audit is used with two alternative prefixes – quality and academic. Quality audit has been defined as a systematic and independent determination whether the planned arrangements (i.e. the quality assurance procedures) are suitable for achieving objectives, the actual quality activities conform to the planned arrangements, and the arrangements are being implemented effectively (Woodhouse 2000, 88). Academic audit implies that the object under scrutiny is an academic programme or institution.

Thus, quality audit does not evaluate quality itself, but refers to evaluation of an organisation’s quality assurance arrangements, including both quality assessment and improvement. Academic audit, as defined by Dill (1998, 70), is an externally-driven peer review of internal quality assurance, assessment, and improvement systems. As a type of evaluation, Finnish university evaluation has been considered as being close to quality audit (Davies et al. 1993, 7). The European University Association EUA has applied the term ‘institutional review’ to refer to university evaluation (Three Finnish Universities…).

The main difference between accreditation and the other types of higher education evaluation is not in the procedure, but in the outcomes: in evaluation it is usually a judgement of quality and recommendations, while the outcome of accreditation is yes or no, which can relate for example, to status or quality label. (For more about accreditation approaches, see e.g. Hämäläinen et al. 2001b, 7.)

Peer Review

Peer review is the traditional evaluation method in higher education, based on the expertise of members of the academic community, and leading to actual judgment about academic quality. It was first used in 1665 for reviewing academic articles (Nininiluoto (1987, 11). Peer review is also used when assessing applications for professorships, and for informing funding decisions. Further, especially at US universities, peer review is used to get feedback for self-improvement (El-Khawas and Shah 1998). In Finland, the University of Helsinki in particular has made use of peer review for self-improvement both at faculty and institutional levels (Evaluation of the Faculty of Science 1994; Tuomi and Pakkanen 2002; Hämäläinen E. 2000). In spite of the different backgrounds of peer review and quality management, peer review continues to be a core element in all evaluation of higher education, but in line with the modern quality concept, experts from other areas and students are increasingly used.

Evaluation and Research

The essence of evaluation is value giving, whereas research is concerned with finding the truth. While research usually (but not always) offers the basis for generalisation,
evaluation does not (Heinonen 2001, 25–27). Rather, evaluation has practical purposes. There is also a difference in the time span of research and evaluation; the time span of evaluation is usually shorter than that of research. However, the difference between the two concepts is far from clear (Sinkkonen and Kinnunen 1994, 30-34), and evaluations vary from mechanical surveys to ambitious research designs (Harrinvirta et al. 1998, 5), and many do not make a difference between the two (Vuorela 1997, 13; Pohjola 1997). The term evaluation research is also used (Vartiainen 2000a and 2000b).

3.3 Purposes of Evaluation

Evaluation is practical by nature. Consequently, it is usually expected to lead to measures of one type or another, and to have positive impacts. Ahonen (1998, 27), for example, argues that mere knowledge that certain policies or programmes have achieved their goals without negative side-effects is not enough. According to Raivola (1995, 21) evaluation aims at improvement and efficiency and includes an inbuilt requirement for change, while Chelimsky argues that evaluations for accountability and knowledge building do not call for change (Chelimsky 1997, 16–17, 21). The practical nature of evaluation is evident because of its close connection with problem solving. Both Shadish et al. and Pawson and Tilley consider evaluation as problem solving, or to be precise, as solving social problems (Shadish et al. 1991, 20–21; Pawson and Tilley 2000, 12). However, evaluation can be considered to be a management tool as well (Seppänen 1997).

The three main purposes usually attributed to evaluation are accountability, improvement, and enlightenment, as discussed above (Patton 1997, 65). Chelimsky refers to three perspectives – those of accountability, development and knowledge (Chelimsky 1997, 9–24). Accountability means that those evaluated are required to report to government or to other higher authorities and / or to tax payers about the measures they have taken as a consequence of an evaluation. Evaluations for accountability are often considered as a means of control. Improvement and development clearly include the idea of change for the better where “better” can refer to quality or efficiency (Sinkkonen and Kinnunen 1994, 53), or to provision of evaluative help to strengthen institutions (Chelimsky 1997, 10), for example. While improvement indicates change, accountability does not necessarily call for it (Chelimsky 1997, 9–24). The knowledge perspective is also called enlightenment or conceptual use (Lampinen 1992, 33–36), or cognitive (Feinstein 2002, 433). Vedung argues that accountability and improvement stand out as the most eminent rationales for evaluation. Basic knowledge is best regarded as a possible, positive side-effect of the other two (Vedung 1997, 101). In addition to improvement, accountability, and enlightenment, new purposes have also emerged. Chelimsky has described the developments as follows:
How are University Evaluations used?

“In this sense, evaluation – not only for the purposes of accountability and good management, but also for knowledge building and sharing, for institutional learning and development, for governmental and democratic reform through the serious examination of public policy – has become a precious and unique tool as we prepare to deal with the new socio-economic, political, and infrastructure needs of the next century. Perhaps it is this more widely glimpsed utility that accounts for the current expansion in many areas now being seen in the evaluation field.” (Chelimsky 1997, 6; *My emphasis*).

In a discussion on the purposes of evaluation Eggers has stated: “I believe that evaluation is learning from experience to improve future work.” He goes on arguing that (i) there should be only one evaluation purpose; and (ii) evaluations not intended for use are pointless. In her answer to Eggers, Chelimsky concludes that there are several levels of purpose. “On the top, I don’t object to putting ‘learning from experience to improve future work.’” However, she pays attention to the duality of the purpose (learning and improving), and to the possible tension between the two elements, that is, finding out (learning) and doing good (improving). The three purposes mentioned above (accountability, improvement, and enlightenment) she placed at the next level, which means that there would be a hierarchy of evaluation purposes with ‘learning from experience to improve future work’ on the top. (Eggers and Chelimsky 1999). ‘Learning from experience’ refers to the NPM ideology and to the idea of evaluation as a learning tool.

In the quotation above, Chelimsky refers to institutional learning and development as one of the new purposes of evaluation. Institutional learning can also be called organisational learning (Argyris 2001). The term used by Fetterman (2001) is ‘capacity building’. In his article on academic audits Dill concludes that the audits offer a valuable model for a ‘capacity building’ approach to academic quality assurance (Dill 2000, 204). Capacity building can be considered to have been the main purpose of Finnish university evaluations.

Evaluations can also be described as being summative or formative. Scriven emphasises that the terms refer to different roles of evaluation, not to different processes. As formative evaluation takes place during the process (project, activity; and could be a programme as well) which is being evaluated, it can be described as intermediate or ex nunc evaluation, aimed at providing feedback (Scriven 1997, 498–499). Both types include a value judgment, but formative is closer to improvement purpose, while summative has an accountability emphasis.

In addition to formal purposes, intended users may use evaluation for other purposes, and there may also be other users of an evaluation using the evaluation as they please. Thus evaluators never know for sure what sort of audiences they will have, or what kind of utilisation evaluation results they will finally have (Pollitt and Summa 1996, 31). On the other hand, an evaluation may have several purposes (Vedung 1997, 101). Feinstein argues that different purposes can be complementary so that accountability acts as an incentive for learning (Feinstein 2002, 433–434). In higher
education, the typical combination is that of improvement and accountability, but it is not considered an easy one. Chelimsky suggests that evaluations with different perspectives (purposes) may be needed at particular times or policy points, and not at others. For example, an evaluation for knowledge or one for development may need to precede an evaluation for accountability (Chelimsky 1997, 10–11).

3.4 External Evaluation of Higher Education

In the following, the practical side of the above changes and discussions is considered, starting with quality assurance agencies, their purpose and approach. Then the so-called general or European model of quality assurance is discussed.

3.4.1 The Appearance of Quality Assurance Agencies

Since 1998, there has been formal pressure on national governments in the European Union to establish quality assurance agencies. According to a recommendation of the European Council, quality assurance systems should be based on the characteristics identified in the European Pilot Project for Evaluating Quality in Higher Education (1995), that is, the four-stage model that emphasises the independence of evaluations from both the State and the higher education institutions (Council Recommendation of 24 September 1998). Later, the Bologna process, with its emphasis on external quality assurance in higher education, has put additional pressure on signatory countries to establish national quality assurance systems (especially the Prague, Berlin and Bergen Communiqués). (For more about the role of the agencies, see Brennan and Shah 2000, 26–28, and Smeby and Stensaker 1999, 9–10).

Different starting Points

While the developments leading to external evaluation of higher education were common to most OECD countries, national starting points varied. Consequently, there were different ways of reacting to the international developments and of arguing in favour of external evaluation of higher education. The different emphases are illustrated by Bauer and Henkel: “...there is perhaps a difference in the purpose of evaluation here between the contrary tendencies applying respectively in Sweden and Britain: decentralisation for better learning, centralisation for better control?” (Bauer and Henkel 1999, 242). Brennan and Shah suggest that in countries with strong university autonomy, such as the United Kingdom, the focus of external evaluation was on control, whereas in countries that traditionally had strong state control, such as Finland and other Nordic countries, external evaluation was more focused on
improvement. In the former, agencies were perceived to challenge well-established traditions of institutional autonomy. (Brennan and Shah 2000, 27).

**The Purposes of National Quality Assurance Agencies**

Based on 36 case studies from 17 different countries, Brennan and Shah (2000, 31–32) recognise ten different purposes for quality agencies (one agency can have several):

1. to ensure accountability for the use of public funds
2. to improve the quality of higher education provision
3. to inform funding decisions
4. to inform students and employers
5. to stimulate competition within and between institutions
6. to undertake a quality check on new (sometimes private) institutions
7. to assign institutional status
8. to support the transfer of authority between the state and institutions
9. to assist mobility of students
10. to make international comparisons.

University evaluations relate in the first place to the eighth purpose, ‘to support the transfer of authority between the state and institutions’. Purpose No. 1, ‘to ensure accountability for the use of public funds’, has a minor role at FINHEEC, as there are other mechanisms of accountability in Finland, such as the university data base KOTA, and the annual performance negotiations between the Ministry and higher education institutions.

An ENQA survey of 2003 considered the functions of agencies, and the objectives of evaluations, but not their purposes. Quality improvement, dissemination of knowledge and information, and accreditation are mentioned as the main functions of the agencies. According to the survey the most important objectives of evaluations are quality improvement, accountability, and transparency (Quality Procedures 2003, 12–14). Functions and objectives overlap. It is obvious that irrespective of different national approaches all agencies wish or claim to improve quality (Quality Procedures…; Harvey and Askling 2003). Enlightenment, learning, or knowledge building are seldom mentioned as purposes in evaluation of higher education. Barnett (1994) is an exception. Perhaps the situation in evaluation of higher education is similar to that in general evaluation – enlightenment is considered a positive side-effect rather than a purpose (Vedung 1997, 101).

Woodhouse (2003, 135–136) lists 18 different purposes for external audits, divided into seven subgroups, namely quality improvement, audit, assessment and standards, accreditation, qualifications, steering and funding, and information. Considering Finnish university evaluations, the purpose of special interest is ‘institutional development or capacity building’, which belongs to subgroup ‘quality improvement’.
Two Birds with one Stone?
As agencies have several purposes, they face the question of how to combine several purposes, usually improvement and accountability, in one evaluation. The question has been much discussed (e.g. Barnett 1994; Brennan and Shah 2000, 51), and it is pertinently illustrated by Vroijenstijn in “Navigating between Schylla and Charybdis” (Vroijenstijn 1995). Brennan states that whatever the purpose of an external evaluation, those on the receiving end consider it as a process of control and interference. However, the interpretation of the purpose depends on the context of the evaluation. Consequently, similar quality assurance methods can be experienced quite differently according to contextual features such as how power is exercised, and the relative importance attached to accountability, the market and trust (Brennan 1999, 222; El-Khawas 2001; Trow 1996). Patton (1997, 65) argues that even though the three purposes are not inherently conflicting, one of them tends to become the dominant motif and to prevail as the primary purpose of evaluation. Feinstein disagrees. He argues that accountability can act as an incentive for learning (Feinstein 2002, 433–434). The perspective of Brennan is that of higher education, while Patton and Feinstein consider evaluation more generally. In spite of the difficulty, practically all quality assurance agencies combine improvement and accountability, not only at the level of principles, but also in actual evaluations (Brennan 1997, 14; Williams 2001, 113–114).

In Finland the strong improvement emphasis in higher education evaluations was made possible, on the one hand, by the national database on the performance of Finnish higher education institutions and the annual performance negotiations between the Ministry of Education and universities. On the other hand, the trust between the Ministry of Education and higher education institutions discussed by Hölttä (1995), and Smeby and Stensaker (1999) made the improvement emphasis possible.

3.4.2 The Four-step Model
Higher education evaluations mainly follow the so-called European or general model of quality assessment in higher education, recommended by the Council of Europe in 1998 (Council Recommendation). Finland is no exception. The model consists of

- autonomy and / or independence of the body responsible for quality assurance as regards procedures and methods,
- self-evaluation,
- external evaluation and a site visit by a peer review team (also called an expert panel or external panel), and
- published report.

The regulatory power of the recommendation is strengthened by the fact that it is included in the membership criteria of the European Association of Quality Assurance
How are University Evaluations used?

Agencies (ENQA, www.enqa.net). The model is used in quality assessment, quality audit, and accreditation. The purpose of an evaluation has an effect in the first place on the criteria and outcomes of evaluation, not on the model itself. Hence, it can be considered to be a kind of multi-purpose instrument. Chelimsky refers to similar tool kits that are used in different evaluations (Chelimsky 1996, 63). Brennan (1997, 11), in turn, considers the model a useful framework to examine the considerable differences between national approaches rather than a real model.

Independence in Methods and Procedures

The requirement for autonomy and independence relates to the legitimacy of the agencies and of the evaluations they carry out. Legitimacy is needed to ensure that evaluations are accepted by those being evaluated (institutions or departments), by government, and by various other stakeholders (Brennan 1997, 12–13; Vartiainen 2004, 115–136). Referring to El-Khawas, Brennan (1997, 12) compares the situation with tightrope walking: One should not have a distinctive tilt towards any of the stakeholders. In addition to independence, the legitimacy of evaluation calls for expertise both in evaluation and higher education (van Vught 1994, 40). In the Nordic mutual recognition project (Lindeberg and Kristoffersen 2002), the close cooperation between FINHEEC and the institutions especially in university evaluations and in institutional evaluations of polytechnics was considered a possible threat to independency. That is, it was seen as a tilt towards higher education institutions (oral feedback to FINHEEC as recalled by Kristoffersen, Wahlén and Liuhanen. For more about the project, see Lindeberg and Kristoffersen 2002). Even though this has not been a problem in Finland, it obviously could be one if considered from a different perspective.

Self-evaluation

The second part of the four-step model is self-evaluation, which is expected to serve both the university or programme under evaluation, and the external panel. This reflects the purposes of improvement and accountability discussed above. Sallinen et al. call it dual strategy (Sallinen et al. 1994, 359). Besides description, self-evaluation is expected to include an analysis. However, high stakes in evaluation, for example accreditation, ranking, or grading, make it difficult for universities to be open about their weaknesses (Williams 2001, 114). Another reason for superficial self-evaluation may be irrelevant self-evaluation questions given from above (Saari 2002, 114; Saari 2003; Stensaker 1997, 64). The problem is how to serve two masters. For Barnett (1994, 177), the question is – who is to be enlightened?

The more comparative or grading elements there are in evaluation, the more important common evaluation questions or criteria become. They serve the institution or programme to be evaluated, the external panel, and in broader terms the transparency of the evaluation. Thus the need for common questions or framework depends on the purpose of evaluation. Accreditation, for example, calls for uniformity, whereas developmental evaluation allows for, or even calls for, a more flexible approach, as
stated above. Smeb and Stensaker consider the need for common criteria from a different perspective. They relate the degree of standardisation in evaluation methods and procedures to available quantitative databases. Those Nordic countries that have well-developed databases on higher education, have the least standardised evaluation methods. “One may speculate as to whether the design and structure of more qualitative assessment systems are not closely related to and balanced with other quality assurance mechanisms and planning tools in each country. Finland, where quality assessment is so far most pluralistic and experimental, is also the country with the best functioning and most thorough quantitative database.” (Smeby and Stensaker 1999, 11).

As the authors of self-evaluation reports at universities are experts in producing texts and in influencing others, Saari argues that instead of being self-critical, self-evaluations tend to produce rhetoric. Further, self-evaluation tends to represent a hoped-for situation (Saari 2003, 21–22). As factors that influence self-evaluation Saari lists the following:

1) ownership, e.g. who makes decisions about evaluation,
2) purpose of evaluation,
3) publicity and audience,
4) available time and other requirements such as fairness, economy, and uniqueness, and
5) the form of self-evaluation (fixed vs. free).

According to Saari the free form is more likely to produce innovations than the fixed one (Saari 2002, 75–76).

The self-evaluation process is often considered the most fruitful part of evaluation, as it offers an arena for discussions (Saarinen 1995, 232), and makes the organisation consider its own operations (Valovirta 2000, 91, 99). In other words, self-evaluation can be a learning experience. Consequently, self-evaluation, or in wider terms, the participation of those evaluated, is considered important especially in developmental evaluation (Chelimsky 1997, 10).

For disciplinary units self-evaluation can offer an opportunity to assert their authority (Brennan 1997, 16). Similarly, in university evaluation, the institution might see the evaluation as providing the possibility for it to strengthen its authority. But self-evaluation can also be a bureaucratic exercise. Brennan makes a distinction between academic, evaluative self-evaluations on the one hand, and descriptive, administrative self-evaluations on the other (Brennan 1999, 222; cf. fixed vs. free above). If evaluation questions are more relevant to the agency than to the university, the emphasis of evaluation is tilted towards accountability, which may undermine the aim of improvement and make window-dressing an attractive alternative (Williams 2001, 114).

In the evaluation of educational fields or degree programmes, the ‘self’ is usually expected to be the department that communicates directly with the external agency and with the peer review team. In university evaluations, the ‘self’ is the evaluated
university. However, as self-evaluation is usually extended to faculties, departments and to other special units, the ‘self’ consists of several layers. Karjalainen and Huusko (2004, 8) take the perspective of staff at lower levels of the organisation, which tend to feel that those from the upper echelons act as external observers. Further, questions that are appropriate at the institutional level may be irrelevant at the department level (Stensaker 1997, 59).

**External Peer Review**

External peer review is based on self-evaluation reporting and site visits, leading to a published report. As with self-evaluation, the site visit has two purposes: to verify the self-evaluation, and to acquire additional information (Kogan 1994, 353). From the legitimacy point of view, two questions are usually considered important: who are the members of the external panel, and who appoints them (Quality Procedures... 2003, 24–26). Other important questions are who should the peers meet, and how long a site visit ought to last. The success of an evaluation depends very much on the expertise and skills of the peer review team.

Traditionally, peer review is conducted by esteemed academics from the discipline to be evaluated. This is still the case in the evaluation of research, but in evaluation of teaching and higher education institutions the situation is different. In addition to academics, the external panels may include representatives from business and the community, experts in university pedagogy, alumni and students (Quality Procedures...). Student participation on the panels seems more prevalent in Sweden, Norway and Finland than in other countries (Froestad and Bakken 2004). However, there were no student members on the external panels assembled for university evaluations. The appointment of external panels is usually an agency decision, but universities or departments may be invited to propose members or they may exercise a right of veto (Quality Procedures...2003). This was the case in Finnish university evaluations.

In addition to the formal tasks of validating or legitimising self-evaluation, acquiring additional information, and making the actual judgment, peer review teams can see their role in a more developmental way (Kogan 1994, 355). This has been illustrated in the report of a university evaluation, and resembles the critical friend approach referred to by Chelimsky (1997, 18) and Fetterman (1997, 383). The alternative or additional roles suggested for peer review are

- a mirror for the university to reflect on its performance
- a revealer of issues that the university .... should pick up
- an educator – providing a source of ideas and good practice elsewhere
- an advisor – delivering positive recommendations where appropriate

*(Davies et al. 1999, 9)*
The study by Stensaker about Swedish audits indicates that the critical friend or advisor approach would have been welcomed. The audits were criticised for having too little time for discussion especially in large universities, and for focusing discussions on verifying facts rather than on improvement. Stensaker concluded that the site visit served the external panel better than it did the university (Stensaker 1997, 60). Brennan (1997, 18) points out that peers are both colleagues and competitors, and Williams (2001, 114) notes that peers can be enemies just as much as friends. When considering the independence of the peer review teams Brennan uses two examples. Peers can be hired to assist the central authority, or they can be in the driving seat (Brennan 1997, 16). Obviously, peers “in the driving seat” have more independence than those hired to help the central authority. In the Finnish university evaluations, the peer review teams occupied the driving seat.

When considering the strengths and weaknesses of peer review, its origins in the evaluation of research should be kept in mind. While there is no general agreement on criteria of ‘good science’ (Niiniluoto 1987, 12) in evaluating higher education (institutions) there usually is at least a project-specific agreement on the criteria to be applied (e.g. the criteria used in Swedish quality audits, Wahlén 2004, 140). Brennan states that in evaluation of teaching, peers are visible and they must justify what they say and do. This holds for university evaluations as well. Thus peer review of teaching and institutions is more transparent than is the case with research (Brennan 1997, 18).

The main strengths of peer review when carried out by academic peers are considered to be flexibility to adjust to disciplinary traditions and compatibility with academic values. As peers understand how universities work, contextual factors can be taken into account (Välimaa 1994, 399). According to Barnett (1994, 173), the validity of peer review is high, but the reliability low. In addition to low reliability, peer review has been criticised for a number of other reasons, namely for being expensive, slow, ineffective, biased, subject to manipulation, random, and often inappropriate (Niiniluoto 1987, 12). Further, it is said to be subjective, self-serving, and often lacking clarity in the criteria used. In addition, there is a danger of the halo effect (Välimaa 1994, 399). Recognising the criticism, van Vught states that there is no substitute for the peer review process, as no objective and widely agreed-upon criteria exist to explain such complex phenomena. Further, he emphasises the need for expertise and competence (1994, 40). Niiniluoto points out that no method of evaluating academic quality is infallible. Consequently, “university faculties, editors of journals, and administrators of funding organisations have to learn the skill of utilising the recommendations cautiously in decision-making.” (Niiniluoto 1987, 12). The advice, however, is problematic in those forms of higher education evaluation that call for compliance, such as accreditation.
Outcomes of the Process
Outcomes of the process refer to the external panel’s report and to other potential consequences of an evaluation such as accreditation or funding. Usually an evaluation report includes an evaluative analysis of the evaluated institution or programme, recommendations for change, and/or a judgment. It is the peer review team’s feedback to the evaluated university or programme and to the central authorities. Based on the report, an evaluation can have other consequences like accreditation (status), funding, or ranking (reputation), depending on its purpose.

An evaluation report can be used to provide guidelines for change, it can pass on a message to outsiders, and it can be used for legitimising current and new practices, and for maintaining the status quo of the university. Further, the advantage of a report is that one can return to it later. According to Brennan and Shah, rewards are one of the mechanisms of impact, and impact through rewards is likely to be a function of the published outcomes of an evaluation, that is, the report (Brennan and Shah 2000, 89).

According to Dill, evaluation reports seem to be a key element in the ability of academic audits to influence academic behaviour. However, there can be problems if reports are written in veiled language, and if there is an absence of specific recommendations. However, he states that the validity of perceptions and the reliability of the method are more crucial to the success of audits than the language used (Dill 2000, 199). As regards Swedish audits, department heads have found it difficult to recognise the university in the external reports, which tend to be too abstract, and for the external reports to lack relevance when considered from the department perspective (Stensaker 1997, 64).

Leeuw argues that it is reasonable to assume that no matter how well-intended evaluation activities may be, they could have unintended and undesired side effects that jeopardise performance and/or quality improvement within the evaluated or audited bodies. From the perspective of higher education evaluation, the types of side effects that deserve special attention are elimination of diversity and the enforcement of conformity, the limiting of educational outcomes, and the tendency to fund what is measured, and to value what is funded (Leeuw 2001a, 104-112). The danger of conformity has also been discussed by Sinkkonen and Kinnunen (1994), Nilsson and Wahlén (1999), Franke (2002), and Stensaker (2000). According to Sinkkonen and Kinnunen, various standards have a strong tendency towards conformity, which to some extent is in conflict with the customer requirement of individual services (1994, 118). Considering the Swedish audits, Stensaker argues that there are normative, conforming elements in them, but also forces that support the maintenance of institutional diversity, at least in the short run (2000, 315). Nilsson and Wahlén conclude that Swedish audits have contributed towards establishing a norm, which may, or may not, suit an individual university or college (1999, 6). Conformity, however, can also...
be an aim. The whole Bologna process is about conformity and diversity, and quality assurance is one of the focal areas where conformity is aimed at.

The evaluation report is often considered to be the key tool leading to utilisation of the recommendations of an evaluation and the impacts which might follow (Vedung 1997, 280–281; P. Virtanen 2001, 149–150). However, the self-evaluation process in higher education evaluation is often considered to be at least as important as the report, as it offers an opportunity for discussion already before the peer review team’s visit (Saarinen 1995, 231; PLS consult 1998, 14–15). Thus, the self-evaluation process and report may serve as an integrating mechanism in loosely coupled organisations. Similar experience has also been reported by other types of organisations (Valovirta 2000, 104). The utilisation of the evaluation process is further discussed in Chapter 4.5 (Process Use and the Use of Results) below.

### 3.5 Approaches in General Evaluation

In the history of general evaluation two waves can be identified. The first took place during the 1960s in the USA, and the second in late 1970s and early 1980s in Europe. The second wave reached the Nordic countries in the late 1980s and in the 1990s. The first phase was related to economic growth, the latter to economic depression. In the US emphasis was on effectiveness (impacts) and in the Nordic countries on efficiency (Sinkkonen and Kinnunen 1994, 62–68; Albeak 1997, 371–372). The external evaluation of higher education relates to the second wave.

**The theoretical Background of Evaluation...**

The theoretical background of evaluation can be found in systems theory. The theory of open systems assumes a self-regulating system for which information gathering and feedback are crucial (Birnbaum 1989; Hölttä 1995; Sinkkonen and Kinnunen 1994, 108). Based on feedback, a rational actor – an individual or an organisation – changes her activities. In other words, a rational actor learns from experience, and acts according to the lessons learned. However, the use of evaluation may also relate to strongly authoritarian thinking, especially if the user of the information is higher in the authority structure than the evaluated (Bleiklie et al. 2000, 111; Sinkkonen and Kinnunen 1994, 71–72). Perhaps the expectation of compliance explains why enlightenment as a purpose of evaluation and the idea of an actor learning from the evaluation are little discussed in the literature on evaluation of higher education.

**... and its methodological Roots**

The development of various evaluation approaches reflects the methodological background of evaluation in social sciences. According to Vartiainen, the basis of evaluation
research is in the paradigm of the positivist philosophy of science. According to the positivist approach, the purpose of evaluation is to produce objective and value-neutral information for decision-makers. Post-modern evaluation research, in turn, applies and develops various qualitative approaches, for example, evaluation as a learning process, process evaluation and participatory evaluation. Typical of post-modern approaches is the hermeneutic philosophy of science that emphasises understanding and interpretation. Analysis is characterised by subjectivism, where the researcher aims at a closer contact with the phenomenon under study. Methodologically post-modern evaluation approaches come close to being idiographic, which means that they focus on single happenings or processes, with the aim of understanding them rather than looking for generalisations. The results of post-modern evaluation are unique, emphasising differences between phenomena. Elasticity of research design in various phases is also typical of post-modern evaluation. Elasticity of research design means that the design can be modified during the research process in order to reach a better understanding of the phenomenon under study. (Vartiainen 2000b, 67–69).

Shadish et al. describe the above development from positivist to post-modern by stating that evaluation approaches have developed from a search for truth about effective solutions to social problems, towards understanding how organisations in the public sector operate, to producing politically and socially useful results, and finally, towards integration of the two stages (Shadish et al. 1991, 67). Pohjola characterises the development as a challenge to evaluators to move from an administrative control mechanism towards supporting the development of practices, which calls for a recognition of the political role of the researcher. This would mean detachment from the rational-technical evaluation model, and a move towards a participatory dialogue, where especially the themes of power and morale are important. (Pohjola 1997, 424; see also Henkel 1998) As an example of the new approaches, Chelimsky mentions process evaluations, which can be necessary adjuncts for accountability studies, but probably have their greatest importance and use in an institution-building, developmental perspective (Chelimsky 1997, 12).

The move from positivist to post-modern evaluation approaches raises the question of truth in evaluation. Vedung opposes objectivity and truth on the one hand, and usefulness and use on the other hand, and warns that researchers risk becoming involved in political processes. Further, in choosing between objectivity and usefulness, the latter may be preferred. According to Vedung evaluation should not be transformed into something “entirely politicised” (Vedung 1997, 284–285). Both Vedung and Shadish et al. (1991, 55) realise the need to take the users of evaluation into account, or even to involve them in evaluation. In involving users they, however, see a danger of losing the truth and objectivity of evaluation, and a consequent weakening of the academic quality of evaluation — perhaps also a weakening of the evaluator’s power. Both consider policy level programme evaluations in the first place, not small scale developmental evaluations. Objectivity and truth are important especially in evalua-
tions that serve national-level policy making and accountability (e.g. Laukkanen 1998, 47-48), while evaluations that aim at developing single organisations are expected to produce new ideas for those very organisations. In accountability evaluation the credibility of evaluation calls for truth seeking, while developmental evaluation allows for or even calls for participatory methods. The choice of method depends on the purpose of evaluation (Chelimsky 1997). Higher education evaluations aiming for both improvement and accountability are expected to fulfil the requirements of both usefulness and of objectivity and truth. (For more about truth in evaluation, see Stake 1997, and Scriven 1997)

In the following, different evaluation approaches are considered from two perspectives. The first perspective is historical, based on the development of social science methodology, where different paradigms (approaches) are often referred to as evaluation generations (Shadish et al. 1991; Pawson and Tilley 2000, Heinonen 2001). The second perspective, used mainly in performance audit, is based on the sources of evaluation criteria. Neither of the two perspectives offers a single classification, but rather several as can be seen in the lists below.

Table 1. Approaches based on the development of social science methodology

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- truth to social problem solving (Scriven, Campbell)</td>
<td>- experimental (Campbell &amp; Stanley 1963)</td>
<td>- experimental</td>
</tr>
<tr>
<td>- use and pragmatism (Weiss, Wholey, Stake, Patton)</td>
<td>- pragmatic (Weiss, Patton)</td>
<td>- pragmatic</td>
</tr>
<tr>
<td>- integrators (Cronbach, Rossi)</td>
<td>- comprehensive (Rossi, Chen)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- constructivist (Cuba &amp; Lincoln)</td>
<td>- interpretative, critical</td>
</tr>
<tr>
<td></td>
<td>- realistic (Pawson &amp; Tilley)</td>
<td>- theory-based</td>
</tr>
</tbody>
</table>
How are University Evaluations used?

There are both similarities and differences in these five lists. As to the methodology-based perspective, experimental, pragmatic, and constructivist approaches are a feature of all three lists. In addition, realistic, theory-based, comprehensive, and an approach characterised as integrators are considered. The last two lists are partly overlapping. Realistic evaluation occurs only on Pawson and Tilley’s list, who coined the term in the first place, and theory-based approaches only on Heinonen’s list. In what follows, experimental, pragmatic, constructivist, and realistic evaluation are considered first, based mainly on Ahonen 2001, Heinonen 2001, and Pawson and Tilley 2000. Then criteria-based approaches are discussed.

### 3.5.1 Methodology-based Approaches

The evaluation approaches considered here represent the development of different evaluation generations from positivist to post-modern, or in other words, from experimental to constructivist. As the lists drawn up by Shadish et al., Pawson and Tilley, and Heinonen indicate, there is no shared understanding about the names of the approaches or the generations. Rather, individual researchers interpret the development and the various approaches somewhat differently.

The experimental approach is the first phase of evaluation, the one that is referred to when speaking about the search for truth and objectivity. Based on the positivist philosophy of science, it looks for causality and assumes a distance between evaluator and the evaluated (Heinonen 2001, 34; Vartiainen 2000b, 68). Higher education evaluations are not about causality, nor do they use experimental design. However, due to the accountability purpose, there are truth-seeking elements in higher education evaluation. Also the requirement of independence in the evaluation of higher education relates to audit, and hence truth seeking. The independence requirement,
however, also relates to the acceptance of the results of evaluation (Council Recommendation).

The pragmatic approach emphasises the usability of evaluation results and takes users as the starting point for evaluation. Pragmatically acceptable knowledge is considered valid (Ahonen 2001, 106). The question about truth is considered from the point of view of usefulness. Evaluation is expected to find useful solutions for developing the evaluated organisation or activity. Hence, recommendations are made. The distance between the evaluator and the evaluated is considered case by case as are the methods. The evaluator’s professional dilemma may relate to whose perspective should be chosen, and the solution is to consider alternative decisions from several perspectives (Heinonen 2001, 34–35). Utilisation-focused evaluation, a term coined by Patton, is an example of the pragmatic evaluation approach (Patton 1997). Heinonen considers pragmatic evaluation as an instrument for improvement, and the role of evaluator as one of facilitator (Heinonen 2001, 35, 42–43). Ahonen sees pragmatic evaluation as a power game. Consequently, evaluations and evaluators have to force their way into the network of power relationships that either prevent or support change. Evaluation is supposed to ‘enlighten’. The development of policies, programmes and projects overlaps with political feasibility, cost-effectiveness and incremental perspectives (Ahonen 2001, 106).

Many of the features of pragmatic evaluation can also be recognised in the evaluation of higher education, where, as discussed above, improvement is the predominant purpose. Power is probably discussed more in higher education institutions than in agencies. Saari (2003) presents a different view when arguing that evaluation fills a power vacuum which emerged when the strong state regulation in education was relaxed. Thus, the interpretations of both Heinonen and Ahonen, who see evaluation as a tool for improvement, and evaluation as a power game, describe both higher education as well as other fields of evaluation. However, as stated above, the enlightenment purpose is not easy to find in the evaluation of higher education, and when mentioned, it usually relates to external stakeholders rather than to higher education institutions themselves. Brennan argues that higher education evaluation is about power and understanding – power at the macro level, and understanding, knowledge and change at the micro level (Brennan 1997, 21). In recent times, the Bologna process has added a supranational level to the power structure. Instead of power, discussion on the evaluation of higher education tends to focus on the acceptability and usefulness of evaluation, that is, to questions typical of pragmatic evaluation. Ahonen (2001, 106) states that as it is dependent on power relationships, pragmatic evaluation can be conservative. Considering the major role of academics in peer review and the consequent emphasis on disciplinary traditions, also peer review can be characterised as conservative (Barnett 1995, 7).
Realistic evaluation, as coined by Pawson and Tilley, is based on scientific realism. The approach looks for generative causation and regularities. Thus the importance of explanatory elements is emphasised. According to Pawson and Tilley, these have been overlooked by the experimentalists. The explanatory elements of realistic evaluation considered by Pawson and Tilley are context and mechanism, and the core of the approach is what the authors call the basic formula: mechanism + context = outcome (2000, xv). With the formula they argue that a programme as such is not enough to bring about expected outcomes, but that both a causal mechanism and a favourable context are needed. While Pawson and Tilley consider social programmes and their outcomes, this study focuses on higher education evaluation, especially on university evaluations and their utilisation. Thus, in the case of university evaluations, the basic questions of realistic evaluation should be formulated as follows: What mechanisms trigger the utilisation of university evaluation, and how does the university context enhance or hamper evaluation use? There are features in the evaluation of higher education that resemble realistic evaluation, namely the attention paid to context and the understanding that it plays a role in evaluation. In addition, there is an assumption – not always manifest – that compliance will be encouraged if there are sanctions connected to evaluation. The question Why?, typical of general evaluation, is not discussed in evaluation of higher education, except perhaps in the case of Sweden (Franke 2002, 27).

The constructivist approach is based on hermeneutic philosophy: knowledge and truth are socially constructed. The evaluator and the evaluated are understood as being interlinked, which means that it is unnecessary for there to be a distance between them (Heinonen 2001, 37–38, 43–44; Suikkanen and Pitäinen 1997). The purpose of constructivist evaluation is to understand (Heinonen 2001, 35), or to empower and educate (Ahonen 2001, 107). Hämäläinen and Kauppi use the terms ‘constructive’ and ‘communicative’ evaluation, and ‘developmental or transformative’ evaluation. Communicative means that the information and knowledge produced and created in evaluation process is taken into the discourse between the actors. Transformative evaluation means that this discourse leads to change (Hämäläinen and Kauppi 2002, 70). Due to the above-mentioned requirement of independence between evaluators and the institution or programme being evaluated, the constructivist approach is unlikely in the external evaluation of higher education. For instance, the Standards and Guidelines for Quality Assurance in the European Higher Education Area emphasise that there should be no conflict-of-interest between the members of the peer review teams and the evaluated institution or programme. However, in internal or internally initiated evaluation, where peer review is conducted for purposes of improvement, the constructivist approach need not be excluded. As to this study, the description of evaluation use that was built up is based on the perspective and experience of the interviewees. Thus it can be characterised as having constructivist elements.
An example of the constructivist approach is empowerment evaluation, where the power that usually lies with external evaluators is given to those evaluated (Fetterman 2001). Heinonen (2001, 43–44) however identifies a difference between constructivist and critical approaches, and includes empowerment in the latter group. According to Fetterman, empowerment evaluation is necessarily a collaborative activity, not an individual pursuit. An evaluator does not and cannot empower anyone; people empower themselves, often with assistance and coaching. Further, Fetterman emphasises the importance of the self-evaluation process (Fetterman 2001, 3). According to him, in practice there is an overlap between the collaborative, participatory, and empowerment approaches. Characteristic to empowerment evaluation are collaboration, participation, and commitment to self-determination and capacity building (2001, 112). Siitonen describes empowerment as an internal feeling of strength (1999, 59), and emphasises that empowerment is an individual process. Strength cannot be given to another person (1999, 14). He also recognises the meaning of empowerment as authorising when considering the task of a manager to enable and enhance the creation of new knowledge in business. This kind of empowerment Siitonen describes as giving power, allowing action and initiative, and providing resources and trust (Siitonen 1999, 83). Thus, there can also be empowering elements in more traditional evaluations (Fetterman 2001, 114). According to Pawson and Tilley, empowerment is one of the possible mechanisms that trigger the utilisation of evaluation (2000, 66).

Stakeholder evaluation can be pragmatic or constructivist. The difference lies in the role of stakeholders in evaluation. In pragmatic evaluation, stakeholders are considered to have relevant information; they are heard. In constructivist evaluation, stakeholders are involved in the actual evaluation. Thus they can directly influence both the evaluation process and the result of the evaluation. (Heinonen 2001, 37; Vartiainen 2001)

### 3.5.2 Criteria-based Approaches

An alternative perspective when considering evaluation approaches is based on the types of criteria used in evaluation. The criteria shown in the two lists above relate to standards and norms, formal goals, effectiveness, inputs and outputs, and the values and preferences of decision-makers or stakeholders (Vuorela 1997, 30-32; Sinkkonen and Kinnunen 1994, 78–79). In some approaches the criteria are preset, in others they are produced during the evaluation process. The latter overlap with the above discussed stakeholder-based evaluations.

The difference between criteria-based approaches and those based on methodology reflects two different traditions, namely evaluation and performance audit or, as Pollitt and Summa (1996, 29) call it, value-for-money audit. Power refers to it as compliance audit (Power 1999). Performance audit is defined as “an objective and
systematic examination of evidence... of the performance of a government organisation, program, activity, or function in order to provide information to improve public accountability and facilitate decision-making” (Wisler 1996, 2). Like quality audit, performance audit also had its origins in financial audit, but it has recently developed towards evaluation. The difference between the two, according to Wisler, is that in (performance) audit, compliance and goal attainment are in focus, while (program) evaluation is more interested in impacts and in descriptions (Wisler 1996, 1–2). The approaches can be characterised through their basic questions: “what” is an audit question, “why” refers to evaluation (Pollitt and Summa 1996, 31). (For a detailed discussion on differences between the two traditions, their future perspectives, and respective professions, see Wisler 1996)

As suggested above, the requirement of independence in higher education evaluation can be considered a feature of performance audit rather than one of the positivist philosophy of science. Further, the comparison of goals and achievements refers to audit. Focused on goal attainment, audit is assessment rather than evaluation. The recent development of higher education evaluation towards preset standards and criteria emphasises the strengthening of audit elements. Considering the central role of economy, efficiency, and effectiveness in the NPM ideology, the attention paid to them in higher education evaluation is surprisingly small.

Among the various evaluation approaches, where do Finnish university evaluations “belong” to? Rather than representing any one approach they seem to have borrowed elements from several different approaches, the most important of them being utilisation-focused, realistic and empowerment evaluation, and the audit tradition (Liuhanen 2005).

3.6 Opposite Trends

In addition to the heavy emphasis on improvement, there is another interesting feature to the evaluation of higher education, namely the absence of enlightenment among the formal purposes of evaluation. Learning and enlightenment are especially referred to in discussions about self-evaluation, but rather as a fortunate side-effect than as a purpose (Vedung 1997, 101). Interestingly, they do not appear on the list of formal purposes attributable to higher education evaluation (Brennan and Shah 2000, 31-32; Quality Procedures…), nor is capacity building discussed or suggested as a purpose of evaluation. Barnett, however, considers enlightenment as a purpose of evaluation. Among the multitude of different forms of quality assessment in higher education Barnett sees a dominant trend that goes 1) from collegial towards bureaucratic power, and 2) from enlightenment of the academic community (faculty, universities) towards a technical approach, in which enlightenment occurs outside the university, in an ex-
ternal agency, external panel, or government, for example (Barnett 1994, 177). At the same time power is transferred from higher education institutions to external bodies that make decisions about the focus, level, method and timing of evaluation.

Brennan considers the dimensions of power and enlightenment useful as they “invite us to consider the links between the macro level of power and the micro-level of understanding, knowledge and change” Brennan (1997, 21). However, he criticises Barnett’s model for being an idealised construct that pitches the “good guys” of academic community against the bureaucratic “bad guys” of the State. Further, Brennan points out that states also have used peer review for their own purposes (Brennan 1997, 21). Even though Barnett’s model was a simplification of the developments in higher education evaluation where self-evaluation is practically always a part of evaluation, and evaluation is mainly conducted within the academic community, it points to an interesting development. It is obvious that since 1980s, the role of external evaluation...
in higher education has grown and that it is still growing. In addition, there are now agreed standards, procedures and criteria for quality assurance in Europe (Standards and Guidelines…). In general programme evaluation an opposite trend can be seen. Evaluation is developing from positivist evaluation for national decision-makers towards participatory and empowerment approaches (Heinonen 2001, 34–44; Vartiainen 2000b, 67–69; Koskiaho 1997; Pohjola 1997).

As to external evaluation of higher education in Finland, the trend has been similar to that described by Barnett. However, the starting point may be further up than is the case in some other European countries.

As the number of actors in higher education evaluation increases (such stakeholders as panel members, the European Ministers of Education expressing their views in their Communiqués, and other interested parties), the power of universities in evaluation decreases. Like Barnett, one can ask who is to be enlightened, or who is to benefit. Is it the evaluated university, government, peer review team, or the evaluation agency? According to Clark’s model, within his ‘triangle of coordination’, power is moving from academia towards the market and government or to intergovernmental actors such as the Bologna process. – From the perspective of general evaluation, Pohjola asks who should benefit. She sees two alternatives. On the one hand, evaluation research can be loyal to official policy and produce information to support and legitimate the status quo. On the other hand, the emphasis can be developmental, and the evaluation can aim at change. According to Pohjola the two different approaches produce two different realities (Pohjola 2000, 5–6).
4

Utilisation of Evaluation – the Name of the Game

In this chapter, first the concepts of use and utilisation are considered and related to some similar concepts. Second the role of utilisation in evaluation is discussed, followed by a discussion on various types of utilisation, and different users. Third, the utilisation of the process and results of evaluation, and factors which can facilitate and obstruct the utilisation of evaluation are considered. Finally, the conceptual framework of the study, based on this and the above chapters, is presented.

4.1 The Concept of Utilisation (Use)

In addition to evaluation, two of the key terms used in this study are *use* and *utilisation*. The terms useful, usefulness and utility belong to the same family. According to the Collins Dictionary, the utility of something is its usefulness, and if you utilise something you use it. Further, utilisation is more formal than use, and utility is more formal than usefulness (Collins 1995, 1851–1852). Thus there are two pairs of synonyms, use and utilisation on the one hand, and usefulness and utility on the other hand. In the evaluation literature, utilisation and use are applied without distinction. As regards utilisation, Vedung states that its extra bulk (eight more characters) may lend it a kind of scientific credibility (Vedung 1997, 289). The word utilisation is, indeed, more common in scholarly texts than use, but the latter has also been applied recently (e.g. Feinstein 2002; Forss et al. 2002). In the following, use and utilisation are treated as synonyms, and therefore used interchangeably. This is partly due to the literature, where both terms are used, and partly to an attempt to avoid repetition.

The concept of *utilisation or use* includes an actor, an individual or an organisation that makes conscious use of something. According to Lampinen this active, purposeful element distinguishes utilisation from two other related concepts, namely dissemination and impact. Utilisation refers to acceptance and/or implementation, while impact refers to the consequences of acceptance and/or implementation with respect to the purposes for which it was accepted. For the utilisation of research, Lampinen sets two criteria. First, social science must have some clearly manifested policy implications.
Second, research must make a difference in the thoughts and actions of those who make policy (Lampinen 1992, 29). The former refers to instrumental and the latter to conceptual utilisation. Both are relevant in the utilisation of evaluation, but there are also other forms of utilisation.

**Dissemination** (transfer), one of the neighbouring concepts of utilisation, means that the information produced by research or evaluation is made available for intended users. The most important tool for the dissemination of evaluation results is usually considered to be a report (Dill 2000, 199; P. Virtanen 2005). However, being available does not mean being used. That is why evaluators and evaluation researchers have developed different strategies for facilitating use (e.g. Shadish et al. 1991, 55; Vedung 1997, 279–287). There are also cases where evaluation comes into use without, or before, the dissemination of results. That is process use. Both strategies and process use are discussed below.

**Impact** refers to the consequences of an evaluation. According to Lampinen, impact refers to consequences of acceptance or implementation with respect to the purposes for which an evaluation was accepted, i.e. to the end results (Lampinen 1992, 29). As regards evaluation, impact does not always call for use. For instance, merely knowing that an evaluation will take place, may have an influence on how people work or behave (Vedung 1997, 278). Example No. 1 in the figure below describes this. It does not necessarily mean that the evaluation or its results were accepted, nor does it mean there was implementation. It might also mean resistance. Impact can also be a consequence of evaluation and dissemination, for example of reading an evaluation report. This is described in example No. 2, where evaluation and dissemination lead directly to impacts, without conscious utilisation. Example No. 3 in the figure describes a situation, where an evaluation is conducted, evaluation report is distributed, and utilisation takes place. This is the case considered by Lampinen (1992, 29). Finally, in example No. 4, evaluation can be used without dissemination, and then lead to impacts. This refers to process use, where evaluation and utilisation take place simultaneously. The focus of the study is on examples 3 and 4, i.e. on evaluation, dissemination and utilisation.

![Figure 3. Evaluation, Dissemination, Utilisation, and Impact: Inter-relationships](image-url)
To describe the utilisation of evaluation, Feinstein introduces three formulae. According to him, the use of an evaluation depends on two factors, namely the relevance of evaluation, and the quality of dissemination \((U = R \times D)\). Relevance, in turn, is timeliness multiplied by credibility \((R = T \times C)\), while dissemination depends on the presentation of evaluation (user friendliness), and ‘means’ or ways or channels for distribution \((D = P \times M)\). Thus, if any of the four factors fails, the utilisation also fails. An alternative way to address the issue of evaluation use is to consider the incentives and capacities to use evaluation, and the supply and demand of evaluations. (Feinstein 2002, 433-439).

**Utilisation and Impact – two related Concepts**

A clear distinction between the concepts *utilisation* and *impact* was made above. However, there is no shared understanding about them among researchers. Rather, the two are given slightly different interpretations, and in some cases they overlap. While Lampinen considers impact as a consequence of utilisation, Valovirta includes *effects* as the last phase in the utilisation process (Valovirta 2000, 72, and 2002a, 64). Brennan and Shah (2000, 89), and Välimaa et al. (1998) use the concepts *direct* and *indirect impact*. As examples of direct impact, Brennan and Shah mention changes that are made as a result of recommendations given in an evaluation report, and financial or other decisions that may follow external evaluations (Brennan and Shah 2000, 90).

In the terms of this study, they represent utilisation.

The different concepts of utilisation and impact lead to different questions. While impact relates to *what and how much happens* as a consequence of an evaluation, utilisation focuses on *how*, that is, for what purposes evaluation is used. However, the utilisation of evaluation is often also considered to refer to change (Marra 2000, 22; Shadish et al. 1991, 37–39, 53, 59).

Vedung operates with the concepts of *direct* and *indirect use*. In direct use, the recipient is personally exposed to the evaluation – either by taking part in investigative work or by being exposed to findings through evaluator briefings or via published reports, whereas indirect use occurs when the transmission is mediated through some third link in the communication process between evaluators and users (Vedung 1999, 277). Frederiks et al. in turn, use the concepts *active*, *passive* and *no utilisation*. Active utilisation means taking measures on the basis of available evaluation results. Passive utilisation refers, for instance, to disseminating evaluation reports within the organisation, discussing the results in formal settings, and making recommendations for future changes. No utilisation means that evaluation results are neglected (Frederiks et al. 1994, 230).

In a study on the impacts of a university evaluation, Välimaa et al. developed a model that consists of the measures taken right after the evaluation, and of functional and structural changes that follow those measures. The changes consist of direct and indirect impacts. (Välimaa et al. 1998). In the light of the definition used in this study,
measures (that is, different decisions) represent *use*, while changes that are consequences of those decisions represent *impacts*.

Valovirta considers the utilisation of evaluation as a process rather than a set of decisions. Utilisation is a process in argumentation. This emphasises that evaluation is seldom used right away, as such. People need to become involved, they need to discuss the results, to make their own interpretations, and to argue about them. Decision-making often comes only after a lengthy process. In Valovirta’s model, effects are the last part of the utilisation process, following interpretation of results, argumentation, and decision-making (Valovirta 2002a, 64).

Brennan and Shah present a conceptual model of the impacts of quality assessment on higher education institutions. There are four different levels of impact, namely system, institutional, basic unit, and individual, and three different mechanisms through which impact occurs – rewards, policies/structures, and cultures (Brennan and Shah 2000, 9–13). The model does not describe utilisation, but the impact of quality assessment. However, due to the closeness in meaning of the two concepts, both the levels and the mechanisms considered by Brennan and Shah are relevant here as well.

![Figure 4](image_url)

**Figure 4.** The impact of quality assessment (Brennan and Shah 2000, 10).

Similar elements – context, input, throughput, output, feedback and utilisation – are also considered by Westerheijden et al. (1994, 230) and Patton. However, Patton adds a fourth element, namely the human or personal factor, which is essential in utilisation. In human factors Patton includes both the evaluators and the recipients of evaluation (Patton 1997, 47). In this study the human factors are evaluators and users. The former are considered as part of evaluation, the latter as individual users. Individual users are, however, at the same time a part of the internal context of evaluation.
For describing the process of utilising evaluation, the concepts of acceptance, dissemination and implementation seem especially applicable. However, the order of the three is not always the same. In general evaluation, dissemination often takes place before acceptance, while in the evaluation of higher education, acceptance and dissemination may happen simultaneously in the self-evaluation process, or acceptance may precede dissemination.

4.2 Utilisation
– the Primary Purpose and the Main Problem of Evaluation

Big Expectations
Utilisation is considered to be the criterion of successful evaluation, but it is also a major problem in evaluation. According to Weiss, utility is the primary purpose of evaluation, ‘the name of the game’ (Weiss 1999, 471; see also Patton 1997, 4–10, 20–22; Rossi et al. 1999, 431). Ahonen (1998, 27) and Virtanen (1998, 265) argue that an evaluation should not be conducted at all if there is no prospect that it will be useful for some audience or other. Utility is also the first of the Program Evaluation Standards (Patton 1997, 17). However, the expectations of immediate utilisation of evaluation in policy-making have not materialised to the expected extent. In Patton’s words, “Evaluations of Great Society social programs were largely ignored or politicized, and the utopian hopes for a scientific and rational society somehow failed to be realised.” (Patton 1997, 7) (See also Pawson and Tilley, 2000, 3–4; Vedung 1997, 265–267, 287 and Forss et al. 2002, 31) Rossi et al., (1999, 431–438), however, point out that evaluation use is better in small scale diagnostic or formative evaluations, and Chelimsky argues that utilisation is more important in developmental evaluations than in those for accountability or knowledge building (1997, 16). The problem seems to be at the policy-making level, where instrumental use of evaluation has not been as extensive as expected (e.g. Harrinvirta et al. 1998, 40). The need to improve the usability and utilisation of evaluations is also reflected in the development of different evaluation approaches discussed above. It is noteworthy that the Standards and Guidelines for Quality Assurance in Higher Education do not mention utility among the many requirements that they identify as being critical for both internal and external quality assurance of higher education institutions, and on quality assurance of the agencies. (Standards and guidelines…)

Is utilisation always necessary?
Chelimsky considers the straightforward expectation of utilisation of evaluation to be problematic, and argues that not all evaluation needs be justified by use. According to her, accountability- and knowledge-perspective evaluations typically seek and accom-
moderate appropriate policy or enlightenment use, but do not require it as justification (Chelimsky 1997, 16). She further claims that focusing too strongly on evaluation use may lead to the danger of avoiding the most needed and difficult studies fearing that findings will be suppressed by those for whom the *raison d'être* is to fight them. Further, expected non-use is a characteristic of some of the noblest types of evaluations: those that question accepted beliefs in a time of ideology or those that threaten powerful entrenched interests. In evaluation, the time span from results to expected use is shorter than in basic research, but Weiss points out that the results of both may turn out to be important in a longer term than originally expected. In addition, as evaluation is mainly expected to be used in decision-making or for improvements, other forms of utilisation may be ignored (Weiss 1999, 471).

**Too much Information**

When considering the utilisation of evaluation and the general challenge of knowledge, Patton identifies evaluation as one of a number of information sources. He states that our problem is keeping up with, sorting out, absorbing, and *using* information, and goes on to argue that often the central problem is getting people to apply what is already known (Patton 1997, 5). Feldman and March take Patton's last point further and argue that organisations and individuals tend to collect more information than they use or can reasonably expect to use, and yet they need more information all the time (Feldman and March 1981, as quoted by Lampinen 1992, 44). An example of excessive information production in Finnish universities is the abundant student feedback collected, but not systematically used (Liuhanen 1997, 9).

4.3 Evaluation is used for Different Purposes

When discussing the different ways of using evaluations, Lampinen (1992, 30) refers to ‘dimensions’ of use, Shadish et al. (1991, 52) to ‘kinds’ of use, and (Vedung 1997, 267) to ‘types’ of use. From the user’s perspective, the different types of evaluation use are purposes of evaluation use. Saari, for example, considers three purposes of evaluation use, namely to legitimise changes, symbolic to demonstrate rationality, and formative, which refers to effecting change (Saari 2002, 69). In general, the term “type” has been used here.

In what follows, seven different types of evaluation use considered in the literature are discussed. They are instrumental, legitimising (or persuasive), tactical, symbolic, ritual, conceptual, and interactive. Conceptual use is also referred to as enlightenment (Weiss 1999, 471, 477), or knowledge perspective (Chelimsky 1997, 13). The types are not mutually exclusive. Rather, an evaluation can be used simultaneously in different ways, or the different types can appear in a sequence (Lampinen 1992, 40; Vedung
In the following, the various types of evaluation use are discussed, and their connection to organisation theories is considered briefly.

**Instrumental Use**

Instrumental use is the “basic” type of evaluation utilisation, the kind of use that was aimed at in early program evaluations – to serve national level policy making (e.g. Shadish et al. 1991, 36; Vedung 1997, 269–272; Weiss 1999, 477). As Marra (2000, 26) puts it, instrumental means that evaluation findings are utilised as a means to goal-directed problem-solving processes – that is, used *strictu sensu*. Rossi et al. (1999, 396, 432) use the term direct utilisation to refer to instrumental use. In higher education, peer review for filling posts or for considering the merits of an article for a publication represents typical instrumental use. Instrumental use is often easy to demonstrate: A decision is made and/or some changes (development, improvement) follow (cf. “measures” in Välimaa et al. 1998). Instrumental use is also the type of evaluation use described by Finnish universities in their self-evaluation reports for follow-up evaluations (e.g. Åbo Akademi 2001, and University of Turku 2003), and in their annual reports to the Ministry of Education for performance negotiations. Utilisation of evaluation is understood as implementation of the recommendations given in evaluation reports. This is not far from what Neave refers to as the assumption of policy as a linear process (2005, 117), and Vedung as the engineering model in the utilisation of evaluations (1997, 268–269).

**Legitimising or Persuasive Use**

Both legitimising and persuasive use refer to trying to convince others about something. According to Shadish et al. (1991, 53) persuasive use means that evaluation is used to convince others of a position already taken or presenting evaluative data in policy debates. Vedung uses the terms ex-post rationalisation and power approach (1997, 275), while Lampinen (1992, 36–37) and Weiss (1999, 477) use the term political utilisation. Valovirta, in turn, identifies a difference between positive, negative and general legitimation. Positive legitimation means that positive evaluation results are used to demonstrate good performance – the ability to “run the business”, or good standing compared to others, for example, that research is at “international level”. He considers this as a form of accountability. Negative legitimation, in turn, refers to the above mentioned ex-post rationalisation. Both positive and negative legitimation have a special aim, for instance in gaining additional resources. The third version of legitimising use, general legitimation, is not targeted to any special recipient, but is aimed at improving the general legitimacy or to strengthening the external image of the organisation. (Valovirta 2000, 78–79). As examples of legitimising use, Niiniluoto, Oksanen and Valovirta mention researchers, who need an external “second opinion” to convince funding bodies and decision-makers of one’s competence in order to get funding or a post (Niiniluoto 1987, 15; Oksanen 2000, 41; Valovirta 2002b, 57).
Persuasion or legitimation need not happen in the direction of evaluation report, which in stead of speaking for change may be used for example to maintain the status quo. Due to its political nature, persuasive use has been accepted more slowly than instrumental and conceptual evaluation use.

**Enlightenment and Conceptual Use**

When describing enlightenment, Vedung – referring to Weiss – states that evaluation is used in thinking, but not in action. It leads to deeper understanding (Vedung 1997, 272–274). According to Marra, enlightenment suggests that detailed findings become generalisations that are eventually accepted as truths, and come to shape the ways people think (Marra 2000, 26). Weiss characterises enlightenment as a percolation of new information, ideas and perspectives into the arenas in which decisions are made (Weiss 1999, 471). Enlightenment is also called conceptual use (Shadish et al. 1991, 53; Lampinen 1992, 33–36; Vedung 1997, 272; Rossi et al. 1999, 396), knowledge perspective (Chelimsky 1997, 13), and demystification (Shadish et al. 1991, 53). Further, it is referred to as institutional learning (Chelimsky 1997, 6). Lampinen describes conceptual use as an interactive kind of utilisation process, which is a disorderly set of interconnections and back-and-forthness (Lampinen 1992, 34). This resembles the argumentation described by Valovirta (2002a), and the interactive use described by Vedung who, however, makes a distinction between interactive and conceptual use (1997, 272–275). As consequences of the utilisation (argumentation) process, Valovirta mentions new shared conceptions and increased awareness (2002a, 64). The experience in one of the Finnish pilot institutional evaluations was that the university became more aware of itself (Sallinen et al. 1994, 367). Oksanen (2000, 40) points out that an evaluation of an individual institution can turn out to be a learning process also for those in a wider institutional environment by clarifying the mutual expectations of partners and other interested parties. Whether enlightenment or conceptual utilisation is individual or organisational is not much discussed in evaluation literature, but the assumption of a learning organisation (Bleiklie et al. 2000, 110–116; Dill 1999; Kinnunen and Sinkkonen 1994, 71–72) obviously refers to the latter.

The problem with conceptual use is that it is more difficult to detect than instrumental use, partly due to its longer time span and the fact that people forget where the knowledge they use actually comes from (Forss et al. 2002, 31). Both Chelimsky (1997, 16–21) and Weiss (1999, 471) criticise the expectation of immediate use. The time span depends on which part of evaluation is used – process, results or both – and on the type of use. Due to the time span, enlightenment is the type of evaluation use that is most difficult to keep separate from impact.

**Symbolic and Ritual Use**

Both symbolic and ritual use represent cultural organisation theories (Albaek 1997, 387–393). In symbolic or ritual use the fact that evaluation is carried out becomes
more important than results. According to Weiss, symbolic use conveys an image or a message (Weiss 1999, 477). The examples given by Weiss and Pollit are very similar. Weiss states that policy makers want to be considered modern, up to date, and well informed (1999, 473). Pollitt in turn argues: “Thus, being seen to have set in hand or demanded evaluation would become more important than what is done with the results of those exercises. In this scenario, evaluation becomes a badge of modernity, a token of ‘good management’ “(Pollitt 1998, 219; italics in the original). Ritual use described by Segerholm adds another reason to evaluation uses where the content of evaluation is of little if any importance. “Evaluation has become so much a part of what is considered the norm that it is perceived as something you just do, without further consideration of the reasons why” (Segerholm 2001, 430). According to Uusikylä and Virtanen, symbolic and ritual use relate to a low level of commitment to evaluation. The difference between the two is that symbolic use relates to external evaluation initiative, while ritual use is connected to internal evaluation initiative (Uusikylä and Virtanen 2000 as referred by Sundqvist in Visionääri 1/2002).

Tactical Use
As described by Vedung, the rationale of tactical use is that an evaluation is set up and is under way, not that it will eventually produce substantive results. Vedung compares it with Potemkin villages – evaluation is needed to gain time or to avoid responsibility (Vedung 1997, 276). This seems to refer to political use (to avoid responsibility), but it might also refer to symbolic utilisation (to give an impression of rationality). Forss et al., in turn, relate tactical use to instrumental and interactive use, referring to strategies for sharing results (Forss et al. 2002, 31). Thus ‘tactical use of evaluation’ is understood in different ways, depending on the author.

Interactive Use
Before evaluation results are used in decision-making, they usually need to be discussed and interpreted. According to Vedung, interactive use refers to a situation where evaluation findings get mixed with other available information and come to use only after complicated negotiation processes (Vedung 1997, 274). For Lampinen, interaction is rather a characteristic of conceptual use: “Conceptual utilisation refers to an interactive kind of utilisation. The process is not a linear sequence from research to decision but a disorderly set of interconnections and back-and-forthness that defies neat diagrams.” (Lampinen 1992, 34) While Vedung considers interaction as a type of utilisation, and Lampinen combines it with conceptual use, Valovirta argues that all utilisation of evaluation is about interaction, or as he puts it, argumentation. The argumentation process may then lead to decisions, learning, and to legitimising use (2002a, 64). Thus, interaction can be considered a premise for evaluation use, or an essential part of it, rather than a separate purpose of use.
Some types of evaluation are considered more acceptable than others. For Vedung, for example, instrumental, interactive, and enlightening use are acceptable. He also understands legitimising: “After all, evaluation should count but politics ought to decide.” As to tactical use, it should not be nurtured by researchers; rather it should be unveiled (Vedung 1997, 287–288). Pollitt also expresses disapproval rather than acceptance when referring to evaluation as a badge of modernity and a token of good management (Pollitt 1998, 214–224).

Albaek (1997) criticises the international evaluation literature for being based on general impressions, occasional evidence and empirical generalisations, rather than on solid theoretical grounds. He argues that research on utilisation aims at finding such “legitimate” types of evaluation use, which serve the researcher’s understanding of his / her role as a tool of increasing “rationality”. Further, many studies on utilisation are prescriptive by nature (cf. the opinions of Vedung and Pollitt quoted above). As the theoretical base for research on utilisation, Albaek (1997) proposes the perspectives of three different organisation theories, namely rational, political and cultural. He points out that one reality, in this case the utilisation of evaluation, looks different according to the theoretical perspective of the observers (see also Scott 1987, and Bolman and Deal 1987, 16).

The defining characteristics of a rational system are goal specificity and formalisation. Goals guide decision making and action, and offer criteria for evaluation. (Scott 1987, 32-35, 268-269) Albaek argues that rational (above ‘instrumental’) utilisation of evaluation is fully compatible with the generally accepted functions of evaluation research as usually described in evaluation textbooks: evaluation is a means of improving efficiency and effectiveness. Evaluation is considered a feedback mechanism the fitness for purpose of which depends on its ability to serve as an effective tool, under given circumstances (Albaek 1997, 375–391).

The political (legitimising) and cultural (symbolic, ritual) types of utilisation have their background in institutional organisation theory, emphasising the relationship between organisations and their environments, and the role of culture in shaping organisational reality. The power of culture is illustrated in the following statement of Powell and DiMaggio: “…while institutions are certainly the result of human activity, they are not necessarily the products of conscious design” (Powell and Di Maggio 1991, 8). Powell and DiMaggio seem to argue that culture overrides rational planning.

To consider an organisation as a political system is to focus attention on uncertainty, dissension, and conflict (Birnbaum 1989, 132). Political organisations are characterised as battle fields of competing interests, and as “configurations of social groupings with basically different life-styles and political interests” – the latter referring to universities (Baldridge 1971, as quoted by Bolman and Deal 1984, 110). Thus, to put it bluntly, political use (above ‘legitimising’ and ‘persuasive’ use) means producing political ammunition. In political use of evaluation the user interprets the
available information to strengthen his / her own opinions, or he/she acquires evaluation results that support his/her own views, in e.g. difficult political decisions (Albaek 1997). Vedung uses the term “power approach”, and Valovirta refers to positive and negative legitimation (2002a).

When interpreted from the cultural perspective, the use of evaluation relates to the symbolic meaning of evaluation (above ’symbolic’ and ‘ritual’ use). Evaluation can give an impression of a serious, responsible, and rationally managed organisation, that is, evaluation helps to make a chaotic world look more organised. Evaluation becomes a rational ritual that can be used for image building. Bolman and Deal refer to evaluation as a ceremony (1984, 180). They further argue that the symbolic function is particularly critical in organisations with outputs that are ambiguous and hard to measure (Bolman and Deal 1984, 168). Universities are good examples of such organisations. Birnbaum (1989, 162) gives an example of such a rational ritual when describing the so-called garbage-can decision making, originally discussed by Cohen, March and Olsen in 1976. (See also Meyer and Rowan in Powell and DiMaggio 1991, 41–62).

Albaek has found an example of cultural use of evaluation in Danish universities, which have introduced the evaluation of teaching as a regular procedure. However, according to Albaek there is no evidence that evaluation would have an impact on decisions concerning the tasks of teachers, curricula or methods of teaching. Perhaps it is not the purpose of the evaluation. Perhaps the purpose rather is to give an impression of responsiveness to both students, who keep criticising the low quality of teaching, and the Ministry that considers it as its task to modernise the Danish education system with rational and productive initiatives such as evaluation. Universities evaluate because all others do so (Albaek 1997).

The possibility of combining several purposes in one evaluation was considered above, but the combining of several types of evaluation use, that is, simultaneous or successive use, has not yet been discussed here. The usual assumption seems to be that the purpose of an evaluation conditions its use, which is often, but not always the case. Marra (2000, 23), for example, states that instrumental use and enlightenment are complementary rather than mutually exclusive, and Feinstein argues that accountability as a purpose can create an incentive framework for learning (Feinstein 2002, 433–434).

The fact that different users might use an evaluation for different purposes at various times, irrespective of the original purpose, means that evaluation use refers not only to measures taken as a direct consequence of evaluation, but also – perhaps mainly – to a mixture of processes in argumentation as suggested by Valovirta (2002a).

Conclusion
Most of the types of utilisation discussed above can be anchored in one of the organisation theories. To put it simply, instrumental use represents rational organisation theory, legitimising use relates to political, and symbolic and ritual utilisation relate
to cultural organisation theories, while conceptual or enlightening use can best be connected to theories on organisational learning or learning organisation. As regards interactive utilisation, rather than being a type of its own, interaction relates to almost all utilisation of evaluation. Only individual, enlightening use does not necessarily call for interaction.

Enlightening and conceptual evaluation use and theories on organisational learning and learning organisation are discussed in Chapter 4.7 below.

### 4.4 Various Users

In utilisation users are essential and most evaluations have several users. These could be individuals, organisations or political decision-making systems (Vedung 1997, 277–278). One or more of them could be intended users, who might be external or internal in relation to the evaluated activity or organisation. Intended users may or may not use the evaluation, and people and organisations that are not intended users might use it (Pollitt and Summa 1996). Thus, the evaluation process and results can be used by a number of users for various purposes at different times.

**Internal or external?**

As various users of evaluative information have different needs to be met from the information produced, the position of the user in relation to evaluated activity needs to be considered (Albaek 1997, 377). There is, for instance, a difference between evaluation of one’s own organisation, and that of a subordinate organisation or activity. In evaluations for accountability, users are external to the evaluated activity, while in developmental evaluations they are mainly internal. In general evaluation, the intended users are often national-level policy makers, and thus external to the evaluated activity. In the external evaluation of higher education there usually are at least two intended users—the evaluated university or a part of a university, and the Government. The university is expected to act based on the outcomes of an evaluation, that is, to implement and improve (Vartiainen 2004, 63), and the Government to take the evaluation into account, which may or may not call for change. In addition, evaluation can be expected to inform other stakeholders such as employers, potential students and their parents (Brennan and Shah 2000, 32; Quality Procedures...2003).

**Traditional and Modern Models**

In external evaluation of higher education, there are traditionally two actors: the evaluator and the evaluated university or department. The latter is both the commissioner and the user of an evaluation, and peers conduct the evaluation (El-Khawas and Shah 1998). Today, a four-actor model is common (Figure 5). As an agent of the government,
a quality assurance agency commissions peer review teams or expert panels to carry out evaluations of universities, programmes, or certain themes. In their report the external panels address their analysis and recommendations to the evaluated university, (which is considered to be the number one intended user), to the government, (the number two intended user), and possibly to the agency as well. The university is accountable to the government, which has regulatory power over the university (the arrow from government to university).

![Figure 5. Actors in external evaluation of higher education](image)

In addition to universities and governments, agencies and external panels are also possible users of an evaluation. The utilisation of evaluation by members of external panels has generated little discussion, but it is obvious that membership of an external panel offers a view to the evaluated activity or organisation and thus a learning opportunity (Forss et al. 2002). Further, an agency can use legitimate evaluations to strengthen its own legitimacy (Brennan and Shah 2000, 15-18; Vartiainen, 2004).

### 4.5 Process Use and the Utilisation of Results

**Two Possibilities**

There are two parts in each evaluation that can be used – process and results (outcomes). The usual assumption is that the results count, but the utilisation of an evaluation might start well before results are available, and it may extend from an early planning phase through findings to implementation and enlightenment. The concern about non-utilisation of evaluation mainly relates to the use of results. The most important means of disseminating evaluation results is through the report. Hence there is ample literature on how to write a report, and especially on how to make the message ac-
ceptable and to ensure it is utilised (Vedung 1997, 279–287, Dill 2000, 198–201; P. Virtanen 2001, 149–150). Other means for improving utilisation have also been suggested. In addition to diffusion-focused strategy, Vedung considers three other strategies to facilitate evaluation use (Vedung 1997, 282–287). Shadish et al. (1991, 55) mention indentifying users early and having frequent contact with them, for example. To facilitate conceptual use, they suggest that fundamental assumptions be challenged. Like the development of evaluation approaches, the idea of the strategies for facilitating the use of evaluation can also be understood as an attempt to cut short the distance between evaluation and its use. The strategies are discussed in Chapter 4.6 on obstacles and facilitators of evaluation use.

**Process Use is a recent Discovery**

Compared with the utilisation of evaluation results, the use of the evaluation process is a recent innovation (Patton 1998, Vedung 2001). Patton (1997, 20) argues that what happens from the very beginning of an evaluation will determine its eventual impact long before a final report is produced, and Vedung (1997, 278) points out that users can learn from evaluation process as well as from evaluation findings. Originally, the utilisation-focused evaluation launched by Patton was not meant for enhancing process use, but the aim was to enhance the use of evaluation results by having the commissioning bodies, that is, the expected users, participate in the planning of an evaluation (Patton 1998, 225). The starting point of the production-focused strategy suggested by Vedung is similar (1997, 282). Patton defines process use as follows: “Process use refers to and is indicated by individual changes in thinking and behavior, and program or organisational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process.” (1998, 225) According to him, evidence of process use is represented by the following kind of statement after an evaluation: “The impact on our program came not just from the findings but from going through the thinking process that the evaluation required” (Patton 1997, 90). Forss et al. define process use as the utility to stakeholders of being involved in the planning and implementation of an evaluation. They identify five different types of process use, namely learning to learn, developing networks, creating shared understanding, strengthening the project, and boosting morale, and state that those who learn most are the evaluators (Forss et al. 2002).

**Differences between the Use of Report and Process Use**

Both research and experience suggest that the evaluation process is often considered at least as important and sometimes as more important than actual results. Having studied the first evaluations of Finnish universities and disciplines, Saarinen states that the evaluation process can create a legitimate arena for communication and knowledge transfer (1995, 232). In evaluations of research, the internal evaluation process preceding external evaluation was considered the more productive and useful
part of the evaluation process (Oksanen 2000, 40). The importance of the evaluation process and self-evaluation has also been reported from the institutional evaluations of Finnish government agencies, and from evaluation of research fields (Valovirta 2000, 91, 104; 2002b, 49). The self-evaluation process is considered valuable especially for learning and innovation (Patton 1998; Sallinen et al 1994, 372). Valovirta argues that evaluations were useful irrespective of what kind of conclusions or recommendations the evaluators finally offered. The mere process of analysis led people to ask questions, and in the interview they had to explain why things are the way they are. It provided new perspectives and helped one to see strengths and weaknesses in one’s own work. (Valovirta 2000, 99; see also Valovirta 2002b, 57–59). Chelimsky argues that evaluations for accountability and enlightenment do not necessarily call for use. According to her, use is a reasonable criterion for developmental evaluation, because it is a part of the evaluation process itself. Although such use may be difficult to measure when it cannot be represented as a discrete event and when the users are a part of the evaluation, it is nonetheless integrally related to the purpose of developmental evaluations. (Chelimsky (1997, 17–18). Patton agrees and emphasises enlightenment in process use; results are used more instrumentally and persuasively (Patton 1998, 225–226). Similar results are reported from the evaluations of Finnish research fields (Valovirta 2002b).

4.6 Obstacles to and Facilitators of Evaluation Use

With utilisation as a criterion, why do some evaluations succeed and others fail? This question has been considered from the perspective of evaluation, its context, and users.

4.6.1 Factors related to Evaluation

Here ‘evaluation’ refers to the different evaluation approaches discussed above, and to the four-step model used in higher education evaluation. Thus, in addition to actual evaluation (value giving) it includes both FINHEEC and the peer review teams, and their reports. Further, it includes the conduct of the evaluation, and the various strategies to facilitate evaluation use. As FINHEEC and the Finnish university evaluations were discussed above, the emphasis here is on the various strategies. In evaluation literature why is mainly considered from the perspective of facilitators for utilisation, i.e. different strategies for enhancing utilisation (e.g. Shadish et al. 1991, 55; Vedung 1997, 280–287).
Feinstein argues that the utilisation of an evaluation depends on its relevance and dissemination. This is presented in the formula: $U = R \times D$, i.e. utilisation = relevance x dissemination (2002, 434–436). Vedung suggests four strategies to enhance evaluation use: 1) diffusion-centred strategy, 2) production-focused strategy, 3) user-oriented strategy, and 4) meta-evaluation. The strategies cover both instrumental and conceptual use, and basic assumptions concerning non-use (Vedung 1997, 280–287).

The **diffusion-centred strategy** relates to a situation where evaluation and utilisation are separate phases. Special attention is paid to dissemination of the results by broadcasting evaluation output as widely as is feasible. In addition, permanent intermediary agents between evaluators and practitioners, for example an advisory commission, can be used. The bottleneck in utilisation is that recipients must know about the results before they can use them, and the basic assumption that faulty utilisation is due to communication barriers between producers and consumers of evaluation. Consequently, the quality of the report is important (Vedung 1997, 280–282). Virtanen (1998, 265-266), and Dill (2000, 199) also stress the importance of the report. Vedung, however, points out that effective diffusion is not enough, as information barriers are not the most difficult obstacles to utilisation (1997, 282). The evaluation report and the diffusion-centred strategy are especially important in cases where the report is the only means of disseminating evaluation results, that is, when intended users are external to the evaluated organisation or activity, and/or have not been involved in the evaluation process.

The assumption of the **production-focused strategy** is that evaluations are dismissed or shelved because they are irrelevant if not faulty (Vedung 1997, 282–285). The strategy suggests that evaluation outputs should be made more user-friendly through efforts directed at the evaluation process. The process as such should meet the demands of the conceivable recipients. The advantages of the strategy are that evaluation will be responsive to the issues that real users want to have illuminated, and that learning may occur during the evaluation process. The major problem that Vedung sees in the production-focused strategy is that researchers become involved in political processes, and that the problems addressed may be of minor interest from a research perspective. Usefulness rather than objectivity may be preferred. “From the point of view of academic research, it is important to separate objectivity and truth from usefulness and use. Invalid knowledge may be used, and valid knowledge may remain unused. Evaluation should not be transformed into something entirely politised.” (Vedung 1997, 284–285) The production-focused strategy has obvious similarities with Patton’s utilisation-focused evaluation (1997), and with the user orientation suggested by Laukkanen (1998, 157). The strategy means that the distance between evaluators and the evaluated activity or organisation becomes shorter than in more traditional approaches.

The FINHEEC strategy in university evaluations can be considered to be production-focused. As discussed above, each university had the right to decide on the theme and focus of the evaluation, as long as the perspective was institutional. Universities
could also choose the timing within certain limits. Further, they were free to conduct
the self-evaluation as they considered fit. The idea of a user-friendly approach refers to
the relevance and acceptance of evaluation discussed above. Relevance and acceptance,
in turn, reflect the importance of legitimacy in higher education evaluation.

The third of Vedung’s strategies, the user-oriented strategy, means making potential
evaluation clients more susceptible to evaluation. The critical factor is considered to be
the user’s capacity to receive and apply information. Thus, the recipient organisation
may need a champion for evaluation, or users should be educated in how to use evalu-
ation. An alternative is to incorporate evaluation into the organisation’s management
system (Vedung 1997, 285).

The fourth strategy is meta-evaluation, of which there are three different versions:
evaluation of another evaluation, summarising and synthesising of several evaluations,
and evaluation of the general evaluation function. Vedung (1997, 286–287) recommends
that evaluators evaluate earlier evaluations (see also Uusikylä and Virtanen 2000). FINHEEC uses this kind of meta-evaluation in the introduction of new evaluators.
polytechnics is an example of the summarising type, and quality audits represent
evaluation of the general evaluation function (e.g. Woodhouse 2000; Kekäle et al.

When discussing factors that support the utilisation of evaluation, Shadish et al.
mention mainly factors that facilitate instrumental use, stating that most of them also
facilitate conceptual use, even when the user cannot act on the results. As reasons for
non-use they note that evaluations can threaten entrenched interests; policymakers and
managers use information in many ways; and local personnel may not adopt evaluation
results even if they were highly innovative. Their list of helpful activities to facilitate
instrumental use is a mixture of Vedung’s diffusion and production-focused strate-
gies. Shadish et al. further argue that conceptual use can be facilitated by challenging
fundamental assumptions about problems and policies (Shadish et al 1991, 54–55).

Involvement helps...

Having used both diffusion-centred and production-focused strategies Marra states
that the best ways to encourage the use of evaluation findings have been to involve
the programme staff in defining the study and helping to interpret results, and to
produce regular reports for the programme staff whilst the study is in progress. She
further argues that this kind of sustained interactivity transforms one-way reporting
into mutual learning. (Marra 2000, 22–34). Due to self-evaluation, staff involvement
is an integral part of higher education evaluation. However, the extent of participation
varies substantially.
... perhaps sanctions as well

External evaluation of higher education aims at serving both formative and summative purposes, as discussed above, but the understanding about how to effect change (improvement) varies from country to country. Finland has relied on developmental evaluation, where no sanctions are applied. Alternatives include accreditation (status received or not; or green, yellow, or red card), funding link, and league tables. In those cases the fear of losing status or reputation makes universities compliant, but simultaneously there is a danger of symbolic use or decoupling evaluation from normal activities (Leeuw 2001a; Williams 2001; Stensaker 2000, 308).

4.6.2 Factors related to Context and Users

Compared with the extensive discussion there has been on evaluation method, the importance of the context of evaluation use has only recently come into focus. According to Cook, Weiss contextualised evaluation when considering it as a part of a larger public policy framework. Cook points out that the decision-orientated models of evaluation neglect the fact that information is rarely used in government as the sole or major input into decisions, and that information provided by social scientists is always contested (Cook 1997, 46). The context dependency of evaluation is also considered by Albaek (1997, 377), Karjalainen and Huusko (2004), Sinkkonen and Kinnunen (1994, 21), and Pawson and Tilley (2000). Another expectation, or rather assumption, is that evaluation results will actually be used by organisations (cf. above “the university became more aware of itself”). Patton, however, emphasises the role of individual users, and considers the personal factor in detail. He argues that evaluators should not target organisations, but specific individuals, as “organisations do not consume information; people do”. Further, evaluators should focus on decision-makers rather than on decisions (Patton 1997, 39–61). The above considerations mainly focus on the utilisation of evaluation results. Process use, which is often enlightening and more difficult to see than instrumental use, usually receives less attention. However, it has a more dominant role in various forms of participatory and empowerment evaluation (Fetterman 2001, 3, 13; Forss et al 2002).

In higher education evaluation, the border line between evaluation and its context is often vague. This is mainly due to self-evaluation, which is part of the four-step model. Conducted by staff within higher education institutions, self-evaluation is strongly influenced by the internal context. Thus, the internal context not only influences utilisation of the evaluation, but also the evaluation process itself. When considering the user-focused strategy, Vedung states that utilisation depends on both users and structural features of the utilising organisation (1997, 285). Here the utilisation of evaluation is related both to organisational characteristics, external context, and individual users.
In the above mentioned meta-evaluation of the evaluations of Finnish government agencies, the most important internal factors that enhanced evaluation use were developmental atmosphere and expectations about evaluation (Valovirta 2000, 86). A good deal earlier, in a major study of some 200 self-evaluation processes in US higher education institutions, Kells and Kirkwood identified three factors that relate to perceived successful self-evaluations. These were the attitude of the chief executive or other institutional leaders, the motivation of institutional actors, and the organisational context (Kells and Kirkwood 1979 as quoted by Westerheijden et al. 1994, 229; see also Kells and Stenqvist 1994, 14–16). In line with Kells and Kirkwood, Hämäläinen and Kantola conclude their synthesis of different evaluations of Finnish polytechnics by stating that there is a direct link from self-evaluation to utilisation. If self-evaluation is carried out carefully, if senior leadership is committed to evaluation, and if the self-evaluation is participatory, the recommendations of the external evaluation panel will be taken seriously and utilised (Hämäläinen and Kantola 2002, 335). Hämäläinen and Kantola focus on instrumental use, that is, how the recommendations of the external panel are followed, but pay no attention to other types of evaluation use, or to individual users. Patton, in turn, emphasises the personal factor and suggests that evaluations should be specifically user oriented – aimed at the interests and information needs of specific, identifiable people, not vague, passive audiences (Patton 1997, 382). The champion considered by Vedung also emphasises the importance of individual users (1997, 285).

As discussed above, higher education evaluation is usually expected to lead to improvement, that is, to change. According to Fullan, there are four factors that have a positive influence on implementing change. These are shared institutional values, organisational structures, (for example decision-making and communication), resources (for example time, people, and materials), and leadership (Fullan 1994, 100). The factors relate to both the internal context and to individuals.

External context is not a common theme in the literature on general evaluation. Utilisation is considered mainly from the perspectives of evaluation approach and internal context and / or users. In the evaluation of higher education the situation is different, and the role of external context is increasing in importance. This relates first to the relationship between higher education and the State, and to the legitimacy of the external quality assurance agencies and hence to that of evaluation. From the perspective of the traditional quality concept discussed by Becher (1997), the legitimacy of external evaluation is not self-evident as it affects the distribution of power (Brennan 1997; Brennan and Shah 2000, 15–18; Vartiainen 2004, 62). Second, international evaluations call for a better understanding of various national contexts (Kells 1999; Crozier et al. 2005; Omar et al. 2007). As Trow states, “... apparently similar methods of quality assessment can be experienced quite differently according to contextual features of how power is exercised and the relative importance attached to accountability, markets and trust.” (Trow 1994 as quoted by Brennan 1999, 221)
With mutual understanding, trust may act as an incentive for listening to the evaluator, and not only for listening to him or her because one is obliged to do so (Leeuw 2001b, 141).

In conclusion, the features of the internal context that are considered as enhancing evaluation use relate to structures and resources, shared values and developmental atmosphere, to expectations and motivation concerning the evaluation, and to leadership. In addition, individual actors have an important role in the utilisation. The role of the external context relates in the first place to national evaluation approaches. In international evaluations it plays an important role.

### 4.6.3 Interest, Information, Ideology, and Institution

The factors presented above as facilitators for, or obstacles to, utilising evaluation can be considered from the perspectives of information, interest, ideology, and institution, known as the Four I’s of school reform (Weiss 1995). Out of the four, the first three – interest, information and ideology – relate to individual decision-makers, surrounded by the fourth, institution and institutional rules and culture, that is, the internal context. There is a continuous interplay between the Four I’s, and every decision is a product of that interplay. **Interests** are defined primarily in terms of self-interest. Most political scientists take for granted that self-interest is the core of politics, including organisational politics. **Ideology** encompasses philosophy, principles, values, and political orientation, and includes a relatively coherent value predisposition. Weiss defines **information** as the range of knowledge and ideas that help people make sense of the current state of affairs, why things happen as they do, and which new initiatives will help or hinder. She emphasises that “information” or “knowledge” does not necessarily connote accuracy or rightness, but it can also mean partial, biased, or invalid understandings that people rely on when making decisions. **Institution** involves the structure, culture, standard operating procedures, and decision rules of the organisation within which decisions are made. It is the context in which evaluations are used, and it has an influence on the three individual I’s (Weiss 1995, 571–592). Thus ‘institution’ corresponds with the ‘internal context’ of the conceptual framework of this study, and overlaps with Fullan’s term ‘organisational structure’ (Fullan 1994, 100). Also, Vedung refers to the dependency of evaluation use on both users and structural features of the utilising organisation (1997, 285). The influence or interaction can be seen for example in that internal context and users are usually considered together. It is difficult to keep the two separate.

**The Four I’s Related to Facilitators for and Obstacles to Evaluation Use**

In Table 3, the factors that are considered by various authors to enhance the utilisation of evaluation are presented in relation to the Four I’s.
Table 3. Factors enhancing the Utilization of Evaluation in Relation to the four I’s of Weiss

<table>
<thead>
<tr>
<th>Weiss</th>
<th>Information</th>
<th>Ideology</th>
<th>Interest</th>
<th>Institution</th>
<th>Aim</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other authors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bauer &amp; al.</td>
<td></td>
<td></td>
<td></td>
<td>capacity for action</td>
<td>transformation</td>
</tr>
<tr>
<td>Brennan &amp; Shah</td>
<td>values</td>
<td>rewards</td>
<td>institutional context</td>
<td>impact</td>
<td></td>
</tr>
<tr>
<td>Feinstein</td>
<td>dissemination</td>
<td>relevance, incentives</td>
<td>capacity to use evaluation</td>
<td>utilisation</td>
<td></td>
</tr>
<tr>
<td>Hämäläinen &amp; Kantola</td>
<td>participatory self-evaluation</td>
<td>leadership commitment</td>
<td></td>
<td>utilisation of recommendations</td>
<td></td>
</tr>
<tr>
<td>Kells &amp; Kirkwood</td>
<td></td>
<td>the attitude of chief exec, motiv. of the inst. actors</td>
<td>organisational context</td>
<td>successful self-evaluation</td>
<td></td>
</tr>
<tr>
<td>Patton</td>
<td>process participation</td>
<td>utilisation-focused evaluation, esp. personal factor</td>
<td></td>
<td>utilisation</td>
<td></td>
</tr>
<tr>
<td>Shadish &amp; al.</td>
<td></td>
<td>entrenched interests</td>
<td></td>
<td>utilisation</td>
<td></td>
</tr>
<tr>
<td>Valovirta</td>
<td></td>
<td>expectations</td>
<td>atmosphere</td>
<td>utilisation</td>
<td></td>
</tr>
<tr>
<td>Vedung</td>
<td>user-focused strategy, esp. user’s capacity</td>
<td>production-focused strategy</td>
<td></td>
<td>utilisation</td>
<td></td>
</tr>
<tr>
<td>Vartiainen</td>
<td>legitimacy</td>
<td></td>
<td></td>
<td>acceptance</td>
<td></td>
</tr>
</tbody>
</table>

When placing the factors considered by other researchers under Weiss’ Four I’s, it is not self-evident where each of them belongs. Several could also be placed in a different column. This refers to the interplay between the Four I’s. Ideology or values play a role especially when acceptance of evaluation is considered. However, the emphasis on them is heavier in external evaluation of higher education than in general evaluation. This relates to the autonomy of academic organisations, and to concepts that
belong to business rather than to higher education terminology (Becher and Kogan 1992, 18; Brennan 1999, 224–229). Interest in turn relates both to evaluation, dissemination, and utilisation. Interest can be a reason to accept or reject an evaluation or to participate or not in the self-evaluation, but can it also be created or raised in the process of dissemination or self-evaluation. As interests are primarily interpreted in terms of self-interest (Weiss 1995, 574), the importance of individuals in utilisation, referred to as personal factor or champion, becomes obvious. Interest is also the underlying assumption in different strategies created for enhancing utilisation (e.g. production-focused strategy, Vedung 1997, 282–285; Shadish et al. 1991, 55). In the evaluation of higher education, incentives or rewards also play a role. They may relate to funding, status and a positive or negative judgement in evaluation. Information, in turn, can refer to informing people in advance about evaluation or to dissemination of evaluation results, but it can also be connected to earlier knowledge and experience. Thus it is not far from capacity or resources that relate to institution.

4.6.4. Acceptance, Dissemination and Utilisation of Evaluation

Feinstein (2002) argues that the utilisation of an evaluation depends on the relevance of the evaluation and the quality of dissemination. The relevance of an evaluation relates to its acceptance.

Lampinen (1992, 29–31) considers acceptance to be an element of utilisation. Here, however, acceptance is considered to be a pre-condition for utilisation. Acceptance and dissemination are presented in an order that is typical for evaluation of higher education. It means a sequence from acceptance through dissemination to utilisation. The critical role of acceptance relates to the autonomy of universities. External evaluation, commissioned by higher authorities does not belong to the tradition of higher education. However, depending on how self-evaluation is conducted, acceptance and dissemination can also be simultaneous processes. In major program evaluations, acceptance usually follows the dissemination of evaluation results. Not all the researchers referred to in Table 4 consider the acceptance or implementation of evaluation, but rather situations that are similar to or can be compared with the utilisation of evaluation. Kekäle (1998) considers the implementation of quality management, and Weiss (1995) individual decision-making in a situation of changing decision-making structures, while Fullan (1994) and Bauer et al. (1999) discuss the implementation of change. Although the utilisation of evaluation does not necessarily refer to a visible change, the expectation of some kind of a change is so usual that the situations considered by Fullan, and Bauer et al., can also be considered relevant in the context of evaluation. The table below illustrates the conditions for acceptance, dissemination and utilisation considered in the above discussed literature.
Table 4. Acceptance, dissemination, and utilisation of evaluation as considered by the different authors presented above

<table>
<thead>
<tr>
<th>Author</th>
<th>Acceptance</th>
<th>Dissemination</th>
<th>Utilisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brennan</td>
<td>legitimacy, external context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brennan &amp; Shah</td>
<td>values</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kekälä</td>
<td>match</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leeuw</td>
<td>trust</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shadish</td>
<td>entrenched interests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vartiainen</td>
<td>legitimacy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weiss</td>
<td>Information, ideology, interest, institution</td>
<td>information, ideology, interest, institution</td>
<td>information, ideology, interest, institution</td>
</tr>
<tr>
<td>Dill</td>
<td>report</td>
<td>change in academic behaviour</td>
<td></td>
</tr>
<tr>
<td>Feinstein</td>
<td>relevance, incentives</td>
<td>presentation x means</td>
<td>capacity to use evaluation</td>
</tr>
<tr>
<td>Patton</td>
<td>utilisation-focused evaluation</td>
<td>report</td>
<td>utilisation-focused evaluation, personal factor</td>
</tr>
<tr>
<td>Vedung</td>
<td>production-focused strategy</td>
<td>diffusion-centered strategy</td>
<td>user-focused strategy, user’s capacity, champion</td>
</tr>
<tr>
<td>Virtanen</td>
<td>report</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bauer &amp; al</td>
<td></td>
<td></td>
<td>capacity for action</td>
</tr>
<tr>
<td>Fullan</td>
<td></td>
<td></td>
<td>resources, values, organisation, management</td>
</tr>
<tr>
<td>Hämäläinen &amp; Kantola</td>
<td></td>
<td>participatory self-evaluation process</td>
<td>leadership commitment</td>
</tr>
<tr>
<td>Kells &amp; Kirkwood</td>
<td></td>
<td></td>
<td>attitude of the chief executive, motivation of institutional actors</td>
</tr>
</tbody>
</table>
Out of the above authors, Feinstein, Patton, and Vedung consider all the three – acceptance, dissemination and utilisation, while the others focus on one or two of them. However, Weiss’ Four I’s are relevant to each of the three phases. In the evaluation of higher education, part of the dissemination already takes place during the self-evaluation process, even though the actual results are available in the report of the peer review team. The diffusion-centred strategy described by Vedung clearly focuses on dissemination. In general, the evaluation report is considered to be the main means of dissemination. Other means of dissemination are little discussed. It is worth noting that even though Dill considers the higher education context, he does not refer to the role of self-evaluation in the utilisation process. In higher education, utilisation is often considered from the perspective of recommendations given, which refers to compliance audit. Except for Patton (1997, 44–61) and Vedung (1997, 285), reference in the literature to the role of individual users in the utilisation of evaluation remains limited.

Acceptance is about there being sufficient compatibility between an evaluation, its context, and users. Considering evaluation in general, Feinstein (2002, 434–435) uses the term relevance, while Kekälä (1998), whose perspective is that of quality management, refers to match – a term also used in the quotation below, concerning one of the pilot university evaluations. The external contextual context, discussed by Leeuw (2001b) and Brennan and Shah (2000, 19–32), is not generally emphasised in the literature. However, it often has an influence on what sort of evaluations are conducted and on the legitimacy of evaluation. Thus, it plays a background role (Trow 1996, 235–239). Further, it is gaining importance, especially in the international evaluation of higher education.

In dissemination the question is how to reach and convince intended users. The usual assumption is that dissemination begins when evaluation results are available. However, dissemination will often have already started in the self-evaluation process, well before the external panel has made its site visit and made their report available. This does not make the dissemination of evaluation results unnecessary. Rather it means that the self-evaluation phase can be considered the first phase of dissemination, preparing ground for the panel’s recommendations, and their implementation. In addition, process participation can be seen as information (knowledge) gathering, creation and sharing. This is a major reason for emphasising the role of self-evaluation. Even though self-evaluation and dissemination are conceptually two different phases in evaluation – the former takes place before actual judgement and the latter comes after it – they are often combined or considered side by side.

Once accepted and successfully disseminated, the utilisation of an evaluation relates mainly to capacity and resources, in other words, an organisation’s capacity to use evaluation. Aaltonen and Koivunen (1998, 13) use the expression ‘evaluation maturity and ability to use evaluations’, which is gained in evaluations. Thus, acceptance and/or dissemination alone do not guarantee utilisation, which calls for special
skills. The role of leadership is considered important (Hämäläinen and Kantola 2002; Fullan 1994), while the role of individuals in general is less discussed, which does not mean that it is unimportant.

Considering the above, three conditions for the utilisation of evaluation can be seen. These are

1) the characteristics of evaluation as seen by the recipient,
2) the quality and availability of information (knowledge / experience) before, during, and after the evaluation, and
3) the capacity of the recipient organisation to act based on the evaluation.

These three perspectives receive different levels of emphasis in the literature on higher education evaluation and in that on program evaluation. While higher education evaluation emphasises acceptance, especially the legitimacy of evaluation, general evaluation pays more attention to dissemination. In the evaluation of higher education, external evaluation can be considered to be interference, a threat to autonomy, while in various policy programmes the higher authority usually has a self-evident right to expect implementation.

The following quotation points to the importance of acceptance. “For a university to benefit from an institutional evaluation, it is important to see the positive potential in the differences: evaluation can benefit every unit and person, if, and only if, the evaluation process and discourse match the traditions and readiness of a unit or person and meet their needs. This is one of the most important challenges of institutional evaluation when it is extended in other universities.” (Sallinen et al. 1994, 372; My emphasis)

4.7 Organisational Learning and Learning Organisation

Enlightenment (knowledge perspective, conceptual use) is one of the three purposes of evaluation in the general evaluation literature as discussed above. Evaluation is characterised as learning from experience (Eggers and Chelimsky 1999), and as a learning process (Vartiainen, 2000b). It is also referred to as institutional learning (Chelimsky 1997, 6), and as capacity building (Fetterman 2001). Woodhouse (2003, 135–136) includes institutional capacity building among his 18 different purposes of external audits, and Patton (1997, 105) argues that self-evaluation serves both learning and innovation. Further, according to Bleiklie et al. (2000), the idea that target groups can learn from experience is part and parcel of this ideology (i.e. New Public Management). Thus, there are several terms that include the idea that learning can be considered to be one of the purposes of evaluation. Further, the terms used by Chelimsky and Fetterman – ‘institutional learning’ and ‘capacity building’ – refer to
How are University Evaluations used?

organisational learning. The point is that there seems to be a connection between the utilisation of evaluation and learning. In other words, learning can be considered to be a purpose of evaluation. In evaluation of higher education learning is not much discussed. Brennan and Barnett are exceptions. Brennan’s argument is that the evaluation of higher education is about power and understanding – power at the macro level, and understanding, knowledge and change at the micro level (1994, 177). The three – understanding, knowledge and change – suggest that as a consequence of an evaluation universities can learn. While Brennan discusses power and understanding, Barnett sees a trend in quality evaluation from self-enlightenment of higher education institutions towards enlightenment of external agencies (1994). Thus, enlightenment also plays a role in the evaluation of higher education, but the question who is (to be) enlightened – higher education institutions, external reviewers or agencies, or perhaps individuals within higher education institutions – is little discussed.

In spite of all the learning that takes place within universities, they are usually not considered examples of learning organisations (Dill 1999, 127–154). Rather, they have characteristics that may obstruct organisational learning and adaptation such as dual structure, loose coupling, and disciplinary traditions. However, as a reaction to new expectations in their environment universities make structural changes, reconsider their policies and strategies, launch new degree programmes, and establish new departments. Does it mean that they learn as organisations? Dill states that the perspective on academic accountability that is influencing the development of quality assurance policies in a number of countries appears to assume that universities – with regard to their core processes of teaching and learning – can become learning organisations. But what makes a learning organisation? Garvin (1993, 80) offers the following definition “as a first step”: A learning organisation is an organisation skilled at creating, acquiring, and transferring knowledge, and at modifying its behaviour to reflect new knowledge and insights."

Dill (1999) and Garvin (1993) consider the concept of the learning organisation, but there is also the approach called organisational learning. Research on organisational learning studies the phenomenon of learning within organisational contexts, while studies on the learning organisation attempt to define and create an ideal type of organisation in which learning is maximised. The history of research on organisational learning is older, and its approach is more theoretical than the literature on the learning organisation, which has an applied orientation (Argyris 2001, Dill 1999, Strandli Portfelt 2006).

As there seems to be a connection between the utilisation of evaluation on the one hand and learning on the other hand, there is also reason to discuss that link in more detail. In the following first the concept of organisational learning is discussed, and then that of the learning organisation.
What is Organisational Learning?
The process use of evaluation was defined earlier in Patton’s words as follows: “Process use refers to and is indicated by individual changes in thinking and behaviour, and program or organisational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process.” (1998, 225; my emphasis) Among the different types of process use Forss et al. (2002), in turn, recognise learning to learn, developing networks, and creating shared understanding. The learning referred to is not only individual, but obviously also organisational, in other words “shared”.

Argyris discusses whether ‘entities defined at relatively high levels of social aggregation’ are sufficient for the study of organisational adaptation and learning, or whether attention should also be paid to how individuals enter into these processes. He argues that a theory of organisational learning must take account of the interplay between the actions and interactions of individuals and the actions and interactions of higher-level organisational entities such as departments, divisions, or groups of managers. He also considers the concept of inquiry, described as “the intertwining of thought and action carried out by individuals in interaction with one another on behalf of the organisation to which they belong in ways that change the organisation’s theories of action and become embedded in organisational artefacts such as maps, memories, and programmes”. According to Argyris it is possible for an individual to think and act on behalf of organisation. Further, when doing so, individuals can undertake learning processes (organisational inquiry) that can yield learning outcomes that are reflected in organisational changes. From the perspective of universities as loosely coupled organisations, the importance of interplay and inquiry is both interesting and challenging. However, organisational learning can also be seen as serving stability rather than change, which means that organisations learn to preserve their status quo (Argyris 2001, 9–11).

How about the Learning Organisation?
There are several definitions of learning organisation in addition to that of Garvin quoted above. Mulford (2000), for example, defines learning organisations as organisations that structure, restructure and develop themselves in such a way that the organisation as well as its organisational members continually learn from their experiences, from one another as well as from the environment. The learning results in effective problem-solving and organisational improvement. Mulford’s definition emphasises the role of structures in the learning organisation. Dill, in turn, argues that Garvin’s definition of a learning organisation (skilled at creating, acquiring, and transferring knowledge, and at modifying its behaviour to reflect new knowledge and insights) requires two essential conditions to be met:
1) for organisational learning to take place new ideas are essential, and
2) these ideas must be a trigger for organisational improvement – that is, the new ideas must lead to accompanying changes in the way the organisation’s work is accomplished.

Thus, while Mulford emphasises structures, Garvin pays attention to an organisation’s capacity to change.

From the perspective of the two conditions, the recommendations for improvement given by the mainly international peer review teams of the Finnish university evaluations can be seen as a source of new ideas, while the re-evaluations (or follow-up evaluations), conducted three to five years after the first evaluation, offered an opportunity to check whether and how the universities had implemented the recommended changes, and whether actual behavioural change had taken place. Numerous recommendations were made to the 20 Finnish universities in the reports of the peer review teams, mainly concerning management and leadership. Considering the institutional focus of the evaluations this is not surprising. However, there were few recommendations concerning information and analysis, which might have been valuable from the perspective of organisational learning. (Liuhonen 2001, 15)

According to Garvin (1993), organisations that are skilled at translating new knowledge into innovative ways of behaving actively manage six learning activities to ensure that they occur by design rather than by chance. Institutional evaluation can be considered to represent point 4, “experimenting with new approaches”:

1) systematic problem solving
2) learning from one’s own experience
3) learning from others’ experience
4) experimentation with new approaches
5) transferring knowledge, and
6) measuring learning.

Strandli Portfelt (2006, 30) considers learning organisations from the perspectives of open systems, learning culture, and supporting learning structures. She has identified five characteristics of a learning organisation:

1) an open system, with the ability to learn from and influence the environment, as well as learning from within itself
2) an organisation where vision emerges from individuals and where the individuals’ vision is linked to the organisational vision, but where organisational vision and missions change as a result of learning
3) an organisation where individual learning becomes collective by dialogue, requiring an encouraging and accepting organisational culture and supportive organisational structures
4) an organisation where there is a well balanced mix of mastery and development-oriented learning which includes a culture that promotes and encourages learning, with flexible structures that have the capacity to change (in order to support future learning), and

5) an organisation where people use system thinking to identify and solve problems effectively.

From the point of view of university evaluations, the most interesting of the above five features is the third, according to which individual learning can become collective by dialogue, provided that there is an encouraging and accepting organisational culture and supportive organisational structures.

The definitions for the learning organisation mainly include similar elements – measuring learning, a culture of evidence, learning from experience, from each other and from the environment, problem solving, adaptation, restructuring and improvement, and a culture that allows questioning and dialogue that support the creation of knowledge and development.

The characteristics of a learning organisation presented above are considered typical of learning organisations in general. Universities, however, have special features that should be taken into account when considering them as learning organisations. In his article on the Architecture of an Academic Learning Organisation (1999) Dill focuses on universities. Based on the reports of 12 universities from eight different countries for an OECD / IMHE (Institutional Management in Higher Education) project Dill recognises five elements of the “architecture” of an academic learning organisation (Dill 1999, 148–151). They are

1) a culture of evidence
2) improved coordination of teaching units
3) learning from others, systematic surveying of programme graduates
4) university-wide coordination of “learning”
5) transferring knowledge.

Culture of evidence is close to Garvin’s “measuring learning” and “systematic problem solving”. According to Dill it is ironic that the most critical factor in transforming the university into an academic learning organisation may be instilling a culture of evidence into academic problem solving. Quoting Sir Eric Ashby (1993), Dill states that the irony lies in that scholars who in their academic work demand proper evidence for making decisions, make decisions on matters such as admissions policy, staff-student ratios, the content of courses and similar issues, based on dubious assumptions, hunches and scrappy data.

As examples of improved coordination of teaching units Dill mentions the appointment of curriculum coordinators, and the creation of academic staff committees to
coordinate the quality of teaching and learning within academic units. This means that teaching would no longer be considered to be as private as it used to be, but rather the quality of teaching would be a concern of academic units as well as to university management.

*Learning from others* means that universities actively seek out knowledge from others, be they other universities or programmes, or their own graduates and their employers. Benchmarking, external reviews, and the internal audits applied in some universities are examples of learning from others. - Dill describes *university-wide coordination of “learning”* as the pan-university level of structures for providing more effective coordination, support and accountability for the systematic improvement of teaching and learning.

Out of the five elements of an academic learning organisation among the 12 universities in Dill’s sample, the least in evidence is processes or structures that encourage the *transfer of knowledge* for improving core processes. Dill states that this weakness is not restricted to academic institutions, but their traditional decentralised structure exacerbates the problem. Further, managerial action should be paid to identifying best practices and moving them around university academic units. – In recent years various national and international benchmarking activities have gained ground in European higher education, but their influence on learning and actual behavioural change are far from clear.

In addition to the five criteria suggested by Dill, two other points made by Garvin and Strandli-Portfelt, respectively, seem important when considering the utilisation of evaluations in universities. They are the requirement of *actual behavioural change*, and *dialogue through which individual learning becomes collective*. Learning, however, is not much discussed when speaking about university evaluations. Rather, the aim is actual behavioural change or *improvement*, to use the usual term in higher education evaluation. Dialogue, in turn, is often connected to self-evaluation, but my understanding is that there is as much need for dialogue both before and after an evaluation. This is because dialogue is considered to be the tool through which individual learning becomes collective (Strandli-Portfelt 2006, 30).

The Bologna process and especially the Berlin and Bergen Communiqués of 2003 and 2005 respectively, requiring that universities establish policies and procedures for quality assurance, have given European universities a push towards a culture of evidence and of becoming learning organisations.

As regards university evaluations and their utilisation in the two case study universities, the perspectives of organisational learning and learning organisations raise two questions:

1) did university evaluation lead to organisational learning in the case study universities, and

2) did any pre-existing elements of a learning organisation support learning and the utilisation of evaluation?
4.8 Conceptual Framework and the Research Questions

The conceptual framework of this study is an adaptation of the model that Brennan and Shah (2001, 10) used to describe the impact of quality assessment on higher education institutions, consisting of method (evaluation), context, and impact discussed above. The framework used in this study consists of five elements:

- evaluation
- users
- utilisation of evaluation
- internal context
- external context.

Some of these elements differ from those formulated by Brennan and Shah. The elements ‘method’ and ‘impact’ used by Brennen and Shah have been replaced with ‘evaluation’ and ‘utilisation’ in the framework for this study. There is also an additional element, namely ‘users’, as there can be no utilisation without a user. Further, ‘utilisation’ is considered to include organisational learning as a type or purpose of evaluation, comparable with enlightening and conceptual use and knowledge perspective. Context, in turn, is divided between the external and the internal context, and ‘learning organisation’ has been considered to be a feature of the internal context.

As to organisational learning, the above discussion indicates that the change expected as a consequence of evaluation often relates to learning, even though the very word “learning” was not used. The usual terms are enlightenment, conceptual use, and knowledge perspective – each including the idea of learning, but without clarifying whether the learning that possibly takes place is individual or organisational.

There is a sequence from evaluation through its users to utilisation, including organisational learning, also called institutional capacity building and institutional learning. As context-dependent the first three elements of the framework – evaluation, users, and utilisation – are placed inside the internal and external context in Figure 6.

Evaluation and utilisation are the key concepts of this study. In the framework they are presented as overlapping to indicate that part of the utilisation of evaluation takes place during the evaluation process itself. As to the other questions, “who” precedes the other five that specify the overarching question “how”. Further, utilisation takes place within the internal and external context.

The concepts of utilisation and impact are similar, as discussed above. The difference lies in the user, an individual or an organisation that makes conscious use of something. The user can be either an organisational unit or an individual. Utilisation is a process, while impact refers to the result of the process. Impact can take place without use. Further, an evaluation can have both internal and external users. Here the
main focus is on internal users, that is, individuals and organisational units in the case study universities. This is because the universities were the principal intended users of the evaluations – they were expected to respond to the findings of the evaluations. The other intended user was the Ministry of Education, considered below.

The external context of Finnish university evaluations consists of the national and international developments at the time of the institutional evaluations discussed above (Context). As to internal context, there are 20 universities in Finland, and despite of any common characteristics, there are 20 different internal contexts for university evaluations. The empirical part of the dissertation focuses on two case study universities.

The utilisation of university evaluations is considered from the following perspectives

- who are the users of university evaluations?
- how are the evaluations used?
- what is used – process, reports or both?
- where are the university evaluations used?
- why are they used / not used?

**Figure 6. Conceptual Framework of the Study**
The above questions specify the research question presented above – how Finnish university evaluations are used. Thus ‘how’ both refers to different types of evaluation use, which from the user’s perspective are purposes of use, and covers all the five sub-questions as the title of the study indicates. ‘Who’ refers both to the intended users, the universities and the Ministry of Education, and more specifically to various users within the universities. ‘What’ refers to evaluation process and to reports, i.e. the two parts of evaluation that can be used. ‘Where’ relates to the primary and support processes of universities: teaching, research and the so-called third task, i.e. external engagement, and management and administration. Finally, ‘why’ covers factors that could facilitate or hamper evaluation use. Thus, ‘why’ relates to all elements of the conceptual framework except for utilisation, i.e. evaluation (method), users, and internal and external context. There is an interaction between evaluation and its users on the one hand, and users and utilisation on the other hand. To emphasise their contextual dependency in the framework evaluation, users and utilisation have been placed within the internal and external context. The perspective of organisational learning can best be related to the question “how” in the meaning “for what purpose?”, while the perspective of the learning organisation relates to internal context.
This chapter introduces the core of the study, the description of the utilisation of institutional evaluation at two Finnish universities. First, impressions I drew from the interviews are discussed. Second, the case study universities are described, and the ways in which the evaluations were used in each of them are discussed. Third, features common to the utilisation of evaluation at both case study universities are considered.

5.1 The Conduct of the Study and Impressions Drawn from the Interviews

The selection of the interviewees was guided by the idea that they should know about the evaluation and its utilisation. This means

- those involved in the planning of the evaluation, and responsible for the implementation of both the self-evaluation and the recommendations of the peer review teams,
- those involved in coordinating the self-evaluation, and writing the report, and
- those whose position suggests that they should have a role in the utilisation of evaluation.

Interviewees included rectors and vice rectors, directors of administration, other institutional level administrators, deans, faculty administrators and key persons from continuing education centres. The department level of the universities was not directly represented. However, among the interviewees there were persons who had moved from positions in departments to positions at faculty or institutional levels after the evaluation. They provided valuable information about the experience from those at the department level. In addition, deans are not only faculty-level actors, but also actors at department and institutional levels.
Almost all those of whom I requested an interview agreed. Those who did not, felt that their knowledge of the evaluation was too sparse for there to be any point in interviewing them. The attitude of the interviewees was positive both concerning the evaluation itself and my study. They generously found time to be interviewed. Many of the interviewees had read the evaluation reports, and one of them had prepared a list of the measures taken at their university after the evaluation. There was one exception from the generally positive attitude. One dean was critical, but his criticism was aimed at the university rather than at my study.

In all 40 interviews were carried out, including three pilot interviews – one with a rector, another with an administrator, and the third with a Ministry of Education officer. Most of the interviews were conducted between September 2002 and June 2003, but the members of the peer review teams and the rectors of three other universities were interviewed between October and December 2003. Most of the interviews took place in the interviewee’s office. However, the members of the peer review teams, one of the rectors and a dean were interviewed in the FINHEEC meeting room. The interviews lasted for 60-90 minutes.

In the table below the numbers of different interviewees and the abbreviations used when quoting the interviewees are presented.

<table>
<thead>
<tr>
<th>Interviewees</th>
<th>Code</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectors</td>
<td>R</td>
<td>10</td>
</tr>
<tr>
<td>Deans</td>
<td>D</td>
<td>9</td>
</tr>
<tr>
<td>Institutional administrators</td>
<td>IA</td>
<td>7</td>
</tr>
<tr>
<td>Faculty administrators</td>
<td>FA</td>
<td>7</td>
</tr>
<tr>
<td>Continuing education centre staff</td>
<td>CE</td>
<td>2</td>
</tr>
<tr>
<td>Ministry officers</td>
<td>M</td>
<td>3</td>
</tr>
<tr>
<td>Members of the peer review teams</td>
<td>P</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>40</strong></td>
</tr>
</tbody>
</table>

The pilot interviews pointed to a need to clarify the concept of utilisation in order to keep it separate from that of impact. As a consequence, at the beginning of the interviews I asked how the interviewee had been involved in the evaluation and what the term utilisation brought to mind. The themes, and a list of more specific questions that served as a check list, are presented in Appendix 1. The interviews did not necessarily follow the order of the themes presented in the Appendix, nor did I always touch upon all the questions on the check list. If an interviewee seemed to offer new
perspectives on evaluation utilisation, I preferred to follow his / her line of discussion rather than stick to the generalised themes. In general the interviewees were quite eager to discuss their experiences, and in some cases to explain why things went the way they did. When interviewing Rectors in particular, I felt like conducting a follow-up interview.

When analysing the interviews, the five research questions who, what, how, where, and why served as a framework. The analysis proceeded through listening to the tape recordings and reading the word processed interviews several times over, marking different types and areas of use, and parts of evaluations that had been used, and drafting individual “user profiles” to find different types of users. Further, I looked for signs of ownership of the evaluation such as how the interviewees spoke about the evaluation – whether it was ‘their own’ or something imposed on them; whether they reacted with enthusiasm and determination or indifference. When finding signs of ownership, I tried to find out whether the ownership was shared or individual.

The Evaluation was taken seriously
Interviewees from both universities told me that the evaluation had been taken seriously, and that its recommendations had been utilised in many ways. When describing attitudes towards the evaluation, the word ‘seriously’ was often used (e.g. R2, R3, R4, R5, R6, R7, IA2, IA4, D1, FA4, FA6). Mostly the interviewees spoke broadly rather than personally, and used expressions such as “we took it seriously”, “it was taken seriously”, or “they took it seriously”. By ‘they’, an institutional administrator was referring to staff in faculties and departments, surprised by the high quality of their self-evaluations (IA3). Only one rector referred to the question in a personal sense, by saying “I took it seriously”.

Some spoke of the evaluation with a sense of ownership and pride (R2, R3, R7, D1, D2, IA5), and others in a-matter-of-fact style (e.g. R5, R6, IA2, FA3, CE2). In particular, rectors and senior administrators were concerned about how the university would perform in the evaluation; they wanted their university to demonstrate good performance, with the hope that the evaluation would have the effect of legitimising their institution (R3, R5, R7, IA5; cf. Vartiainen 2004, 136).

What does the Utilisation of Evaluation bring to mind?
Responses to this question varied substantially. Some described the measures taken at their university as a consequence of evaluation (R1, IA4), others described discussions concerning the concept of regional impact (R3, IA5). There were also those who thought about the numerous recommendations, and of their university’s response to them (IA4), and those whose focus was the main users – who or where they were (R4, FA3, FA4). Further, image building and increased cooperation with local authorities were mentioned (R4, IA6, D5, FA4). Utilisation was also thought of as the use of evaluation reports – as a source of information about the university (D2) to
be used internally, but also by colleagues from other universities (R6, R7, IA3). Due to the variation in perceptions and understandings about evaluation use, the picture was not consistent across either of the case study universities. Rather, people within one university can and do see the use and benefits of the evaluation differently. For example, the different perceptions related to the usefulness of the evaluation and of the evaluation reports, to the relative importance of each of the measures taken based on the evaluation, and to the types or purposes of evaluation use – what might be an improvement for one, might be a public relations activity for another. A coherent description is possible only from a distance. Two examples from one of the case study universities illustrate the situation. A dean: “An excellent learning opportunity. The self-evaluation report is the first comprehensive report ever about the university as an organisational whole; it can be used to increase organisational self-understanding.” (D2). A faculty administrator: “It was conducted under stringent control by the central administration; there was no room for faculty-specific issues.” (FA4)

5.2 Two Universities as Users of University Evaluation

The following description is based on interviews with rectors and deans, and institutional, faculty and continuing education administrators. The perspective of departments was not directly represented, nor were the views of individual teachers and administrators sought. This shortcoming is to some extent made up for by comments received from the deans and faculty administrators interviewed, about the use or absence of use of evaluation in the departments. Deans operate at three levels within their universities. As members of the University Council they represent the institutional level, as deans the faculty level, and as professors the department level. In the interviews they mainly considered the evaluation from the faculty perspective, that is, they spoke as deans. Those interviewees who had moved from one position to another within the university after the evaluation had valuable information about, and experience at, the department level. Among those who had changed position were a former department and faculty administrator, now an institutional-level administrator; a former department head, now a vice rector; two former rank-and-file professors, now deans; two former vice deans, now deans; and finally a faculty administrator who had moved from one faculty to another. In addition, a former vice rector had become a City Mayor after a period as rector. These persons were able to consider the utilisation of evaluation from several perspectives. It is obvious that a description based on the perspective of departments would differ in detail from the one given here. However, “there is not a right or wrong description of any university, but only different variations, partial pictures” (K. Virtanen 1999, 6).
5.2.1 The University of Kuopio

**Internal Context**

The University of Kuopio was established in 1966 as a result of regional higher education policy, primarily to educate physicians for the region. Today the university has a clear profile in health and in environmental science. The University’s self-evaluation report states that in addition to its public service role, the university has taken on the task of serving economic life in the private and public sectors. (Kinnunen et al. 1998b, 3) In a comparative analysis by the Association of European Universities, the University of Kuopio represented a typical peripheral university / context. (Kinnunen et al. 1998b, 39) In the context of the three different university ideals - autonomous cultural institution, public agency and corporate enterprise – presented by Bleiklie et al. (2000, 46–51), the University of Kuopio can be considered to be both a public agency and a cultural institution. Its history refers to the former, its strong research orientation to the latter.

At the time of the evaluation, it had a student population of about 4400, and a staff of 720. At the time of the follow-up evaluation the figures were nearly 5800 and 1000 respectively (KOTA). There were four faculties at the university: Medicine, Pharmacy, Natural and Environmental Sciences, and Social Sciences. After the evaluation, a Faculty of Information Technology and Business Administration was established. At the time of its establishment, however, the structure of the University was less traditional:

1. “... there were no Faculties here, but big departments. Resources were not allocated faculty-wise, nor were teachers. They were a shared resource. This university was started with a new idea, and faculties were not introduced until 1989. For quite some time we managed without faculties, which to my mind led to a situation that people at the University of Kuopio have never been as narrow-minded as in other universities.” (R3)

According to the self-evaluation report, there was a strong tradition of involvement in interdisciplinary projects at the University, which had been helpful in implementing internal reforms (Kinnunen et al. 1998b, 69). Prior to the university evaluation, the University had already conducted both university-wide and smaller-scale evaluation projects. (A list of those evaluations is available in the Finnish language version of the self-evaluation report, Kinnunen et al. 1998b).

2. "Before the evaluation we had already established a mechanism for evaluating research, … In doing that, we also built the basis for further evaluations. … Something else that we had done was to establish a framework for allocating basic resources. We had a working group

---

How are University Evaluations used? – 103
that quickly reviewed the cost structures, for example the cost of one study point (credit) in different faculties. We also estimated how much teaching was given across the faculty borders. In this co-operation we learned about evaluation as well … At the same time we also started the quality project of the University.” (R3)

The self-evaluation report pointed to a need to strengthen academic leadership and management (“the professionalisation of the Rectorate”), and to offer training courses in management. It also mentioned a lack of time for holding joint discussions and for the exchange of information. Further, it referred to the mismatch between the needs of the region and the academic profile of the University. (Kinnunen et al.1998b; abstract, 68–72).

**Evaluation**

The evaluation was a joint project of three Eastern Finland universities. It focused on the regional role of the universities and was conducted in 1998. The universities and the Ministry of Education had agreed about it in 1996 in negotiations on outcomes for the years 1998-2000. In spite of the fact that the project was a joint effort, each university carried out its own self-evaluation, and prepared a self-evaluation report of its own. The objectives of the three-university project and those of the University of Kuopio were as follows:

*The joint project*: “The objective of the project is to analyse the role and status of each of the three universities from the local viewpoint, in relation to the regional development in Eastern Finland, and in a broader national and international context.”

*University of Kuopio*: “The objective of the evaluation of the regional role was to determine the direct and indirect effects the University of Kuopio has had and currently has on the local, regional, national and international levels.” (Kinnunen et al. 1998b, 9–11).

Towards the end of 2002, just before the interviews for this study took place, a follow-up evaluation of the 1998 evaluation was carried out. At the time of the interviews, the University was waiting for the report of the peer review team.

As to university management, several changes had taken place between 1998 and 2002. A new rector had been elected, then re-elected for a second term, and one of the vice rectors had been elected right after the evaluation. The present director of administration was appointed to that position soon after the evaluation, having previously been head of the Continuing Education Centre, in which position she had been deeply involved especially in the self-evaluation. All the deans were new, but some of them had earlier been vice deans.

It was not clear to the interviewees where the idea of a joint project focused on the regional role had originally come from. Some of them thought the idea had come from the Ministry, others suggested the Rector. There was no criticism of the theme.
There was a joint steering group for the three-university project, in which the universities’ senior management and other key stakeholders were represented. As the FINHEEC representative, I was a member of the joint steering group, but not of the university-specific steering groups. The main responsibility for the management and reporting of the self-evaluation project at the University of Kuopio lay with two persons – an associate professor who was an expert in evaluation, and later professor and dean, and the head of the Centre for Training and Development, who was responsible for the University Quality Project, and the adaptation of quality award criteria to university context. In addition, the Self-Evaluation Report lists some 20 other people, half of them academics, who were involved in collecting and analysing the evaluation material. One of them was the rector of the University, described as “a true specialist in service policy and, in particular, a promoter of the idea of universities’ regional responsibility in Finland and Europe” (Kinnunen et al. 1998b, 11–12).

The self-evaluation consisted of two parts. On the one hand, the regional impact of the University was analysed, based on surveys and documents. On the other, an internal analysis based on quality award criteria was carried out by the faculties. In the faculties, the responsibility rested with deans and faculty administrators. It was the first stage in launching the University Quality Project to faculties and departments. (Kinnunen et al. 1998b, 9-12). The Head of the Continuing Education Centre with another staff member from the Centre attended a meeting in each faculty to present the quality award framework and to provide instructions for self-evaluation in the departments (CE1, e-mail correspondence). Almost all departments conducted a self-evaluation, and reported on it. The self-evaluations served both the departments themselves and those preparing the University’s self-evaluation for the peer review team.

An international peer review team was appointed by FINHEEC after consulting the three universities. As it was a joint project, there was a three member core team that visited all three universities. In addition, three other members were appointed, each of them to match with the profile of one of the universities. In the team for the University of Kuopio, there were two Finnish and two international members. Later, a two-day joint seminar was arranged by the three universities to discuss the results and future steps based on the evaluation. The participants were rectors, deans, and key administrators, and members of the peer review teams.

The recommendations made in the report can be divided into three groups: those to the Ministry and FINHEEC, those common to all three universities, and those addressed separately to each of the universities. Those given to the University of Kuopio included:

- that the University direct more resources to activities supporting local industry. The role of the Centre for Training and Development and of the Open University was emphasised.
- that steps be taken to attract more money for technology research and to encourage
joint applications by the University and local industry to TEKES (Technology Development Centre) and other technology development funds

• to review its current arrangements for technology transfer and for patenting and licensing the products of its research and for the establishment of suitable mechanisms for the protection and exploitation of its intellectual property

• that a system be developed for ensuring that the overall training and environment properly meet the needs of all graduate students. Furthermore, that the graduate students be fully informed of the regional policy to enhance the ownership of the regional policy at this level.

• that mechanisms for strategic planning be consolidated, particularly in regional development. (Dahllöf et al. 1998, 61)

The Utilisation of the Evaluation at the University of Kuopio

Earlier Evaluations and Good Team Work

An atmosphere of enthusiasm, cooperation and pride came through most of the interviews. The interviewees mainly spoke about “us”, usually referring to the whole university, not so often to a faculty or a department.

3. “When thinking how we used it (the evaluation) all through the process, on the other hand it was for effecting discussion about the regional perspective, and on the other hand, as we had taken the quality perspective, how to improve the University…” (IA5)

The University already had experience of earlier institution-wide evaluations (Kinnunen et al. 1998a, appendix 2a). It had also developed criteria for evaluating research which is clearly appreciated in the university (D3, R3). A system for allocating basic funding had also been established. In addition, the University was planning an internal quality assurance system, as mentioned. Thus there was experience of institution-wide evaluations at the university (IA5, D1, CE2, R3). After the evaluation a well functioning management team and various seminars helped to keep up the discussion about the regional role of the university (R3, check IA1 ja IA2).

4. “Haven’t seen such team work often. Many of those who came to see us said afterwards that it was obvious that we made a good team. There was absolutely no friction. I felt that producing ideas – there were the signs of a well working team in it as well.” (R3)
Visits to the Faculties
During the self-evaluation process, those responsible for the adaptation of the quality framework to be used in the self-evaluations visited each faculty, to introduce the quality framework that was to be used in the self-evaluation, and to answer questions. Several seminars about the regional role were arranged for the University Council. In addition, deans had their own meetings about the evaluation (IA1, IA2, IA5), which strengthened knowledge sharing and the commitment of the participants (IA1). The discussion continued at strategic seminars conducted for the University Council (IA2). Generally, there is a culture of discussion at the University, which may relate to its small size and to its original structure where there were no faculties; people are used to crossing departmental borders (R3).

5. “This is still a rather small community and direct contacts between people work well, including the central administration. So I think that if there is something that is considered to be good and important, discussion starts quickly, and we are able to take necessary measures.” (D3)

One of the key actors in the self-evaluation considered the process quite orthodox in that it went right through the organisation (D1). An example of evaluation use at department level was also reported.

6. “… there were questions – now I’m talking about the department level – whether our degree programme is relevant. There were many situations that made us ask whether our programme has the right emphasis. Do we have enough contacts with the community? Of course the self-evaluation was an extremely important process for us, especially for the Department, but also for the Faculty. It offered appropriate information for addressing the most important questions, information that people make use of unless they are stupid. We really got feedback. So I think that the report was useful at the department level…” (R6)

According to two rectors, this was not the only example (R3, R6). Institutional administrators, however, were not convinced that the evaluation would generally reach down to the department level (IA1, IA2). One of the deans was happy to take on board the regional role and impact when his Faculty’s strategy was prepared, but he challenged the interpretation of the concept, and offered an interpretation of his own. (D7).

7. “The biggest problem … was that I see regional impact especially as support for entrepreneurial activities, but this self-evaluation report has a far more extensive perspective. It takes a wider sort of view and there are various ways of affecting impact. That was the biggest question, the different perspectives. … From the faculty perspective I started build-
ing regional impact from scratch. I wanted to develop measures to have regional impact at Faculty level.” (D7)

The understanding that the evaluation encouraged staff at both institutional and faculty levels to engage in strategic work is shared by institutional administration (IA2).

**Differences between Faculties**

The most active faculties in the evaluation process were Social Sciences and Natural and Environmental Sciences. In the interviews, their positive attitude and activity were explained at least partly by discipline: Due to their research focus and traditions they already had contacts with local stakeholders (D1, D7). In addition, there was substantial evaluation expertise in the Faculty of Social Sciences.

8. “Perhaps in the Faculty of Social Science these social questions are considered sort of self-evident, and the connection with, say, local industrial policy or regional development authorities or other stakeholders is taken for granted and more natural, as the focus of our research and development work is in the surrounding society, unlike those who focus on a cell.” (D1)

9. ”My background was in science education, and my view clearly relates to the fact that I always have the utility perspective in mind. There is always a decision, followed by a consequence, and the consequence should always be in line with the utility point of view. That’s my basis for thinking this regional impact as well.” (D7)

In spite of the eagerness of some faculties and departments to utilise the evaluation, institutional administrators felt that it was perhaps rather an institutional-level development tool (IA1, IA2).

**The Concept of Regional Role and Discussion about it**

Considering the location and history of the University one would expect that the regional role of higher education institutions would be well understood within the University. However, the concept was perceived as being confusing, and the theme challenging. The difficulty related at least partly to a mismatch between the University profile and local expectations (D1, P2). The concept was introduced and defined in discussions.
10. 
“... we were puzzled, we didn’t quite know what this regional task and impact, or interaction with the region as it was called, what it was. I remember the discussions that arose and these served as sort of definition. I think that the theme of the evaluation should be considered in relation to time. Now, when we talk about regional impact and the third task everybody understands right away that it means this. But at that time, when somebody spoke about the third task, it was not understood or its existence was simply denied. There were still these two old academic basic tasks, they were taken for granted. In that sense the evaluation process brought about ideas... It gave concepts and names... for this, what this was all about.” (D1)

It was pointed out that regional impact has always played a role in the University, but under different Rectors it had been given different emphases (IA1). In the interviews three different concepts were used – the regional role, regional impact, and regional development. Another area where the evaluation has been used as intended is in the building of a quality system. It was considered difficult as well. The heavy emphasis on the regional role first raised concerns that the University profile might be changed so that it was more like that of the polytechnics (IA5, D7), and the idea of a university-wide quality system others concerning excessive systematisation (IA5). However, the concept of the regional role was not a problem for all. Rather the evaluation of the university’s regional role was seen as an opportunity and a challenge (R3 below). The two quotations below give an example of both the doubtful and the confident attitudes.

11. 
“We had two extremely difficult issues to address at the University: quality and quality system, and regional impact. They both raised strong feelings and the question about where the university was being pushed. Were we moving away from basic research? Who is going to manage all this, and can all this be done systematically?” (IA5)

12. 
“It was very easy (to accept regional role as the theme of the evaluation). --- Focusing on regional impact, I thought that we were taking the most difficult task considering universities….tackling the most difficult task relevant to universities. I admit counting on that the university can not lose anything. Such a difficult task is worth taking, because we can get positive feedback, that’s sure, acceptance in the university without losing anything. Pioneer work is pioneer work, it develops thinking.” (R3)

The discussion about the concept of regional role started in the self-evaluation process and continued through the evaluation until the follow-up evaluation. The evaluation

---

1. Third task is the expression used to describe the importance of cooperation with business and community – in addition to the traditional teaching and research roles.
process was seen as attitude work (R2). The follow-up evaluation also inspired new deans to join in the conceptual considerations, even in the conceptual debate (D3, D7). The building of a quality system was also underway (CE1), but in the interviews it had a minor role.

13.
“The first thought about how to use the evaluation was to effect internal discussion about regional impact. Thinking of what we used it for all the time was affecting discussion, on the one hand about the regional perspective, and on the other hand, as we had included the quality perspective in the evaluation as well, about how can we improve ourselves with the emphases that we have chosen. Then, considering the regional perspective, the evaluation triggered the question, how the University really should act in this region. It triggered the discussion how should our own processes be improved.” (IA5)

Regional role was not the only theme in the discussions. Evaluation more generally was being debated continuously. That related also to the pre- and during-the-evaluation debates concerning the university-wide quality system then being implemented. The process of evaluating the regional role provided a good lesson in how to conduct and use evaluations. (IA1, CE1). After the process, the discussion about evaluation in general and as related to the quality system spread to departments and to university management and administration. It is also an issue in annual action plans and reports. However, the inclusive feedback system that was planned is not yet fully in place (CE1).

**Utilisation of the Reports**

Besides adding to institutional self-understanding, and offering necessary information to the external panel, the *self-evaluation report* served as an information source, when the University was setting up projects with the forthcoming EU funding schemes in mind. Further, both the self-evaluation report and the report of the peer review team have been distributed to external stakeholders and partner universities (IA5, R6). Perceptions about the internal utilisation of the external report differ within the University. Where one sees active utilisation, others see surprisingly little of it. Rather than being about the actual “amount” of utilisation, this probably identifies different expectations and positions, and consequently different experience and information about the situation. According to a dean the external report had been in active use at the University, as it revealed the important point that the profile of the University did not meet the needs and expectations of the region (D1). An institutional administrator, in turn, sounded somewhat disappointed with the little use made of the external report inside the university (IA5). The rectors, however, have use for the external report; they often refer to it in their speeches (CE1).
A Need for Leadership Training

In the self-evaluation process, a need for training in management and leadership was discovered as both university management, deans and department heads had to consider how to persuade people to accept the many problems revealed and changes required by the evaluation (IA5). The question was raised whether anything could be achieved without management/leadership, and for the first time professional management (ammattijohtajuus) came into discussion. Somewhat later, courses in leadership and management were introduced (IA5, D1). Further, it was noticed that the University was not as well-known as was assumed. Hence the image of the university became an issue, and it was understood that the University should learn more about marketing. Consequently, public relations activities were included in the University strategy (IA2). Marketing is not limited to general legitimisation of the University, but the University is now used as a flagship for the town, to improve its image as a location for industry (D5). One of the rectors wondered whether the University should have been more active in using the evaluation for public relations purposes (R2).

Intellectual Support and attitude Work

According to one of the rectors, the evaluation offered intellectual support for tackling the weaknesses pointed out by the external panel, and for reconsidering the tasks of the University, especially the third task. The involvement in the evaluation process of the key faculty personnel made them consider the same questions. All this had an influence on the strategies at institutional, faculty, and department levels.

14. “… as to utilisation, actually it took place all along the process, and there was an awful lot of attitude work in the process. - - - You see, the evaluation pointed to some problems and kind of organised them, and gave mental support … for me as Rector so that it was easier to answer the question about what is right and what is wrong and what is a task of the University and what isn’t. It is not a minor point, considering the utilisation of results. As faculties and their key personnel were involved in the evaluation process and had to consider these questions, it is my experience that these considerations are visible in the strategies of the faculties. The change in thinking occurred not only at the institutional level, but also at the faculty and department levels, even if perceptions were somewhat different in different parts of the University. So it is not homogenous, but nor is there a need for it to be either.” (R2)

Responding to local Needs

In order to respond to local expectations, the University has started new Master programmes, and established a new Faculty (IT and Business Administration). Further, it has established two development units to focus on research cooperation with local stakeholders. As a consequence of the evaluation,
15. “... researchers kind of noticed that the University of Kuopio, which traditionally has not had close connections with the region, because we focus on Medicine and related fields, so perhaps people realised that as a matter of fact we do have research groups that could have contacts with local business, if we only were able to find them. It quite shook up our research.” (IA5)

Although changes occurred following the recommendations of the evaluation, not all those changes can be attributed exclusively to the evaluation. For instance, an important inducement to change came through FIM 10 million (approx. 1.7 million), made available through EU Objective One funding (R6). In addition to this financial carrot, the discussion reflecting different understandings about the regional role of the University acted as a facilitator for evaluation use. The discussion was enabled by the interactive culture and the small size of the University (P2), and by the earlier evaluation experience. Further, a key person in the University administration ensured continuity, while Rectors and Deans changed. On the topic of evaluation use, it was suggested that the change was a positive factor.

16. “So... it was absolutely critical that you had someone coming from outside the university with strong private sector experience. Someone who knew about change management.... The interesting thing about Kuopio was the change of the Rector during the process, which was probably a good thing because it meant that other people had to accept ownership of the evaluation project...” (P1).

In addition to the conceptual considerations (that serve future orientation), the evaluation was used to develop strategic thinking and evaluation expertise, to give new routes for thinking (R3), and to strengthen networking with both regional development agencies and local authorities (IA5, R2, D1). The theme of the evaluation is alive and well in Kuopio, and the University has taken the role of an animateur. The follow-up seminar indicated that follow-up evaluation had been accepted by the key players throughout the system. This acceptance was easy for the University of Kuopio, because the university had always been focused and it did not have to make any fundamental shift in its direction. (P1, P2). In addition, the possibility of receiving EU funding, and the mismatch between local needs and the university profile encouraged the University to assume the active regional role recommended by the peer review team. The only problem referred to when discussing the utilisation of the evaluation was the nature of the University’s organisation (R3). However, its small size, interactive organisation culture, and earlier experience of university-wide evaluations helped to overcome at least some of the problems posed by the organisation.

In its report on the re-evaluation of Eastern Finland’s Universities, the peer review team commended the university of Kuopio for paying regard to its regional obligations without being overwhelmed by them (Goddard, Asheim et al. 2003, 56).
Critical Factors at the University of Kuopio
The interactive culture and small size of the University combined with evaluation expertise and experience of change management were the most important factors that enhanced the use of the evaluation at the University of Kuopio. In addition, there was a strong expectation among regional stakeholders that the university takes a stronger role in the region than before.

Biggest Benefits of the Evaluation
Increased understanding of the concept of regional role and the related internal discussion, progress in strategic thinking, and the building of new networks for cooperation were considered to have been the biggest benefits of the university evaluation at the University of Kuopio. The evaluation gave a name and a face to the regional role of the University (D1). In the process, knowledge of, and expertise in evaluation, and new ways of working together were developed (IA2, IA2). Further, the evaluation helped the University to prepare itself for its forthcoming application for funding under the EU Objective 1 scheme (R6, IA2). These benefits resemble the five types of process use discussed by Forss et al (2002), namely learning to learn, developing networks, creating shared understanding, strengthening the project, and boosting morale.

17. “... right now the biggest benefit is that by thinking about and undertaking such evaluations, by considering these questions and by reacting to the recommendations, it has become part of the process that produces understanding about what regional impact in this region could be, and what the role of the University could be in it.” (D3)

Purposes of Use
The main types of evaluation utilisation at the University of Kuopio were conceptual and interactive, followed by instrumental. I use the term conceptual instead of enlightenment as the emphasis was on defining a concept, that is, regional role or impact, and on coming to terms with it. (IA2, IA5, D1, D3, D7). Academic management and leadership, including deans, were deeply involved in the evaluation. However, only occasionally did the interactive use reach the department level, where ritual use seems the most common type of evaluation use (IA1, CE1, FA2, FA7). In the following I use the term ‘conceptual use’ to refer to conscious efforts to create a definition for a new concept or to redefine an old one.

Considering the four evaluation strategies described by Vedung, the University of Kuopio with its interactive approach and visits to all the four Faculties comes closest to the user-oriented strategy.
5.2.2 The University of Turku

**Internal Context**

The University of Turku was established in 1920. The University has six faculties: Humanities, Mathematics and Natural Sciences, Medicine, Law, Social Sciences, and Education. At the time of the evaluation in 1999 there were about 14000 students and 2200 staff. In 2003 when the follow-up evaluation was conducted, numbers were about 15500 and 2700 respectively (KOTA).

In the self-evaluation report, the University is characterised as an autonomous national institution of higher education with only limited regional goals and accountability. The question of how to respond to regional needs was a relatively new challenge for the University. The report states that the increased activity and willingness to co-operate is not only attributable to pressures from outside, but they also reflect a change within the University. “Increasing university self-regulation, and changes in the university funding base and in the higher education environment, which contribute to accelerating competition for students and staff, are all factors that call for greater external involvement.” (Puukka 2000, 1) As to the three types of higher education institutions considered by Bleiklie et al. (2000, 46–51), the University of Turku is closest to being describable as a cultural institution. According to one of the surveys conducted as part of the self-evaluation, the University is considered to be a traditional university in both the positive and the negative senses. All in all, it is often perceived as an old, traditional seat of learning with a static image and an air of conservatism. It is evident that for multidisciplinarity to work as an effective image, more attention should be given in future to highlighting the University’s manifold strengths and expertise (Puukka 2000, 15).

“The purpose of the evaluation project was to strengthen the strategic capacity of the University and to promote the development of the university-environment interface. The following terms of reference were adopted for the exercise:

- to assess the University’s external impact – whether cultural, social or economic in local, regional, national and international contexts;
- to identify the existing mechanisms and processes involved in the interaction between the University and its region, as well as the feedback systems from the region;
- to generate information for the revision of the University Strategy and the strategies of the faculties, departments, and special units, and to generate information for the negotiations regarding the three-year agreement with the Ministry of Education for the period 2001–2003;
- to promote the University’s external impact and social relevance relating to the provision of teaching, research and other services;
• to support the University in its effort to develop and improve co-operation with society; and
• to translate the results of the evaluation into effective management action and measures and procedures for the development of the University.” (Puukka 2000, 3–4).

The terms of reference further state that the evaluation will be first and foremost a process of learning and development resulting in improved interplay between the University and its stakeholders. The perceived multiple impact of the University will improve the self-knowledge of the University and help regional stakeholders to see the scope and extent of the University’s operations better (Puukka 2000, 4). Self-knowledge of the University by its internal and external stakeholders is likely to improve once they see the extent and ways the University has had an impact on the region it serves. The self-evaluation report states several times that the University has not yet been able to use its increased autonomy to the full extent. This is expressed in various ways, for example “...is still on its way from administration to management”, “A concrete move from administration towards management has yet to be made”, and “... the University of Turku has not fully exploited the new opportunities.” (Puukka 2000, 47, 55, 58)

Evaluation
The theme and timing of the evaluation were agreed upon between FINHEEC and the University in 1998. However, the idea of considering external impact dated back to an IMHE seminar that took place in Turku in the early 1990s. (IMHE is abbreviation for Institutional Management in Higher Education, which is an OECD project). The background was familiar to some interviewees, but not to all of them.

18.
“Actually, the idea dates back to 1993 and 1994 and relates to the management seminars organized by the OECD’s IMHE programme, and the then CRE (now European University Association), and one of them which .... took place here in in Turku in the summer of 1994. -- -- We thought that this (Response of higher education institutions to regional needs) is a theme that is gaining importance, and might be useful for the University of Turku as well as an old University as even though we had some kind of a regional role in the early 90s, it was far more vague than now. In that respect this has been a good theme for us.” (IA3)

At the time of the interviews, no changes in key management positions had taken place at institutional level after the evaluation. At the faculty level, four out of the six deans were new. There were no major changes among the faculty administrators.

The responsibility for the evaluation project lay with a steering group chaired by the Rector. The seven members of the group represented academic leadership (two
deans, and a professor), institutional administration (Director of Administration, Head of Planning, and the Coordinator of the self-evaluation project), and FINHEEC (represented by the author). The self-evaluation consisted of a self-evaluation of all faculties and special units, based on a questionnaire, and of five sub-projects that covered educational impacts, research impacts, impacts of mature age education, economic impacts, and social and cultural impacts. The sub-projects involved surveys being carried out, supervised by respective project teams. There was one Finnish member in the five-member peer review team. A four-day site visit took place in late 1999. A follow-up evaluation was conducted in 2003.

The report of the external panel produced about 30 recommendations, presented under eight headings:

1) **the context of the evaluation and approach**, including a recommendation to create internal mechanisms to ensure that the information gathered in the self-evaluation reports are widely disseminated inside of the institution

2) **strategic direction**, emphasising the need for external engagement

3) **the role of the rectorate and faculties**, recommending among other things a review of the structure of faculties and departments, and the appointment of a third Vice Rector, and the establishment of a Regional Development Office

4) **the changing role of learning and teaching**, suggesting the establishment of a new Learning and Teaching Council, and pointing to the need for quality assurance procedures

5) **the role of research and research services**, recommending the establishment of a Research Council to have oversight of research strategy and policies, and the engagement in multidisciplinary research

6) **relations with the external environment**, suggesting i.a. that the University take the lead by establishing, in partnership with the relevant public and private organisations, a forum in which a long term strategy for the development of a knowledge base learning region can be worked out

7) **the role of the centre for extension studies**, emphasising the need to integrate the Centre with the academic heartland, and

8) **administration and management**, recommending a clarification of the roles and responsibilities of the Faculty offices vis à vis central administration of the University.

(Goddard, Teichler et al. 2003, 43)

The eight headings above describe the broad approach of the peer review team. As the title of the report suggests, the idea of the peer review team was that impact is created through “External Engagement and Institutional Adjustment”.

116 – Anna-Maija Liuhanen
The Utilisation of Evaluation at the University of Turku

At the University of Turku the interviews were conducted about three years after the evaluation. The University was preparing for a follow-up evaluation, to be carried out some months later. To my question about what the utilisation of the evaluation brought to mind, the interviewees often responded by referring to the measures taken at the university as a consequence of the evaluation. Each of them considered the questions from his/her own perspective – rectors and institutional administrators from the university perspective, deans and faculty administrators mainly from that of their own Faculty. Several interviewees discussed the purpose of the evaluation, thinking that it was an operation of the University management or FINHEEC, that is, something demanded from those above (R4, IA7, IA4, FA4, FA5). It was assumed that the purpose of the evaluation was to strengthen the University’s image and external relationships (IA6). However, the evaluation was also expected to provide an opportunity for learning and development. The interviewees’ expectations were matched by the terms of reference presented in the self-evaluation report, which in addition to the external impact of the University included the following:

- to generate information for the revision of the University Strategy,
- to promote external impact and social relevance,
- to develop and improve co-operation with society, and
- to translate the results into effective management action
(Puukka 2000, 3-4).

A Management Instrument

The evaluation was an instrument of the operational management, the measures of which needed the blessing of strategic management, that is, the University Council (R1). The Council was advised about the evaluation project and its theme when it was launched. Later, the Council’s discussions concerning the evaluation were combined with discussions and decisions about university strategy (IA3). The broad lines of the evaluation were decided by the seven-member steering group mentioned above. The main responsibility for the implementation lay with the project coordinator. In addition to official bodies, at the time of the evaluation there was an informal group of academic leaders, consisting of the Rector, Vice Rectors and Deans.

19.
“For a while a group of academic leaders emerged to consider what should be done about the problems of the university.” Is it still alive? “Well, of course it left some traces, but it is perhaps not as intensive as it had been.” (D2)
Top-down Information Flow

The deans’ descriptions of their participation in the evaluation process varied from it being seen as a duty and necessity, through to active and determined participation and ownership of the evaluation (D2, D6, D8, D9). The University Bulletin and other channels were used to inform university staff about the evaluation. Such was the eagerness that the expression ”all information channels” was used to describe it (FA3, R1). However, options for interactive utilisation of the evaluation process remained limited. In most faculties, deans and faculty administrators were responsible for information gathering and reporting. In spite of that, some faculty administrators felt that they were outsiders or merely data collectors (FA1, FA3, FA4). During the self-evaluation, the project coordinator acted as a link between institutional administration and other levels, and visited the faculties, which was clearly appreciated, but she alone could not make the process more interactive. (R4, D9, FA5, FA7). According to an institutional administrator the self-evaluation was a learning process for university management, whose commitment developed during the process (IA6). In spite of there having been a concluding seminar (attended by more than 200 participants), and some later follow-up seminars, more discussion at the end would have been welcomed (R4, IA6).

20.

“Well, we made a classic mistake here in the sense that we haven’t discussed this (external report) as much as we perhaps should have since it was published. That’s what easily happens in these cases. I mean that in the wider university community it has not been discussed very much. I think the management should have been more active to ensure that the discussion kept going.” (R4)

In line with considering the evaluation as a management tool, the departments were not perceived as users of the evaluation, but rather as informants and recipients of information. There were some staff in the departments who would have welcomed at least some feedback concerning the self-evaluation reports that they had prepared material for (R4, FA1).

Differences between Faculties

The Faculty of Social Science was described as being generally critical – evaluation was no exception, while the Faculties of Education and Science had a more positive attitude (R7, D2, FA4). In the Faculties of Medicine and Law, which already had well-established forms of cooperation with external stakeholders, the evaluation was considered to be late (FA1, FA5). To build the self-evaluation on dialogue, the Faculty of Education invited all staff to contribute via an intranet inquiry. That way, it was thought, the Faculty would get essential information about its own strengths and weaknesses, and perhaps move towards a ‘learning organisation’ (Niitemaa 1999). The self-evaluation questions were considered applicable, even though they were focused
on general issues rather than on the work of individual teachers or administrators. A dean suggested that it was useful if people feel that they belong to the University community, and take responsibility for it; the evaluation offered an opportunity to be heard and express one’s opinion on problems and developments in the University (D2). As part of the self-evaluation, the faculty also collected data about the research cooperation between its staff and external partners. The extent of it was a positive surprise, the information has turned out useful, and it has been used to establish a database to be regularly updated (FA5). This indicates the emergence of a culture of evidence and knowledge transfer in the Faculty.

Utilisation of Self-evaluation and of the Evaluation Reports
Although the administrative approach of the self-evaluation was criticised (FA1, FA3, FA4), there were also those who considered it useful and felt that they learned from the evaluation. It was in the process where both the self-understanding and the understanding of the importance of cooperation with external stakeholders were gained (D9, FA6). The evaluation process also encouraged strategic thinking (IA4), and gave impetus for planning, which had not been done properly before (D4). Further, the writing of the self-evaluation report was rewarding, as people noticed the achievements and strengths of their department (CE2).

21. “... when first looking at it, we asked ourselves is our work really this great. Many felt this way. And when looking at their own text, many said that I thought that the story was quite modest, but when the story was ready, it seemed quite good. - - - Then, when one of the panel members told us that as he saw it, and he has seen and evaluated a number of European universities, that regional development work at our continuing education centre is at a better level than in Europe in general. For us this was of course very nice to hear.” (CE2)

The self-evaluation report is analytical, and the challenges are recognised. The author is well acquainted with the theme. However, the report was not discussed or widely debated at the University. Although the aims of the evaluation were carefully formulated in the self-evaluation report, people were not clear about them. Rather the aims were considered vague (R4, IA7). For institutional administrators, the self-evaluation report worked as an awakener (IA4, IA7) and served the enlightening purpose hoped for by some faculty administrators—”that they would better understand us” (FA1, IA4). However, the self-evaluation report of the university seems to have been less used than the external panel’s report. The evaluation project was described as a public relations exercise, which was not aimed at any change (IA6), and perhaps even as an opportunity to show off (D8). The way of organising and conducting the evaluation project indicated that senior management did not expect that the project would be ‘owned’ by everyone at the University (R1, IA3).
Right after the evaluation, the report of the external panel served as support and as an awakener for institutional management in launching the concept and ideas of external engagement, and in implementing the recommended changes (R1). However, external engagement was not a brand new idea at the University. There had been extensive research cooperation with external partners well before the evaluation, but it had been based on the interests of individual researchers or research groups rather than on an institutional policy (D4).

22.
“... almost all who are appointed as professors want to be researchers in the first place. Even teaching is kind of extra business, and for some the contacts with local business and industry are more important in the sense that they often relate to their research interests, and I would say that almost, that in our Faculty there are quite a number of people for whom external engagement and contacts with local business are quite natural. But so far it has not been understood that it could be part of this planning, that it would be considered as part of this community's activities.” (D4)

The peer review report has been actively used by senior management to help in sharpening the University's image and to support change. In particular, rectors and key administrators use it for public relations purposes and they have sent it to colleagues at other universities. (R7, IA3, IA4)

Since the evaluation, the University implemented a reasonable proportion of the changes recommended by the external panel. It became involved in the Turku Science Park, and started new degree programmes in two nearby towns in which it already had some activities before the evaluation. Further, it established two Councils, one for Teaching and another for Research, and appointed a third Vice Rector, in accordance with the recommendations of the peer review team (Goddard, Teichler et al. 2000, 43–47). The position of deans as academic leaders has been strengthened, and new arrangements in institutional administration have been made (R1, CE2, IA4).

On the Focus of the Evaluation
The theme of the evaluation, external impact, was chosen by institutional management to improve the University's external impact and relationships. However, there were also other interpretations about the reason for choosing this theme. According to one of these interpretations, the evaluation was focused on external impact in order to avoid discussion of internal problems, which refers to an assumption of political utilisation. According to another, external impact was only a label under which a more inclusive evaluation was conducted. (D8, D2).
As the theme of the evaluation was external impact, many thought of a summative evaluation. Hence, the approach of the peer review team, emphasising the importance of internal processes, was a surprise to many. The title of their report, External Engagement and Institutional Adjustment, is descriptive of the approach.

**Critical Factors**

From the point of view of evaluation use, the loose coupling and the size of the university can be considered to be problems, especially as the “walls between departments are made of concrete” (P2). This, in turn, relates both to the size and the traditions of the university. No major efforts were made to overcome the loose coupling, or to involve people in the evaluation process. Rather, institutional-level measures were expected to have an impact on other organisational levels in due course (R7). Discussion or debate concerning the evaluation and its theme remained limited (R4, P1). The group of academic leaders served as a forum for discussion only for a short while during the evaluation process, not after it (D2).

In addition to the loose coupling and individual pursuit for glory, the role of “right persons” and the change of key personnel were considered critical from the point of view of evaluation use. The latter speaks about the lack of continuity and integrating mechanisms. Hence, ignorance can become a barrier to the use of evaluation. On the other hand, the utilisation of evaluation was better in cases where a “right person”, that is, someone interested in the evaluation, took over the deanship or the position of department head. (D6). The decisive role of right persons was described as follows:

"...evaluation becomes useful, if there happen to be suitable persons to use it, but evaluation as such does not lead anywhere, unless it kind of hits some people who take it on board and work for it.” (D6).

It was suggested that institutional administration should see to it that staff at the lower levels of the University understood the purpose of evaluation, and the benefits that could be expected of it. With some marketing, perhaps the usual complaint about the central administration and their assumption that staff in departments are not already
fully occupied (“even more work without more resources”) could be avoided. (FA7). It seems that in spite of the major information activity before evaluation, people were not aware of the background and possible benefits of the evaluation.

On the Consequences of the Evaluation
The main owner and user of the evaluation was the Rector, for whom the external report was an important tool both inside and outside the University. The evaluation did not go very deep into the organisation, and there was no clear change management (P1). The University followed the recommendations of the external panel so carefully that a faculty administrator spoke about obedience (FA5). However, external engagement is not integrated into either research or teaching, nor have the new Research and Teaching Councils taken on that role. When speaking about the Councils, none of the interviewees referred to external engagement. Thus, the connection between structural changes and external engagement is not clear (P2). Locally, the University has not assumed the role of a key actor. Hence external engagement is not a living part of university policy, even though engaging in the ‘third task’ is rather easy for the University due to the strong region and industry. All that is needed is that staff get involved (P2).

In their re-evaluation report, the peer review team states that despite the enhanced university-region interface, the University has not assumed a leadership role in technology transfer, commercialisation of research results or generation of new businesses. (Goddard, Teichler et al. 2003, 18).

Biggest Benefits
The major benefits of the evaluation were considered to be the improved visibility and sharper image of the University, followed by increased cooperation with regional stakeholders (IA3, IA6, D9, R7). This is in accordance with one of the aims set for the evaluation – “to promote the University’s external impact.” (Puukka 2000, 151). After the evaluation, the importance of cooperation was better understood (R7, D8), and the self-understanding about the University had improved (IA3, D6, FA6, D9). It was emphasised, however, that the University had only just started, that it had not come very far yet, and a lot remained to be done (R1, R4).

25.
“I emphasise that we’ve only just started, that we haven’t come very far as yet. - - - Another point (merkitys) might be that it kind of kick-started certain things, I mean towards a change... - - - Perhaps it will also familiarise the university with a different culture.” (R4)

Further, due to focusing the evaluation on external engagement, the University felt perhaps better prepared than other universities when the regional role of higher education institutions became an issue on the national higher education agenda (R1).
**Purposes of Evaluation Use**

There is not a single type of evaluation utilisation that would characterise the whole University of Turku, or even all those involved in the evaluation. Institutional management used the evaluation mainly for instrumental and legitimising purposes. The strong emphasis on instrumental use in the interviews may to some extent be due to the short time since the evaluation; at the time of the interviews the decisions made and the measures taken were fresh in each interviewee’s memory. An alternative explanation is symbolic use. The measures based on the evaluation were taken to give an impression or to send a message rather than to effect real changes (IA6, D8, P1). At the faculty level, enlightening use seems the most common. I use the word ‘enlightening’, and not conceptual, as there was no conscious effort within the University to define or to discuss the concept of external engagement or the role of the university. Evaluation use in the departments and in most faculties remained small. It was described by some as answering questions that had been asked by staff at the institutional level rather than actual evaluation of the department and its work (R4, FA5), which refers to ritual use. Some faculties took a more active approach, and it is of course possible that some departments were also able to make use of the evaluation process. Thus, the self-evaluation was mainly administrative in that the project was like any other information gathering exercise required by central administration. However, as described above, in one faculty a more interactive self-evaluation process was used.

Considering the whole University, the utilisation of the institutional evaluation can be described as split – internally administrative (instrumental and ritual), and externally political. Administrative utilisation refers to the emphasis on top-down information during and after the evaluation, and political to legitimising use, especially in relation to stakeholders. Individual enlightenment was mentioned in several interviews, while a shared learning experience was less discussed even though one of the aims set for the evaluation was a “process of learning and development”. The approach of the University with the top-down information flow can be described as being a combination of the diffusion-centered and the production-focused strategies described by Vedung (1997, 265–266).

5.3 Common Features

In spite of obvious differences in the utilisation of institutional evaluation in the case study universities, there were also similarities. These are discussed here. First, an answer to each research question is given, and then the role of evaluation and context in the utilisation of university evaluations is discussed. The relationship between utilisation, organisational learning, and learning organisation is discussed in Chapter 5.4.
5.3.1 Users and Internal Context

For each university, the evaluations can be described in the first place as being an instrument for use by rectors, vice rectors, and key administrators, that is, institutional level management and leadership. As users, they represent the organisation. Despite the differences between the two universities, the further down the hierarchy one looks, the more peripheral the use of university evaluation becomes. Obviously, the institutional focus of the evaluations is closest to rectors and institutional-level administrators. However, some of the deans and faculty administrators turned out to be what could be described as active users. They took ownership of the evaluation process (D1, D2, D4, D6, D7, FA1, FA5), and considered it beneficial for their faculties (FA1, FA2). Other deans and faculty administrators described the evaluation from a distance, in some cases with an air of disappointment or frustration or in a rather cynical tone (D8, FA3, FA4, FA7). Without active participation by deans, participation by faculty administrators remained limited (FA5, FA7). However, the example of evaluation use at the department level indicated that an institutional evaluation can reach departments and could also be useful there (R6). For that to happen, it is necessary to have people to carry the message.

It is usually assumed that the user of an evaluation will be an organisation. This was also the case in the university evaluations. However, the role of individuals was important, as pointed out by Patton (1997, 44–61) and Vedung, who refers to champions (1997, 285). Individual staff members, also called ‘the right persons’ or ‘animateurs’, were needed for three main reasons – to encourage others to get involved in the evaluation, to take care of dissemination, and to represent continuity. ‘The right persons’ were characterised, for example, as people who take the message and make things happen, as change managers and as animateurs (P1). Ideally they should be either academic leaders or high-ranking administrators (FA7, D2, FA5, IA5). Representing both the academic and the enterprise dimension of universities, deans and department heads were in a key position, but not all deans, let alone department heads, assumed an integrating role in the evaluation.

Lack of continuity was one the problems related to evaluation use, exemplified by change of key personnel. This relates to the role of individual actors, but it also refers to the management system of Finnish universities, where rectors are elected for five years, and deans and department heads usually for three years. The change of key personnel was mainly spoken of as an obstacle, but obviously it could be a facilitator as well. To put this in the terms of Clark and Weick, there is a lack of integrating elements in loosely coupled organisations (Clark 1983, Weick 1976). Another example of loose coupling is weak vertical information flow (R4). In both universities it was difficult to perceive the situation from institutional level down to the departments. The faculty level is closer and thus better understood (IA2, IA4). Institutional level management assume, erroneously, that the message about evaluation and its theme will find its way through the organisation, as faculty administrators represent continuity (R7). They
do, but in disseminating the message and especially in effecting change, they need the support and commitment of academic leaders (FA3, FA4, FA5, FA7).

In an organisation where people are mainly interested in personal “pursuit of glory”, the task of “selling” something like a university evaluation is not an easy one. When speaking about selling or marketing the evaluation to the academic community, the interviewees actually said that in order to get people involved in and committed to the evaluation, those people needed to understand the reasons for, and the aims of, the evaluation. They should also be able to appreciate the possible benefits for the university, for their faculties or departments, or for themselves (IA7, D6, R3). The key role of individuals and the difficulties in dissemination and implementation mirrored the experiences of Swedish quality audits and US institutional accreditation discussed above. (Nilsson and Wahlén 1999, 12; Dill 2000, 187)

Experience in, and knowledge of evaluation and a positive attitude towards evaluation facilitated its acceptance and therefore its utilisation. This relates to the user’s capacity to receive and apply information, a point emphasised in Vedung’s user-oriented strategy (1997, 285), to the resources available as considered by Fullan (1994, 100), and to the capacity to use evaluation, as discussed by Feinstein (2002, 436–437). Other related terms include ‘capacity for change’ used by the Finnish Ministry of Education, ‘capacity for action’ (Bauer et al. 1999), and ‘change management’ (Fullan 1994). Further, in addition to individuals, a team of committed individuals was important in communicating the message and making changes (P1).

The internal context – especially the loose coupling and individualistic character of universities – was an obstacle that the active users and facilitators of evaluation use had to overcome. (Cf. Vedung 1997, 285, about the role of users and structural features in the utilising organisation) Evaluation capacity, interactive culture, and interested individuals helped in overcoming obstacles. As regards the participation in and utilisation of university evaluation, the message of the universities was not obligatory, but left room for individual choice. In the circumstances, the dissemination of evaluation through the whole organisation was difficult, and – considering the size of the universities – the number of users of evaluation remained small.

Ministry of Education and Members of the Peer Review Teams as Users of Evaluation

As was the case with university personnel, people at the Ministry of Education also see the usefulness of university evaluations and their importance in the governance process somewhat differently, each from their own perspective. The external view of the reports of the Peer Review Teams is appreciated and most of the information offered by the reports is considered to be relevant, in spite of the fact that the reports are addressed to the university in question in the first place (M3). The reports inform the national governance process, but they cannot always be used directly in the annual performance negotiations between universities and the Ministry of Education. Typically they are useful when the profile or initiatives of a university are on the agenda (M2). In
addition to accountability, the evaluations serve the enlightening purpose at a general level by helping the Ministry to understand the universities better.

From the Ministry perspective however, there are some problems related to university evaluations, especially concerning reports of the external panels. First, if the recommendations are very general and the style of the report is too ‘polite’, it reduces the value of the report, as it cannot be used for working up a development plan for the university in question. Consequently, follow-up by the Ministry becomes more difficult (M1, M3). Both Dill (2000, 1999) and Stensaker (1997, 64) criticise evaluation reports also, because of the veiled language used in their writing. However, it was assumed that from the university point of view this could be a source of relief. First, if the recommendations are so general, the university could perceive that nothing needed to be done (M3). Second, evaluations tend to take too much time to serve national level decision-making (M2). Third, the utilisation of evaluation reports in the university governance can be considered problematic. The problem relates to the autonomy of universities, and to the role of the Ministry. How detailed can and should the control be? Where is the border line between the Ministry’s power and the university’s autonomy? (M1) A more practical problem is the process of ‘steering-by-results’ as such, which leaves little time for becoming properly acquainted with the recommendations emanating from evaluation reports (M1).

Sometimes the Ministry expects the universities’ reactions before universities are ready to implement change. When major structural changes have been recommended, it seems to be years before the universities are ready to take the suggested measures. (M3). Further, the dialogue between universities and the Ministry tends to remain at the institutional level. Whatever the message, it is difficult to get it through to the last department, and even if the management is active, it takes time (M2). The problems of dissemination seem to be well known at the Ministry as well. It was emphasised that evaluation use extends beyond decisions based on evaluation, and both internal discussions (which are essential), and external discussions in the local media, are important from the point of view of social relevance. (M2). Thus, the Ministry also sees the difficulty of dissemination and the importance of dialogue.

The above assumption, that universities are happy with general recommendations which can be interpreted as suggesting that nothing apparently needs to be done, does not quite go together with the understanding about trust between the Ministry and universities referred to by Hölttä (1995, 34). Further, the above comments suggest that Rekilä is correct when arguing that the role of evaluation has remained unclear in the governance process of Finnish higher education (Rekilä 1996, 86).

While the evaluations were used mainly for accountability and to some extent instrumentally and for enlightenment by persons from the Ministry, members of the Peer Review Teams used them primarily for enlightenment and networking. They learned from their involvement with the universities under scrutiny and were able to take good practices and ideas to their own universities. The evaluation also offered
an opportunity to network and to share experiences internationally, perhaps even leading to new career developments. It was also said to be energising (P1, P2). This corresponds with the types of process use discussed by Forss et al. (2002). In addition, peer review can also be considered as an option for benchmarking.

5.3.2 What is used – Process, Reports, or Both?

As both parts of an institutional evaluation – self-evaluation, and external evaluation – consist of a process and a report, there are two processes and two reports that can be used. In both case study universities there were more people involved in the self-evaluation process than in the external evaluation process, that is, in the interviews conducted by the external panel. Most interviewees did not differentiate between self-evaluation and external evaluation processes, but considered them together as a single evaluation process. Process and reports, in turn, were mainly considered separately, but there were also those for whom the evaluation was considered as a whole. Generally speaking, each of the four parts of the evaluation – the processes of self-evaluation and external evaluation, and the respective reports – has been utilised in both universities. As the self-evaluation process, including data gathering and reporting, lasted for several months, and kept dozens of people busy at both universities, it can be assumed that it offered more opportunities for making use of the evaluation process than the interviews conducted by the external panel. Compared with the self-evaluation process, the two or four days’ site visit by the peer review team was very short, and the number of those interviewed small. On the other hand, rote responses to given questions do not encourage evaluation use.

An Assumption

It was suggested that different parts of evaluation are useful for different levels or parts of an organisation – evaluation process for faculties and the academic community, the self-evaluation report for those responsible for various special functions (student affairs, continuing education, external relations, research services etc.), and the report of the external panel for university management (D2). The assumption is quite telling, but there is more to it, discussed below.

In the Process

Process participation was highly valued by those who participated (cf. Marra 2000, 33–34). It was also the part of the evaluation that offered an opportunity for dialogue and learning, served the dissemination of the evaluation theme and message, and paved the way for implementation. Thus many interviewees considered the evaluation process more useful than the reports. The reason is the very task of evaluating that makes people stop, think and write about their work. Daily routines leave little time for
reflection (CE2). The biggest benefits come from a thorough self-analysis, which may open eyes to new things – at least for those involved in the process (R4). Accordingly, several interviewees described how they had learned about their own unit or about the University. “You learn from the process”, they said (IA4, D6). The deeper people were involved in the self-evaluation, the better they understood the reports, as they were aware of the background and the discussions that led to the final report. (D6).

Especially useful seems to have been participation in the self-evaluation process as most of those who have experience of it valued it more than reports. Process participation was described as a rewarding experience (CE2, FA1), and as a challenge (cf. Valovirta 2000, and 2002b). Further, interactive self-evaluation made it possible to create shared understanding, to strengthen the project, and to boost morale (Forss et al. 2002). Most of those who were either not involved at all in the process, or who had a minor role in it understandably preferred reports. (R7, IA1, IA2, FA3).

The interview by the external panel was appreciated as an opportunity to engage in discussion with esteemed academics, a discussion which challenged one’s own truths, and provided new perspectives (D1, D6). This opportunity was available only for a few. The possibility of creating internal and external networks was another benefit of the whole evaluation process (D6), also limited to a rather small number of people compared with those involved in the whole process. Further, learning from the process was referred to, but there were two kinds of learning, individual and organisational (IA4, and IA5 and D2, respectively).

The emergence of shared understanding suggests that the self-evaluation process can act as a means of dissemination and prepare the ground for the message and recommendations of the external panel, and for further discussion concerning the evaluation theme. The dissemination that took place during the self-evaluation process was mainly internal. However, the contacts with external stakeholders during self-evaluation also offered them an opportunity to learn about the theme and the university, and probably created expectations as well. As suggested by Saarinen, evaluation acted as an arena for discussions and dissemination (1995, 232).

Utilisation of the Reports
For many, the self-evaluation report was the first comprehensive picture of the university ever – not a history or an annual report, but a description of the whole university as it is now (R7, D1, D2, IA3). For others, it offered more detailed information about other parts of the university (R7), and worked as an awakener (IA4).

The nature of the self-evaluation report as a general picture of a university also makes it useful outside the university (R6, R7, IA3). The reports have been made available to esteemed visitors, to provide an opportunity for colleagues from other universities and to local partners to learn more about the University (R6, IA5, FA4). Further, it is always possible to refer to the reports for checking (R3, R5). Faculties, however, have not made use of self-evaluation reports to learn from each other (R4,
They knew very little if anything about how other faculties carried out their self-evaluation, or how they utilised the evaluation process and reports. Institutional administration is poorly informed about possible use of the evaluation in faculties and departments. “I know very little about departments and how they reacted to the evaluation; my knowledge ends with knowing the deans”, said an institutional administrator (IA2), providing an example of loose coupling within the university.

The report of the external panel is the mirror that is often spoken of in connection with evaluations (e.g. Davies et al. 1999, 9). As there was no major criticism of the judgments and recommendations of the Peer Review Teams, it seems that the reports, including conclusions and recommendations, were considered relevant and acceptable by the case study universities. (Feinstein 2002, 434–435; Lampinen 1992, 29) Consequently, the reports had been used for both developmental and public relations purposes. As described in the university-specific chapters, both universities have taken a majority of the recommendations on board, and made decisions in accordance with the reports. Further, rectors often refer to the external reports in their speeches and in various staff development seminars (D6, R6, CE1, IA6, R7). They also go back to the report every now and then to refresh their memory or to check something (R3, R5, R7). The report of the external panel can provide support for rectors when they are arguing for and implementing change (R1, R2).

**How about the Assumption?**

The above assumption about the usefulness of different parts of evaluation to different stakeholders is correct in the sense that rectors use the external report more than others. Further, as assumed, deans and some faculty administrators have benefited from the process; their use of the external report is not as visible. As few staff participated in the self-evaluation, enlightening and conceptual use are mainly limited to staff operating at the institutional and faculty levels. Those interviewees who represented special functions (continuing education, study affairs etc.), did not place special emphasis on the utility of the self-evaluation reports. Those who did make use of the self-evaluation report were mainly institutional and faculty administrators, the latter mainly to learn about their own faculty.

Even though many considered the evaluation process more useful than the reports, both are essential parts of the evaluation. The self-evaluation report describes the university “as it is now”. The external report, in turn, is a reaction to the self-evaluation report, and presents a possible future. For the case study universities the external reports offered challenging ideas about universities and their external engagement. Further, the external report includes an external opinion about the evaluated university, and it is used as support when implementing change. Both reports contain detailed information about the project and about the university that can be used later. The dialogue continued in the follow-up evaluation. Thus the challenge mentioned above also relates to the reports of the peer review teams.
As Forss et al. (2002) argue, it seems possible that those who learn most are the evaluators. From the perspective of the evaluators, evaluation can be characterised as a task with a view, but they are not the only ones who learn. There were a number of those at both universities who said that they had learned in the process. At the University of Kuopio, much of the learning related to the concept of regional role, and how to carry out and utilise evaluations (D1, IA1). At the University of Turku, learning was more individual. As argued by Forss et al. (2002, 39), networking related mainly to the members of external panels and to university management, including the internal coordinators of the evaluations.

In conclusion, external reports have been used mainly for developmental and legitimising purposes, internally and externally, but also for enlightening purposes. The self-evaluation reports, in turn, mainly serve the information needs of the external panel, and to enlighten institutional administration.

5.3.3 Different Types of Evaluation Use – how?

All the different types of evaluation use discussed in Chapter 4 – instrumental, enlightening and/or conceptual/organisational learning, legitimising, tactical, symbolic and ritual, and interactive – can be recognised in the interviews. From the user’s perspective, the ‘types’ or ‘dimensions’ of using evaluation are purposes of use. For example, when utilising the evaluation process or results in a legitimising way, a university or an individual uses it for legitimising purposes, irrespective of the evaluation’s original purpose. In the following, the utilisation of evaluation in the case study universities is related to the different types of evaluation use discussed above.

The types of evaluation use that were described most often are instrumental, enlightening or conceptual, and legitimising. As to instrumental use, both universities took measures based on the recommendations of the external panels. Legitimising use, addressed to both internal and external audiences, is also common. At first, both symbolic use in the sense of conveying a message, and ritual use as doing one’s duty, that is, cultural utilisation of evaluation (Albaek 1997, 387–393), seemed practically non-existent. However, the way in which the self-evaluation was carried out in the departments and in some faculties – answering questions from above – represents ritual utilisation (Segerholm 2001, 430). Thus, behind instrumental, conceptual, and legitimising use, there is a layer of ritual use as well. This means that even though institutional evaluation formally reaches down to the department level, it is so strongly seen to be a central administration exercise that it is difficult to make it into a development tool (Saari 2003). In those cases, self-evaluation hardly assists with dissemination. As regards symbolic use, those who referred to it did not report their own way of using evaluation, but spoke about others’ use (e.g. R4, IA1, IA6, IA7, P1, P2). This relates
to the valuing of different types of utilisation discussed by Albaek. Cultural utilisation is not highly valued.

Instrumental use, the type of use originally aimed at when speaking about the utilisation of evaluation (e.g. Vedung 1997, 269–272), seems the easiest to recognise and to report. It is the rational type of evaluation use, expected to lead to improvement. Typically this will involve measures to create, improve, or renew something, be it structures, processes, products or programmes (Albaek 1997). Thus, it has visible consequences. However, perceptions about reasons for, or developments leading to, such measures vary according to what people know. Further, it cannot be taken for granted that measures based on evaluation automatically lead to the expected improvements. For example, the establishment of the Faculty of Information Technology and Business Administration at the University of Kuopio was meant to enhance external engagement. However, if the university staff do not see the connection (FA7), they may interpret it as symbolic use. On the other hand, the effects of the decision can only be expected in the long run (Weiss 1999, 471). Meanwhile, the decision can be used to demonstrate that the University takes seriously the expectations of local stakeholders, i.e. for legitimising purpose.

The external report and the recommendations it provides offer support for implementing change, and for polishing the University image. Thus, it is used for both instrumental and legitimising purposes. When referring to the report, people refer to the expertise and authority of the panel members to convince others. They use the report, and the external panel, as backing (R1, PP, R7), or as an authorisation (D1) when speaking in favour of change towards regional responsiveness or external engagement. It can be considered to be a combination of instrumental and legitimising use. Rectors spoke about the intellectual support that the evaluation gives to measures that otherwise might be questioned in the university (R1, R2).

26.
“Well, this pointed to some problems and gave intellectual support so that it was much easier as a rector to answer questions like what is right and which tasks belong to the university and which don’t. It really isn’t a minor issue, when thinking about the utilisation of the results of the evaluation.”(R2)

In the main, those who used the evaluations for instrumental and legitimising purposes, were rectors and deans. They have the power to make major changes in their universities, and they – in the first place – are responsible for the stakeholder relations and for the image of the university.

Instrumental use and legitimising use are often combined. Changes (improvements) made as a consequence of evaluation are used for positive or general legitimation of the university (R7, R2, R1). When the recommendations of a Peer Review Team are used for effecting change, it seems to be an example of instrumental (rational)
use, but it can also be seen as political or symbolic use – evaluation as a pretext for tackling issues that do not relate to the evaluation at all (D8). An observer interprets the purpose of the decision-maker from his / her own perspective, based on his/her knowledge of the situation, and possible biases. Only the users know their own purposes. In the meaning of gaining time or avoiding responsibility political (tactical) use seems practically nonexistent.

At the department level ritual utilisation was probably the most common way of using institutional evaluation. However, it was not described as evaluation use simply because people did not feel like using it, or seeing others making use of it. Rather, it was nothing but one more data collection and reporting exercise among all the others. As to symbolic use, the fact that university evaluations were required by a higher authority suggests the possibility of symbolic use – you evaluate because you have to or are expected to do so (Saari 2002, 69; Sundqvist 2002). As a whole, it seems that the two universities were able to turn an obligation into an opportunity, and to avoid mere symbolic use. However, it was suggested that some of the measures based on the evaluation were taken to give an impression rather than to effect real changes, which refers to symbolic use (P1). Another explanation is also possible. There is not enough capacity to implement necessary change. The point is that symbolic use was not the decision-maker’s purpose. The problem was perhaps that there was an insufficient understanding of change management.

Conceptual and Enlightening Use
Both conceptual use and enlightenment refer to learning. However, in this study the two are used as separate concepts to emphasise the difference between the two universities in what can be referred to as learning purposes. Conceptual utilisation refers to an approach that included the definition of the concept of regional role, which can be considered an academic element, and intensive dialogue all through the process. The quote below is descriptive of the approach.

27.
“The first question concerning the evaluation was how to start internal discussion about regional impact. That’s what we used it for – to start a discussion about the University’s regional role. The other question was, how we could improve ourselves with the disciplinary profile that we have. It shot off a discussion on how this university should operate in this region, and raised the question about about how to improve our own processes.” (IA5)

Enlightenment is used more generally to refer to increased understanding or, indeed, to seeing something in a new light. It helped people to understand their own university better or “to see the whole for the first time” (D2). Others learned about the region (IA5), or understood the importance of co-operation with local stakeholders (e.g. FA5, FA7, D7, D8). Evaluation also acted as an awakener – some realised the importance of
self-evaluation or monitoring (D6), others the need to “go out” and to network with local and regional stakeholders (R1). What I call enlightenment was often described by words such as understand, notice, realise, perceive, and awake – all referring to increased understanding or learning. In the evaluation process people also learned about evaluation itself (IA1). Both conceptual utilisation as an attempt to come to terms with the concept of universities’ regional role and enlightening utilisation as a more individual type of learning can be seen as serving organisational learning.

Decisions are seldom made “just like that”. It was emphasised that the external report should not be read as if it was the Bible. Rather, the recommendations needed to be carefully considered before implementation (FA3). In practice this means that before actual decision-making takes place, dialogue is needed. Further, it refers to interactive use (Vedung 1997, 274). Interaction can relate to using different information sources, or it can relate to social interaction, or both. Interaction belongs to all phases of university evaluations – planning, self-evaluation, dissemination, and implementation. It need not be a disorderly back-and-forth process (Lampinen 1992, 34), but it can be used to support other types of evaluation use, especially to create shared understanding, in other words organisational learning, develop networks, and extend communication (Forss et al. 2002). Interaction seems especially important in dissemination. The need for and the major role played by interaction refer to communicative and transformative evaluation (Hämäläinen and Kauppi 2000) discussed above.

The different types of evaluation use are not organisation-wide, shared experiences. Rather evaluation use may be limited to a small group of key persons, or even to a few individuals. Enlightenment was often described as an individual experience. Considerable time may be needed before it becomes observable, and even then outsiders may not connect it to the evaluation (R7). This relates to the closeness of the concepts of utilisation and impact, which is evident especially in enlightening and conceptual use. One of the interviewees described the use of evaluation pertinently and pithily. He first spoke about the users of the evaluation, and referred to enlightening and legitimising use. Then (italics) he referred to symbolic use, “...we have been evaluated...”, and further to the strengths and weaknesses of the University, which can be read as general legitimation. Finally, “now we are improving this way”, reports instrumental use, which serves legitimising purposes.

28.
“Well, I think this is more a tool for use at the institutional level. Perhaps faculties get something out of it as well for their own evaluation – background and framework, but still I think that it’s more an instrument for the whole university – for instance in relation to other universities and partners, different ministries and local authorities, etc. We can show that we have been evaluated, and these are our strengths and here we have some challenges. And now we are improving this way.” (IA1)
When asked about the usefulness of the process and reports, an institutional administrator emphasised the importance of participation in the evaluation and the learning that occurred during the process. However, when describing evaluation use, he mainly described instrumental use based on the external evaluation report, and presented a list of the university’s reactions to the recommendations of the peer review team (IA4). Participation in the self-evaluation process seems to have been personally useful, while the various decisions reflect utilisation by the university – “actual utilisation”, as it seems.

As discussed above, the literature mainly assumes that utilisation is organisational, and that organisations learn. Individual users are discussed less. Interactive self-evaluation processes clearly promoted both organisational learning and the dissemination of evaluation. However, to create an interactive self-evaluation process, influential individuals (animateurs, champions) were needed.

**Simultaneous and Chain-like Utilisation for several Purposes**
The different types of utilising evaluation are not mutually exclusive. On the contrary, an evaluation or measures based on it can be used for several purposes, simultaneously or sequentially. In addition, people interpret the purpose in various ways, depending on their knowledge and expectations.

University evaluations were used for several purposes – simultaneously or in a chain-like way as described in the above quote (IA1, above). The most usual combination was one of instrumental and legitimising use. For example, as a consequence of evaluation, measures are taken to improve external engagement. When reporting to the Ministry of Education, the University states that it has followed the recommendations from the evaluation, and is thus improving its external engagement. That is rational use expected by the Ministry (M3), combined with positive legitimation. In addition, the message is sent to other external stakeholders, which can be considered either general or positive legitimation (Valovirta 2000, 78–79). Another usual combination is that of interactive and enlightening or conceptual use, discussed by Lampinen (1992, 34). However, interactive use, often in a supportive role, can relate to instrumental and legitimising use as well. The combinations suggest that it is indeed possible to use one evaluation for several purposes, in other words, to kill several birds with one stone. Decisions about major changes as a consequence of evaluation usually call for negotiation and consideration, that is, interactive use. Further, evaluative information is not the only information used in decision-making or in public relations activities, but it is mixed with other available information and knowledge. Thus, if the different types of utilisation are expected to refer to the purposes of utilisation, interaction as such should not be considered to be a type of its own, but rather something that enhances utilisation or is needed for it – or characterises it. An example of such interactive utilisation is the “arenas for discussion” mentioned by Saarinen (1995).
Conceptual Overlap

By conceptual overlap I mean that some of the types of evaluation use considered above actually refer to one phenomenon, or one type might relate to several different purposes. In practice they may be difficult to discern, as the user’s purpose is not necessarily visible to an outsider. An example is the term tactical use, which is used to refer to rational, political, and cultural use of evaluation. In another example, legitimising and persuasive both refer to the same purpose (legitimising is the term used e.g. by Valovirta 2000, 78–80, and persuasive by Shadish et al. 1991, 53). The conceptual overlap can be avoided by referring to the types of utilisation based on organisation theories (Albaek 1997, 371–397). When describing evaluation use based on organisation theories, a researcher does not ask whether the user of an evaluation has acceptable or less acceptable purposes, but rather considers the utilisation of an evaluation from different perspectives.

Considering the various types of utilisation discussed above, all of them can be recognised in the case study universities. The most usual types of evaluation were instrumental, conceptual / enlightening, and legitimising. The interactive utilisation considered by Vedung (1997, 274) is close to the argumentation considered by Valovirta (2002a), and to the extended communication referred to by Forss et al. (2002, 39). All the three can be characterised as dialogue, which in turn seems to be a sine qua non in the utilisation of an evaluation, and an essential element of a learning organisation.

5.3.4 Areas of Evaluation Use – where?

The question of ‘where’ refers to the basic tasks of universities, that is, to teaching and research, and university management and to the theme of the evaluations – the regional role or external impact of universities, also called the third task. At both universities the use of evaluation extended to all the above mentioned areas. As one would expect, most of it is somehow related to the third task, but in some cases the connection remains rather thin. I start from the utilisation of evaluation in external engagement, and proceed through management to teaching and research.

External Engagement and Regional Role – two Approaches

Before discussing evaluation use in the area of external engagement or regional role, the very theme needs to be considered. This is because there were different interpretations of it. Most of the interviewees at both universities thought of regional or external impacts. The impact approach appears also in the titles of the self-evaluation reports: “The Regional Role of the University of Kuopio” (in the Finnish language version, however, Kuopion yliopiston vaikuttavuus, i.e. Effectiveness of the University of Kuopio), and “The External Impact of the University of Turku”. The concept of both Peer Review Teams was broader. In addition to the impacts of a university’s
teaching and research, and its presence in a certain locality or region, the peer review teams were looking at how the universities operated in relation to their regions and stakeholders, and how they took into account the needs of the region. It was not only the impact that they were interested in, but also the mechanisms and processes of external engagement. The approach is obvious in the reports and recommendations of both peer review teams as well as in the titles of the reports: ‘Towards the Responsive University’ (Dahllöf et al. 1998), and ‘External Engagement and Institutional Adjustment’ (Goddard et al. 2000). The emphasis of the external panels on internal and boundary mechanisms or developmental peripheries (Clark 1998, 6) was somewhat surprising for the universities.

As no common self-evaluation questions or guidelines were available, each university had the right to define the evaluation theme, to tackle the questions it felt were relevant, and to organise the self-evaluation in a way that suited it best. Further, the approach gave the external panel the option to offer their interpretation about the evaluation theme. From the university perspective, it was challenging to opt for a theme that was new and understood in different ways. On the other hand, the approach offered an opportunity for dialogue and learning.

Among the various measures taken to improve external engagement perhaps the most visible are the new development organisations, established to bridge university research with local and regional needs, the new faculty at Kuopio, and the new degree programmes at both universities. The universities have strengthened their cooperation with regional stakeholders, and developed local networks (e.g. IA4, IA5). Most of the measures represent instrumental use at the institutional level. The universities have also put effort into disseminating the message about their regional role or external engagement, both to their staff and to external stakeholders and partners – locally, nationally, and internationally. However, mere dissemination of evaluation results is not enough to affect attitudes within universities or to establish arenas for discussion (D6, D7). If the idea of external engagement or regional impact does not reach teaching and research, external engagement remains a symbolic activity – discussions decoupled from the basic processes of the universities. Hence, when considering the acceptance of external engagement as a normal part of universities’ activities, one of the Rectors spoke about a need for a change of attitude (R2). This, in turn, pointed to a need for more interaction between the University and local stakeholders.

Management
Most Finnish university evaluations focused on university management and leadership. In evaluations with a special theme, such as those conducted at the Universities of Kuopio and Turku, management and leadership were considered from the perspective of the theme. The question was how the university management took care of the theme or task being evaluated, in this case external engagement. In addition, other more general questions of university management were considered. As described above,
university evaluations were used mainly by the management. Here evaluation use in management, that is, to improve management is considered.

When analysing the present situation and the future of the universities during the self-evaluation process, people at both universities became aware of a need for strategic thinking and management, and for stronger leadership. The process resulted in new perspectives and ideas for planning (IA4, D4, D7). Later, the external report gave additional impetus for the ideas.

The utilisation of evaluation in relation to various external stakeholders is obvious, but the use of evaluation in relation to the Ministry of Education was not mentioned at all without my prompting. Only rectors and key administrators were able to discuss the role of evaluation in management-by-results, others had only vague ideas. In the main, the interviewees did not think that university evaluations had a significant role in result-based governance (R7, R9). When asked whether the university had been able to utilise the evaluation in its relationship with the Ministry of Education, an institutional administrator answered:

29.

“To tell the truth, I don't think we have discussed it at all. And we haven't been able to make much use of it (evaluation). We should talk a good deal more with the Ministry about regional impact and about development and institutional evaluation and its utilisation. There has been very little discussion.” (IA5)

External engagement and regional role have appeared, however, on the agenda of the internal performance negotiations of the universities (IA2, CE2), which can be considered quite an effective way to disseminate the message about evaluation and especially about external engagement to faculties and departments. It can also be considered as a sign of organisational learning. The ignorance about the role of evaluation in Ministry – University relationships can indicate that university evaluation is considered an internal operation rather than an instrument for accountability. Since the evaluation both universities have received positive feedback from the Ministry of Education concerning the documents prepared for the annual negotiations (IA1, R1). This suggests that in the evaluation process the universities have, indeed, learned to understand themselves better.

**Teaching and Research**

As to evaluation use in the areas of teaching and research, both universities introduced new degree programmes and established new units or bodies to strengthen their external engagement. Further, they became involved in or established various boundary organisations. However, for those who were not involved in the process, the connection between the evaluation and the measures is not as clear. Where a new body or programme is established or a new procedure set up, people may or may not know
that it had come as a consequence of the evaluation, nor do they necessarily know what the rationale for it was. When, for example, the external panel recommended that the University of Turku establish a Teaching and Learning Council, which it did, the rationale was how to build a learning region and how to assure the quality of teaching and learning across all programmes (Goddard et al. 2000, 19, 23). The former, building a learning region, was not mentioned by any of the interviewees, even though several of them mentioned the establishment of the Council. Thus, from the external engagement point of view, the establishment of the Council seemed to be symbolic use of evaluation. However, there was also a different interpretation of the reasons for recommending and establishing the Council:

30.
“The Teaching and Learning Council is something that really doesn’t relate to external impact in any way, but rather to this general evaluation conducted as part of the same process. The international panel saw that the University hadn’t put enough resources into developing teaching and considering the problems encountered in teaching, or on establishing support services to improve the quality of teaching. Then, to some extent the relation to external impact comes in when planning curricula. How we take into account the developments outside the university, in addition to the development in scholarship itself.” (D2)

Likewise, at Kuopio the connection between the evaluation and the new faculty and new degree programmes was not clear for those who were not directly involved in the evaluation (FA7). The ignorance about the connection between the evaluation theme, and the measures taken in teaching and research clearly indicate that the message about external engagement has not reached the basic units – at least not yet.

**The four Areas of Utilisation**

Considering the four areas of utilisation – external engagement, management, teaching, and research – external engagement has obviously received more attention than the other three. However, even though it is called the third task, universities’ external engagement is not meant to be separate from the other tasks. Rather, the expectation is that it should play a role in all the universities’ functions. Utilisation of evaluation in the area of external engagement includes measures through which external engagement is expected to become inculcated into the basic processes of the universities, that is, teaching and research, and management. In addition, the message about THE regional role / external engagement has been disseminated both internally and externally. However, the change discussed – accepting external engagement as one of the case study universities’ tasks, and implementing related measures – is so demanding that major changes in attitudes should not be expected too soon. The time span needed actually refers to the impacts of evaluation rather than to utilisation alone.
5.3.5 Obstacles to and Facilitators for Evaluation Use - why?

As to why the evaluation was used, there were obstacles to, and facilitators for, evaluation use in each of the four parts of the conceptual framework – evaluation, internal and external context, and users. Out of the four, evaluation is the only one that can be tuned; context and users can only be taken into account.

The most important facilitating features of the evaluations discussed in the interviews were the theme and timing of the evaluation, and the expertise of the Peer Review Teams, that is, the perceived relevance of the evaluation. Also, the self-evaluation process was considered to be an important facilitator that in addition to organisational learning, served also dissemination, and implementation. As to users, the evaluation was in the first place a management instrument. From the perspective of wide participation and effective dissemination this can be criticised. However, it is in line with one of the purposes of university evaluations, namely to strengthen universities at the institutional level. In dissemination in particular, the role of the so-called ‘right persons’ – animateurs and gate keepers – was considered important. Further, well functioning teams were an asset in dissemination and implementation. The need for selling and marketing the evaluation also refers to dissemination. The character of university organisations was an obstacle rather than a facilitator for evaluation use. However, earlier experience in evaluation and an interactive approach helped to overcome the problems posed by loose coupling and individual pursuit of glory. External context was directly referred to only when considering the timeliness of the evaluation. In the background there may, however, be the above mentioned trust that is typical of the Nordic countries.

In the following, the question why (evaluation was used or not) is first considered from the perspective of acceptance, dissemination and utilisation, and then from that of the Four I’s of Weiss.

For rectors and others in management positions, the theme and timing of the evaluation were easy to accept as they had chosen them. In the departments the situation was different; people did not know enough about the evaluation to get interested. It was suggested that with some “selling and marketing” the situation might have been better. As regards dissemination, the size and internal culture of the universities played an important role. As to utilisation, it was mainly successful as far as instrumental and legitimising use by university management was considered. The question as to whether external engagement is taken into account in teaching and research does not relate to utilisation alone, but also to dissemination. Further, it can be seen as an impact question, due to the time needed to effect real changes in academic matters. In dissemination and in implementation of the recommended changes, the universities’ capacity to use evaluation or capacity for action was critical. (Vedung 1997, 285; Bauer et al. 1999, 35)
Considering the steps from acceptance through dissemination to utilisation and organisational learning, in the case study universities the easiest was acceptance, for reasons discussed above. While acceptance depends mainly on the relevance and legitimacy of evaluation, successful dissemination and utilisation of evaluation relate in the first place to internal context and users. It was easier for universities to accept evaluation, but more difficult to disseminate results, or to implement the recommendations of evaluation. However, the universities were able to implement a good part of the structural and other changes recommended by the external panels. Thus dissemination seems to be the most difficult part of evaluation.

Each of Weiss’ Four I’s considered above can act either as a facilitator or as an obstacle to evaluation use. Starting from ideology, although the values of the evaluations had only a minor role in the interviews, a conflict between traditional academic values and the utilitarian values of the evaluations was referred to. The value conflict became visible (and audible) when people spoke about the danger of abandoning basic research, moving towards a polytechnic profile, and resisted a quality system or were doubtful about it. Further, the conflict can be seen in the questions related to who was to control the change (IA5), and how vigorously the university should pursue change (R1). However, such reactions were few in number. This could be due to the fact that people did not know enough about the evaluation to criticise it, or that they did not care. That is, they lacked information or interest. As the theme of the evaluation was chosen by university management, a value conflict between management values and university management was less probable than conflict between management values and academics. In other words, who ever chooses the approach and hence the values, tends to choose values that are not in conflict with his/her own values. Thus, in this respect, had interviews been conducted in the departments, results would probably have been different.

As the whole university as an organisation is the focus of a university evaluation, it is obvious that those who are responsible for the whole organisation – rectors and other institutional-level personnel – have the biggest interest in it. Their interest was strengthened by the fact that they had chosen the theme of the evaluation. One might speak of “owner’s interest”, as expected by FINHEEC. Further, they had the best information about it, and had no ideological conflict with the evaluation. Academic staff are usually more interested in the evaluation of research and teaching, because they have a stake in it (D3, FA1, FA4, FA5, D6). However, in both case study universities there were deans and department heads, and perhaps other academics, who were interested in the evaluation and active in its utilisation. They were described as interested in the common good and as people who think ‘outside the box’ – and as “right persons” (R1, IA7, D6, P1). However, interest need not be in accordance with the aims and values of the evaluation; it may as well be in someone’s interest to avoid extra work or to resist change. Further, lack of interest can relate to lack of information. Thus, interest has an effect on who the users are, but it also affects the purpose.
of evaluation use, that is, how evaluation is used. For example, it is in the interest of rectors to use the evaluation to strengthen the university’s position and image and for developing it, that is, for legitimising and instrumental purposes, while others may be more interested in learning about the university and in networking.

Expertise and experience in evaluation and familiarity with the evaluation theme represent information in the sense of resources or capacity. It was easier for those who had such information to get involved in the evaluation. Another type of information is that collected and interpreted during the evaluation process and published in the reports. In dissemination, the interactive way was more effective than top-down information flow. Another effective way of dissemination was the self-evaluation process (cf. PLS Consult). In the process, the participants learned about the university and about evaluation, and their interest increased. Further, it was easier for those involved in self-evaluation to understand the evaluation reports, as they knew the background better than others. Thus, at its best, the self-evaluation process can be a source of both information and interest. Possibilities for conceptual or enlightening use were clearly less for those who were not involved in it. In Feinstein’s terms, there were problems in both presentation and means (Feinstein 2002, 436). In other words, the message was not clear enough, and the information channels could have been better.

Out of the Four I’s, institution was the most difficult part in the utilisation of university evaluation. Even though the evaluation was accepted in the main, it was difficult to extend it and its message to the department level. Similar results from various institutional evaluations have been reported by Dill (1998, 2000), Nilsson and Wahlén (1999), Harvey and Askling (2003), and Vartiainen (2004, 136). In contrast with the above Välimaa et al. (1998) report department-level changes as a consequence of university evaluation. Based on the above discussion concerning two different self-evaluation processes, the decisive difference seems to be in ‘the interactive model of self-evaluation’ – or in dialogue – used at the Universities of Jyväskylä and Kuopio (Sallinen et al. 1994; Kinnunen et al. 1998a, and the above description of the self-evaluation in the University of Kuopio).

The institution itself was also the most important of the Four I’s because it had an influence on the three other I’s – ideology, information, and interest as Weiss points out. The loose coupling and individualistic character of universities do not support joint efforts, and the bigger an institution, the more difficult such efforts become. An institution has an influence on who has enough information based on their position in a hierarchical system, which in turn can have an effect on interest, and on one’s options for using evaluation. Perhaps somewhat surprisingly there was no major conflict between academic and utilitarian values. Rather the conflict acted as an intellectual challenge, more easily coped with than the institution itself. In spite of the similarities, there were also major differences between the two universities in their utilisation of the evaluation. Earlier experience of institution-wide evaluations, which can be called evaluation capacity, was clearly an advantage.
5.3.6 Evaluation

Although the themes of the two evaluations – regional role and external impact – raised questions and even caused some embarrassment at both universities, no criticism was presented against either. Understandably, the theme and timing of the evaluations, the very issues that the universities could decide upon were considered to be facilitators rather than obstacles. They were especially appreciated by institutional management, that is, by those who had chosen them. This relates to the relevance of evaluation discussed by Feinstein (2002). Institutional management also took ownership of the process in the sense expected by FINHEEC. Institutional management at both universities saw an opportunity to demonstrate their performance in the evaluation, and expected an external expert opinion on their impact. This suggests legitimising utilisation of the evaluation. Further, the task of defining or interpreting the concept of regional role or external impact was quite challenging (R3).

26. “That we chose regional impact, I presumed from the outset that we were taking the most difficult task to be considered by universities. I admit that though that we could not lose because it was a difficult task, so it was worth doing, for sure, exactly because of the difficulty. We'd get positive feedback for doing it, acceptance in the university without losing anything. Pioneer work is pioneer work; it develops thinking.” (R3)

The managerial and utilitarian values (ideology) of the evaluations were not directly criticised, but some doubts were expressed about them at both universities. The doubts were visible in the fear of a polytechnic profile, in the doubts concerning far going systematisation, and in the question “should we go out as strongly as that?” The absence of strong criticism can at least partly be explained by the position of the interviewees: most of them were in management positions where managerial values are probably more easily accepted than in other parts of the universities.

The expertise of the peer review team, representing the credibility of evaluation and thus a part of its relevance, was appreciated at both universities. The general esteem in which the team members were held gave additional weight to the recommendations, and to the whole report, which was described as an “authorised opinion” (D1). Further, the interviews by the peer review team, or “discussions with esteemed experts”, offered challenges and new routes for thinking (D1, D6). The challenge offered by the theme was strengthened and further developed in the interviews by the peer review teams. The discussion continued in internal seminars (IA5, R7).

FINHEEC suggested that the ownership of the evaluation process should rest with the evaluated university. This idea was accepted by those involved in deciding about the theme and timing of the evaluation, and in proposing members to the peer
review teams. There was also ownership of evaluation in those faculties that saw an opportunity in the evaluation and were able to make use of it.

Opinions about the use of an international panel and hence a foreign language were split. Most of the interviewees either had not noticed any major difficulties or they thought that it simply had to be accepted. The reasons for this acceptance were first, that Finnish teams would not be perceived as being impartial or credible, second, that “we need to be international”, and third, because “we are part of an academic community” (R1, R6, IA3). However, there were also those who thought that having foreign language skills was a requirement for reading reports, participating in the interviews, ensuring that people felt some commitment to the evaluation, and utilising the evaluation (IA6, IA2). Further, the language had to be taken into account when inviting interviewees. The cultural difference, visible in recommendations that do not fit the Finnish context, was considered problematic only by some (R1, D2, FA3). Thus, the language of evaluation relates to both its acceptance and dissemination.

The follow-up evaluation, also called the re-evaluation, gave additional impetus to the implementation by offering a check-point and continuity. Further, it helped those who were not involved in the first evaluation to catch up, and join in the discussion (D3, D7). Thus it worked as an incentive, and assisted in the dissemination of the message about external engagement.

As there were no concrete sanctions or rewards (such as funding or accreditation) as an outcome of the university evaluations, it is difficult to say whether the existence of these would have lead to better utilisation. This was suggested by one of the interviewees (P1), while some authors (including Leeuw 2001a, Williams 2001, and Stensaker 2000, 308) assume that sanctions tend to lead to symbolic use. At the institutional level, the ownership of the evaluation process offered by FINHEEC, and the consequent opportunity for universities to demonstrate their improved performance and to convince external stakeholders of their willingness to increase their external engagement probably worked as incentives. At Kuopio the forthcoming EU-funding period was an additional incentive.

In conclusion, the perceived relevance of the evaluation was assured by allowing the universities to decide on the theme and the timing of the evaluation. As they were also entitled to propose panel members who had been characterised as esteemed experts, both elements of relevance of the evaluation, namely timeliness and credibility were present (Feinstein 2002). Further, interviewees expressed no doubts about the legitimacy of the evaluation, this being important in the evaluation of higher education (Vartiainen 2004, 115–136). Hence, the evaluation was easy to accept, at least for those who had “invented the theme” as one the Rectors put it (R7). As to other interviewees, no open resistance was referred to, but there was some indifference. Acceptance, the first step on way to utilisation, was easy. Dissemination and implementation of the evaluation were more difficult.
The experience of relevance as a factor that enhances evaluation use is familiar at other universities as well. “At the University of Oulu, those recommendations made by the external panel which were easiest to accept were those of which we had some pre-understanding, or had perhaps even been the subject of some tentative planning”. (Aaltonen 1999, 5)

5.3.7 External Context

The most important of the external trends that facilitated evaluation use were the appearance on the national higher education policy agenda of the regional role or third task, and in the case of the University of Kuopio, their expectation of gaining EU Objective 1 Programme funding. As the recommendations of the Peer Review Teams were in accordance with the external trends, they were easy to accept. At Kuopio, an additional incentive came from the expectations of local stakeholders (Feinstein 2002, 436). Due to their similarity, it was difficult to tell whether it was the external pressure or the recommendations which were the more important (R1).

Like evaluation, external context also relates in the first place to acceptance, particularly acceptance of the legitimacy and relevance of evaluations. The external trends and emerging ideas that were mentioned in the interviews related mainly to relevance, and especially to timing. Trust as a feature of the national context or evaluation was not discussed in the interviews, but there were no signs of mistrust either, except perhaps the comments from a Ministry of Education interviewee, which suggested that the veiled language used in the external panels’ reports might provide relief to the universities as nothing needed to be done (M3; cf. Leeuw 2001b).

5.4 Learning Organisation and Organisational Learning as Part of the Utilisation of University Evaluation

Above, when discussing the concepts of the learning organisation and of organisational learning and their relationship with the utilisation of university evaluations, the following two questions were raised (Chapter 4.7):

1) Did university evaluation lead to organisational learning in the case study universities, and

2) did the possibly existing elements of a learning organisation support learning and the utilisation of evaluation?
Two questions in the interviews were directly related to learning from evaluation: Did you learn something during or following the evaluation, and did the evaluation provide new knowledge about departments, faculties or universities? Several interviewees characterised the university evaluation as an excellent learning opportunity and as a learning process (e.g. R2, R3, R8, R9, R19, D1, D2, IA4, IA6).

Who learned and how? One answer to this question was given above: those involved in the evaluation process, in one way or another, learned most. The deeper their involvement, the more possibilities they had for learning. However, there were also those who learned from the reports, and obviously others who used both the process and the reports. The most important source of learning was other staff members and the peer review teams, not reports or other documents. The peers, characterised as esteemed academics, were said to challenge one's own truths and to provide new perspectives both in interviews and in their reports. Thus, the evaluation can be described as an arena for dialogue as suggested by Saarinen (1995). The more people were involved in various discussions concerning the evaluation, the more were there possibilities for dialogue and organisational learning. However, another group who learned a lot may have been the evaluators as Forss et al. argue (2002).

What was learned? The interviewees said that they learned about the university, about their own department and about the regional role of universities. It also helped to understand the importance of the regional role of universities and of cooperation with stakeholders. Not so surprisingly, people also learned about evaluation. Further, during the self-evaluation many noticed the achievements and strengths of their own department, which was a rewarding experience. A different lesson was the understanding that the university needs to learn about marketing, and another that leadership training is needed.

The evaluation was a learning experience for many staff members, but – to put this in Argyris' words – was there during the two institutional evaluations a sufficient interplay between individuals and the different “entities defined at relatively high levels of social aggregation” to speak about organisational learning? Or were there rather individuals who learned “on behalf of the organisation”? (Argyris 2001, 4–9). The role of individuals in evaluation was discussed above. It seemed especially important at the University of Turku due to its obviously loose coupling, but individuals at the University of Kuopio had a role as well. The “right persons” and the “animateurs” or “champions” discussed above can be seen as those who took care of “the inquiry and of the intertwining of thought and action carried out by individuals in interaction with one another on behalf of the organisation to which they belong in ways that change the organisation’s theories of action and become embedded in organisational artifacts” (Argyris 2001, 9). From the evaluation point of view, active individuals – be they called champions, right persons, “those who carry the message”, or animateurs – were needed to encourage others to get involved, to take care of dissemination, and
to represent continuity. From the perspective of organisational learning they were needed to act as a link between individual and organisational learning.

In the following, the utilisation of university evaluations is considered from the perspective of a learning organisation, applying the five elements of an academic learning organisation defined by Dill (1999):

1) culture of evidence
2) improved coordination of teaching units
3) learning from others
4) university-wide coordination of “learning”
5) transferring knowledge

There were several signs of a culture of evidence in the case study universities, such as self-initiated evaluations, criteria developed for evaluating research, a system for allocating resources, a systematic survey of graduates at the University of Kuopio, and the extensive surveys carried out among a range of stakeholders for the institutional evaluation at the University of Turku. In addition, the register of the contacts of staff members with local and regional actors in one faculty established during the self-evaluation can be considered as indicating the existence of a culture of evidence.

As to improved coordination of teaching units, it was difficult to find signs of it in the case study universities. However, the institutional approach was not the best possible tool for finding out about possible department-level measures in this respect.

People in universities learn from others all the time and seek new knowledge from colleagues at other universities, but it is not systematic, nor does it usually become organisational. However, when preparing the university evaluations, both case study universities made use of the reports of the institutional evaluations carried out in other Finnish universities in order to learn from others.

At the time of the evaluation a teaching council – a body responsible for the university-wide coordination of learning – had been in existence for a good while at the University of Kuopio. For the University of Turku the peer review team recommended that a Learning and Teaching Council be established. The team further recommended that each faculty establish a similar body with representatives from all its departments (Goddard, Teichler et al. 2003). Thus, they suggested the coordination of teaching and learning at both the institutional and the faculty levels, in accordance with Dill’s points 2 and 4.

Some mechanisms suitable for transferring knowledge existed at both universities: the regular meetings of deans at Kuopio, and the team of academic leaders which was in existence during the evaluation at Turku. The Faculty of Education at Turku built the self-evaluation process on dialogue in order to combine faculty staff members’ expertise in external engagement. Obviously, other regular meetings such as those of the councils at university, faculty, and department levels can also be used for knowledge
transfer. At the University of Kuopio the tradition of university-wide interdisciplinary projects supports knowledge transfer.

At the national level there are some arrangements in Finland that serve knowledge transfer between universities, and universities’ learning from others. Examples include the various national seminars about e.g. quality assurance and student admissions, arranged by FINHEEC or by the Ministry of Education. Another example of knowledge transfer is the nomination of high quality units in teaching by the Ministry of Education every three years. As an outcome of the process the descriptions of the best cases were published both as a book and on the FINHEEC website (www.kka.fi), which offers an opportunity for benchmarking. (Hämäläinen and Moitus 1999) Further, those responsible for each university’s quality assurance system, meet regularly.

From the perspective of the five elements of an academic learning organisation, both case study universities showed a lack of coordination of teaching units and the University of Turku lacked the pan-university level coordination of learning.

The other three elements were different at these universities. Due to its small size, originally unusual structure, and interactive culture, the University of Kuopio had an advantage and it could be described as being well on its way towards being an academic learning organisation. The University of Turku, on the other hand, is a rather large, “old, traditional seat of learning with an air of conservatism” (Puukka 2000), and with very self-sufficient departments (“walls between departments are made of concrete”; P2). It did not have structures to support organisational learning and to make a learning organisation. However, there can be departments or faculties within a university that are closer to a learning organisation than the whole university. The Faculty of Education at the University of Turku is a good indication of this. Thus, there seems to be a need for more of those who are able to combine individual learning and the learning of different groupings within the university.

Above, when discussing the concepts of organisational learning and learning organisation and their relationship with the utilisation of university evaluations, the following two questions were raised (Chapter 4.7):

1) did university evaluation lead to organisational learning in the case study universities, and

2) did the possibly existing elements of a learning organisation support / enhance organisational learning and the utilisation of evaluation?

The qualified answer to the first questions is yes. Several interviewees thought that those who were active in the self-evaluation probably learned more than others. Thus, the learning that took place did not reach all university staff right away. Rather the “animateurs” or the “right persons” discussed above had an important role in both universities in taking the message about evaluation to those corners of the university.
where evaluation had been conducted rather superficially or where its utilisation seemed to be ritual or nonexistent.

The answer to the second question is yes, the elements of a learning organisation did support both individual and organisational learning, and the utilisation of evaluation. However, none of the case study universities had all the five elements of an academic learning organisation in place. What could be found – at least in some extent – was culture of evidence, learning from others, and transferring knowledge. What was missing were the improved coordination of teaching units, and, at the University of Turku, the university-wide coordination of learning, both recommended by the peer review teams. At the University of Kuopio its culture of evidence, based on earlier evaluations, played a role all through the evaluation – from planning, through self-evaluation, and external evaluation to implementing changes recommended by the peer review team. The small size and interactive culture supported both knowledge transfer and learning – i.e. the intention to create a shared understanding about the concept of regional role.

5.5 Conclusion – the Whole Picture of Evaluation Use at the Case Study Universities

At the case study universities, who and what were critical as to how and where institutional evaluation was used (that is, the users and the part of the evaluation that was used). As regards who, rectors and institutional administrators as process owners showed the greatest propensity towards utilising the evaluation for all the above discussed purposes, and in all the four areas of teaching, research, management and external engagement. However, irrespective of position, those who were not involved in the self-evaluation process did not describe enlightening or conceptual utilisation. Those involved in the process used it either for enlightening or conceptual purposes, i.e. for learning, or – at the department level – for ritual purposes. The lower down the hierarchy the user was, the more probable was ritual use. Instrumental and legitimising utilisation were management tools. Instrumental utilisation occurred at both institutional and in some cases at faculty level. Legitimising utilisation was almost exclusively prevalent at institutional level. When symbolic utilisation was referred to, it was an observer’s interpretation of how rectors or institutional administrators used the evaluation.

To put it briefly,

1) high (academic) position combined with process participation offered the best possibilities for utilising the evaluation in different ways,
2) instrumental (rational) and legitimising (political) utilisation were typical of management,

3) process participation offered a possibility for conceptual and enlightening use and for organisational learning,

4) symbolic utilisation was seen as related to institutional management, and

5) the lower down the hierarchy a process participant was, the more probable was ritual utilisation.

Considering the five points above, some comments are required. First, it is important to emphasise that even though those responsible for management (rectors, deans, institutional level administrators) had access to all available information concerning the evaluation, the possibility to learn from the evaluation was weakened if they did not actively participate in the evaluation process.

As to the second point, it was in the management's interest to use the evaluation especially for instrumental and legitimising purposes, as management bears the main responsibility for the university's development and image. They also have the authority to take any measures that are needed. Symbolic utilisation was not often referred to, but there was an assumption that some of the instrumental or legitimising measures may have served symbolic purposes as well. The fact that some interviewees interpreted rational or political utilisation as symbolic and expressed disapproval, indicates, as Albaek (1997) suggests, that symbolic utilisation is considered less acceptable than instrumental or legitimising use. As to ritual utilisation, it refers in the first place to the secretarial or clerical work needed in various parts of the universities for the self-evaluation reports. In those cases there is reason to ask whether ritual use can be considered conscious utilisation at all.

As regards participation in the evaluation process, i.e. participating in various discussions and debates or in writing a report, they were an essential factor from the perspectives of enlightening or conceptual use and organisational learning. Existing elements of a learning organisation also supported both individual and organisational learning.

In conclusion, there were four critical factors in the utilisation of university evaluation at the case study universities. They were institution, individuals, interaction (dialogue), and capacity to use evaluation, which can also be called capacity for change.
This is a qualitative multiple case study that aims to describe the utilisation of university evaluation at two Finnish universities. Due to the context dependency of evaluation, the phenomenon under study – utilisation of evaluation – is closely intertwined with both its internal and external contexts (Yin 1994, 13). The research material consisted of 40 interviews: staff from the case study universities (32), rectors of other Finnish universities (3), members of external panels (2), and representatives of the Ministry of Education (3). In addition, the universities’ self-evaluation reports, and reports of the external panels were used. Thus, triangulation was applied. The study has resulted in two university-specific descriptions and a consideration of common features between the two, a design that in Yin’s terms is an embedded multiple-case analysis (1994, 39). The reliability of the study is dependent on the quality of the case study descriptions, which in turn depended on the author’s interviewing skills and understanding of the research objective (Miles and Huberman 1994, 38).

If the utilisation of university evaluations were to be studied from the perspective of university departments, several different pictures would emerge. At one end there would probably be a department in which people had not heard much about the evaluation, where most knew nothing about it. Perhaps only the secretary would remember that there was a questionnaire which she or he filled in some time ago, the same as any other questionnaire sent to the department by the administration. The department head might remember that someone had spoken about an evaluation on external engagement, but that she or he hadn’t paid much attention to it, as the department already had good contacts with local businesses. Why would things be made more complicated than they were? At the other end of the spectrum there could be the department referred to earlier, where the theme raised the question about the link between the programmes offered by the department and the labour market relevance of teaching in those programmes, and the consequent discussion about the evaluation theme (R6). The theme would have been of interest, as it was closely related to the research interests of the department, and the graduates of the department were mainly employed by local organisations.

The criteria usually applied when evaluating the reliability of qualitative research are credibility, transferability, auditability, and confirmability (Miles and Huberman 1994,
Auditability is also referred to as dependability, and confirmability as objectivity of the instrument. These terms are interchangeable with those used in quantitative research, namely internal and external validity, reliability, and objectivity (Saari 2002, 229). To the criteria of qualitative research Miles and Huberman add a fifth criterion, namely utilisation / application / action orientation. The mentioned criteria are, however, not mutually exclusive.

Credibility
Credibility is good if the reconstructions presented by a researcher correspond to the original constructions of those studied (Saari 2002, 229; Eskola and Suoranta 2000, 211). In this study it would mean asking the interviewees whether they could recognise their own university in the description. That question however, was not presented to interviewees. Instead, I applied other means to strengthen credibility. These were rich and extensive research material, triangulation, and familiarity with the object of study. The research material consisted of interviews, self-evaluation reports and reports of external panels prepared for the university evaluations considered. As regards knowledge about university evaluations, my experience of them in different roles provided an excellent basis for understanding the evaluations and their utilisation. As mentioned above, I was a member of the steering groups of both evaluations considered here. In practice this means that I was in contact with the universities from the early planning phase until the concluding seminars, and again at the time of the follow-up or re-evaluations. In all, this covered three to four years, which according to the literature means that my contact with the universities can be considered long and intensive enough from the point of view of credibility. This experience provided an internal perspective of the universities. A third element that adds to the credibility of the study is the interview quotations used to support my interpretations. These quotations also strengthen the auditability of the study.

As discussed under “Research Design”, personal involvement in the evaluations was an asset, but might also have been seen as problem, if pre-assumptions and expectations had been created because of my role as a FINHEEC advisor rather than being in a position where it was possible to comprehend the reality faced by the universities. However, the experience gained as the internal coordinator of an earlier university evaluation allowed what might be characterised as a discussion between the roles of coordinator and researcher, and it helped me to recognise different pre-assumptions about the case study universities, evaluation and its utilisation (Eskola and Suoranta 2000, 17). Another critical point in the study was the interviewees. As they mainly represented university and faculty management, the voice of departments was not very loud. Interviewing people from the departments would have resulted in a more inclusive description of ‘reality’.
Transferability

Above I described the evaluation method, and the internal and external contexts in which the evaluations were used, to help the reader understand the two case studies (Stake 1995, xi). The aim of the study was to describe and understand the utilisation of university evaluation by two Finnish universities. The description revealed two discrete ways of using university evaluation, but there were common features as well, which were discussed above. The common features suggest, first, the possibility of analytical generalisation in the sense of propositions (Yin 1994, 9–11, 30–32, 36). Second, they suggest the possibility for ‘petite generalisations’, that is, situations where refinement of understanding can be reached through the case study (Stake 1995, 7–8), or extrapolation instead of generalisation, which means “thinking about what specific findings mean for other situations” (Patton 1997, 258).

To test the transferability of the results, the rectors of three other Finnish universities were interviewed. As to the size of those universities, one is smaller than the case study universities, and the two others are between them, in terms of enrolment numbers. Before the interviews, a tentative description of the use of institutional evaluation in the case universities was sent to the rectors (Appendix 3).

All the three rectors recognised the general description. One rector said the situation described was recognisable (R9), while the two others felt that it reflected the patterns they had seen in their own universities (R8, R10). An important difference related to the size of the institutions. In smaller universities it is easier to involve a higher proportion of staff in the evaluation than in the larger ones. Further, as there is no faculty level in the smallest universities, the dissemination of evaluation results is easier than in bigger universities. The rectors agreed that the evaluation was in the first place a management instrument. The weak interaction was easily recognisable in the largest of the three universities. According to the rector it related to the independence of faculties and departments (R9) – that is to the loose coupling within the organisation. The same holds for the random character of utilisation, in other words the dependence of utilisation on individuals. More than was the case with most other interviewees, the three rectors emphasised the learning that takes place during an evaluation, especially in the self-evaluation process, and the challenge included in the evaluation. The challenge was more relevant in evaluation that leads to accreditation.

As regards the tentative description, an important feature was missing. At the time of writing, I had not noticed the central role played by individuals, even though I did see the random nature of utilisation.

Auditability (dependability)

In qualitative research, auditability is the counterpart of reliability in quantitative research. When discussing auditability, Miles and Huberman (1994, 278) refer to “quality control”. The auditability of this study has been based on tape-recording and
typing of all the interviews, and on the description of the conduct of the study and the analysis. The interview quotations serve the requirement of providing auditability. Had someone else carried out the interviews using the same themes as I did, the detail of the result might have been somewhat different, as two different interviewers will rarely agree in their understanding of the message given by a third person. Also the time that passed since the evaluation might have caused opinions and understanding about its utilisation to change. Thus, re-interviewing the same persons later would probably bring about somewhat different views about how the evaluation was utilised (Hirsjärvi and Hurme 2001, 184–190).

**Confirmability**

When discussing confirmability, Miles and Huberman (1994, 278) refer to relative neutrality and reasonable freedom from unacknowledged researcher biases, and explicitness about the inevitable biases that exist. Saari (2002, 231) uses the expression *objectivity of the instrument*. Eskola and Suoranta consider confirmability from a different perspective. According to them, confirmability means that the interpretations made by a researcher get confirmation from other studies on the same phenomenon (2000, 212).

My personal involvement in the university evaluations and the conduct of the study was discussed above. The direct voice of the interviewees was presented in the quotations all through the report (Miles and Huberman 1994, 278).
Discussion: Major Changes ahead
– Capacity for Change needed

This report started with a discussion about changes in the higher education scene. It ends with another change. The ‘steering-by-results’ era will soon be 20 years old, and a range of evaluations have been conducted in Finland and in other European countries. However, the role of evaluation in the government control of Finnish universities is still somewhat unclear. As was the case 15–20 years ago, the nature of the changes that Finnish universities now face have been “imported” in the sense that they are not typically Finnish, but rather relate to developments in the broader European higher education sphere, and to globalisation, marketisation, and the demands for higher rates of productivity.

Recently, the Finnish government introduced a productivity programme to enhance the competitiveness of the public sector, including higher education institutions. One of the points made is that Finland has too many small higher education institutions, with insufficient critical mass. This indicated the possibility of institutional mergers and other structural arrangements now being planned and implemented. Another requirement is that universities should recruit more foreign students, and establish international joint degrees with partner institutions – that is, to become more international. For the moment, tuition fees are not charged by Finnish higher education institutions. However, there has been discussion about introducing fees for students who come from outside the European Union. Further, the employer union has suggested that if Finnish students want to get two degrees, they should pay for the second one. So far, however, the legislation prohibits the collection of tuition fees, but the Ministry of Education is preparing to change it in such a way that Finnish higher education institutions would be allowed to offer higher education programmes to student groups – not individual students – coming from outside the European Union and the European Economic Area, and to charge fees.

All the above requirements call for a range of changes within universities. Planning joint degrees and institutional mergers takes time, and no matter how smoothly it is done, there will be tensions between different national regulations and organisation cultures. There will also be a need for understanding of both the partner institution and one’s own organisation, and a willingness to cooperate and preparedness to change. This was clearly indicated by an ENQA project aiming to trace various factors in the participating countries that either promote or obstruct convergence, and
How are University Evaluations used?

by the Nordic project on comparing quality assurance in four Nordic universities (Crozier et al. 2005, Omar et al. 2007 respectively). Even though there was good will and understanding about the aims and the most important practical points in both projects, for historical and cultural reasons and because of different national regulations it took quite some time to reach a shared understanding about what needs to be done and how. Such endeavours tend to be more complicated and time consuming than usually expected.

As suggested at the beginning of the study, evaluation of higher education in general and hence the Finnish university evaluations in particular can be characterised as the intersection of various evaluation approaches. There were obvious audit elements in them, but there were also characteristics that refer to realistic evaluation, stakeholder evaluation, and even empowering evaluation. However, if the university evaluations were to be connected to one type of evaluation, it is the pragmatic, utilisation-focused evaluation. The elements of realistic and empowering evaluation, such as the idea of ownership of the evaluation process, were there more or less to ensure the utilisation of the evaluations. However, as work in the above chapters indicated, those elements did not guarantee an even spread of the utilisation of evaluation in Finland’s universities. Rather the evaluation was a management instrument – in accordance with the original aim of strengthening the institutional level and capacity for change of Finnish universities.

The second round of institutional evaluations of Finnish universities is being implemented. As a reaction to the Berlin Communiqué’s requirement of accreditation or comparable procedures, FINHEEC has introduced a programme of auditing of quality assurance mechanisms at all Finnish higher education institutions. It means a more structured approach than was the case in the university evaluations. In addition to being a reaction to the Berlin Communiqué, the audit approach as such can be considered to be a sign of diminishing trust or of “rituals of verification” – a term used by Michael Power (1999). One might ask whether the government no longer trusts universities as much as in the past, or whether it no longer trusts the control mechanisms established by the Ministry itself. However, more probably the answer is the obedience with which Finland follows the demands of the Bologna process. In general, the government’s grip on universities is becoming stronger and more detailed. An example of that grip was presented in the report of the evaluation of administration at the University of Helsinki (Hallinnosta hallintaan, 2005, 58–60). There are dozens of different laws, decrees, and other regulations that guide the Finnish universities’ daily work.

Balancing between improvement on the one hand, and control (transparency, accountability) on the other hand, is demanding. If Finnish higher education institutions are able to turn an obligation, that is the audit, into an opportunity, and take ownership of it, the audits will serve the aim of improvement – and at their best also organisational learning. So far, however, there is little evidence of the capacity of vari-
ous external evaluations or quality audits targeted to the institutional level to influence the quality of teaching and learning. The numerous simultaneous demands discussed above may lead to legitimising, symbolic and ritual use. When the power related to evaluation moves from higher education institutions to FINHEEC or to various auditing or accrediting agencies, to the Ministry of Education, or to other national, international or supra-national organisations, depending on European developments, the autonomy of universities in evaluation decreases. The developments boil down to the questions whose quality and quality for whom.

Briefly, there will be demand for both capacity for change and management of change in Finnish universities.
References

Literature


Alasuutari, P. 1999. Laadullinen tutkimus (Qualitative Research) 3. uudistettu painos. Tampere: Vastapaino


Evaluation of the Faculty of Science 1994, University of Helsinki. Helsinki: Yliopistopaino

Evaluation of higher education – the first four years. NA. Finnish Higher Education Evaluation Council. Vaasa: Ykkös Offset


KOTA-työryhmän muistio 1985 (Memorandum of the Committee on Evaluation of the Performance of Higher Education Institutions), Helsinki: Ministry of Education 26


Ministry of Education. 2004:6. Korkeakoulutuksen laadunvarmistus – työryhmän muistio (Memorandum of the working group on quality assurance of higher education institutions)


Reponen, T. 1999. Is leadership possible at loosely coupled organizations such as universities? In: Higher Education Policy 12, 237–244


Stenqvist, P. 1993. Korkeakouluarvioinnin järjestäminen Suomessa (How to organise the evaluation of Higher Education Institutions in Finland) In: Korkeakoulutieto 4, 42–52


Valtionhallinnon arviointityöryhmän loppuraportti 1999 (Final report of the working group on evaluation of state administration) Työryhmämuidstoiota 14/99 Helsinki: Valtiovarainministeriö


considerations based on multiple constituency evaluation) In: Hallinnon tutkimus 1, 30–40


Legislation and standards and guidelines

Asetus korkeakoulujen arviointineuvostosta. Nro 1320/1995 (Decree on FINHEEC)


Yliopistoasetus 115/98 (Universities Decree)

Yliopistolaki 645/97(Universities Act)
Evaluation Reports


Helsinki University of Technology. NA. HUT Self-Evaluation Report. CRE Institutional Review Programme. (handout)


Interviews

University of Kuopio
Rector, two vice rectors, and a former rector
Four deans (Medicine, Natural and Environmental Sciences, Pharmacy, Social Science)
Four institutional level administrators
Two faculty administrators
One staff member from the continuing education centre

University of Turku
Rector and two vice rectors
Five deans (Education, Humanities, Law, Medicine, Mathematics and Natural Science)
Three institutional level administrators
Five faculty administrators
One staff member from the continuing education centre
Others
The rectors of three other universities: University of Joensuu, Helsinki School
of Economics, and University of Art and Design
Two members of the peer review teams
Three Ministry of Education officers

Ministry of Education
Director, 2000, August 10 (notes by the author)

External panel’s oral feedback to FINHEEC
(as recalled by Kristoffersen, Wahlén and Liuhanen; part of the Mutual Recognition project, Lindeberg, T. and Kristoffersen, D., 2002, ENQA Occasional Papers 4)
Appendices

Appendix 1
The Interview Themes at the Universities

1 intro

- My study is about university evaluations, and I am about to describe how the evaluations were used at two universities. Interviews provide a better base for the description of evaluation use (= utilisation) than documents would.

- I hope that tape-recording is ok.
- Your name will not be mentioned in the study. I will use codes.
- Any questions?

2

- First, please tell me about the nature of your involvement in the university evaluation.
- How was it conducted? How extensive was the participation by staff in self-evaluation?
- Were there differences between the faculties and departments?

3

- What does "utilisation of university evaluation" bring to your mind? I am interested in how you utilised the evaluation or what sort of benefits you think it has brought to the university. All kinds of utilisation are as welcome as no utilisation. I am not looking only for utilisation that is in line with official aims. Think both about the self-evaluation process, the site-visit of the peer review team, and the report. I'll then continue with more detailed questions.

The focus of utilisation - where

- in which functions of the university has the evaluation been used most
- teaching, research, management, administration? Any examples?
- anywhere else? Outside the university? In the annual result-negotiations?

Process and results

- how was the self-evaluation conducted?
- how useful were the evaluation processes and the reports?
- have you used the reports since? How?
Types of utilisation
– has there been discussion about the evaluation afterwards? In what kind of situations?
– has the evaluation helped in making decisions or implementing changes? How?
– when negotiating with the MoE? For example?
– in cooperation with stakeholders? For example? What about strengthening the image of the university?
– did you learn something during or from the evaluation?
– did the evaluation offer new knowledge about departments or faculties?

Users
– whose project was the evaluation at the university? Who or which units have been the most eager users? Why?
– have you personally used the evaluation or have you found it useful? How about others?
– did the evaluation cause some harm? What kind of harm, and where can you see it?

Expectations, fears, rumours
– what did you expect of the evaluation? Fears, prejudices? How about others?
– did the university set aims for the evaluation? Were they fulfilled?
– did you actively search for information about evaluation? Where?
– could you make use of earlier evaluations (in your own university or elsewhere?)
– how did you disseminate the ”message” of the evaluation? Seminars, discussions, articles?

Other questions
– do you think that this university / faculty / department is different than others in how it makes use of the evaluation?
– if you compare university evaluation with other evaluations such as evaluations of degree programmes or research, do you see differences?
– how do you think the MoE has utilised the evaluation of this university?
– what were the biggest benefits of the evaluation for your university?

WAS IT WORTH DOING?
1 "... täällä ei ollut tiedekuntia, vaan täällä oli isot laitokset. Ei resursoitu yhtä tiedekuntaa varten opettajakaan vaan ne oli yhteisiä. Tämä lähti uudella idealla, ja sen jälkeen 1989 tiedekuntajärjestelmä tuli meitä vastaan. Eli toimittiin pitkään ja se teki sen, että mielestäni Kuopion yliopiston väki ei ole koskaan ollut niin suppeaa kun mitä näkee muualla.” (R3).

2 "Olimme tehneet ennen sitä (alueellisen vaikuttavuuden arviointia) tällaisen tutkimuksen arviointijärjestelmän, ... Ja aika paljon luotin sitä arviointipohjaa siinä siihen mennessä. --- Toinen, mikä me tehtiiin ja oli tehty edellisinä vuosina ja jatkettiin, niin tehtiiin tällainen perusvoimavarojen jakojärjestelmä, jossa oli työryhmä ja käytiin hyvin nopeasti läpi tällaiset kustannustrakenteet, mitä opintoviikko maksaa eri tiedekunnissa. Ja arvioitiin, miten paljon annetaan tiedekuntien yli opetusta. Tällainen yhteistoiminta, ja koko se porukka sai hyvin opin arvioimaan, mitä tämä toiminta on. --- ... silloin aloitettiin myös tämä laatuyliopistohanke,...” (check). (R3)

3 Jos ajattelee, että mihin me sitä käytettiin ihan koko ajan, niin kyllä se oli se keskustelun herättäminen toisaalta alueelliseen näkökulmaan ja toisaalta, kun meillä oli kertytty siihen, että teillä sujuu kaikki toiminta, niin siihen, että miten me voisimme itse itseämme kehittää... (IA5).

4 ... ei sellaista tiimityöskentelyä, ei sellaista ryhmää monessa paikassa. Hyvin moni, joka tuli meidän porukkaan katsomaan, huomattiin jälkeenpäin, että kaikista näkee, että teillä sujuu kaikki hyvin. Se sujuu ehdottomasti ilman mitään kitkaa. Tuntui, että se ideointi oli, siinä oli nämä hyvin toimivan tiimin merkit myös. (R3)

5 Kyllä minulla on sellainen käsitys, että Kuopion yliopistossa on, johtuen siitäkin, että tämä on vielä aika pieni yhteisö ja täällä ihmisten välisien suora kontakteja pelaa aika hyvin hallinnossakin, niin luulen, että jos sieltä (?) tulee jotain, joka koetaan täällä tärkeäksi ja hyväksi ideaksi, niin luulen, että ensinnäkin se keskustelu käynnistyy aika nopeasti siitä, että koetaanko se täällä hyvänä ja sitten vaadittavia toimenpiteitä kyllä pystytään toteuttamaan. (D3)

6 "... sellaisia asioita ja odotuksia oli, että nyt puhun yksikön tasolla, miten me suuntaamme meidän... onko relevanttia meidän toiminta koulutuksen osalta. Meillä oli aika paljon sellaisia tilanteita ihan
How are University Evaluations used? – 177

oman alanikin osalta, että onko sen koulutuksen, mitä annetaan, suunta oikea. Olemmeko me kiinni tässä yhteiskunnassa ja kaikkea tällaista. - - - - Tietysti se itsearviointi oli hirveän tärkeä prosessi meille tai erityisesti laitoksen kannalta, mutta (myös) tiedekunnan kannalta. Siellä tuli näihin laitoksen ydinkysymyksiin tietoa, jota sitten jos ei ole tyhmä, niin käyttää hyväksi. Koska siellä oikeasti tuli sitä sellaista palautetta. Että silloin voisi ajatella, että laitostasolla se oli hyödyllinen...

"Suurin ongelma oli ennen kaikkea, että minä näen alueellisen vaikuttavuuden ennen kaikkea yritystoiminnan sukupuutkien ja tämä (itsearviointi) taas lähtee hyvin laajemmalta näkökulmalta eli se tavallaan katsetelee niitä asioita laajemmalta ja siellä on erilaisia vaikuttavuustapoja ja tavallaan tuo oikeastaan se suurin asia, mikä siinä oli. Elä tavallaan tuo näkökulmien ero. - - - - Minä lähdim rakentamaan alueellista vaikuttavuutta tiedekunnan kannalta tarkasteltuna aivan tavallaan nollapisteestä. Halusin lähteä kehittämään tiedekunnan alueellisen vaikuttavuuden toimenpiteitä. (D7)

"Ehkä yhteiskuntatieteilijässä tiedekunnassa ikään kuin tällaiset yhteiskunnalliset asiat jollain tavalla pidetään itsestään selvästi ja otettuina ja nähän niin kiinni jotakin tämä yhteys, vikaa paikalliseen elinkeinopolitiikkaan tai aluekehitysviranomaisiin tai erilaisiin toimijoihin jotakin luontevampaa, koska meidän tutkimuksen ja kehittämisen alue on tässä ympäröivä yhteiskunnassa toisin kuin jollakin se on solu?" D1

Minä olen luonnontieteilijä peruskoulutukselta, niin minun näkemys varmasti johtuu siitä, että minulla on aina ennen kaikkea se hyötynäkökohda. Aina on olemassa päätös ja siitä seurauksia ja sen seurauksen täyttäisi aina olla hyötynäkökohdan mukainen. Siltä pohjalta minä myöskin ajattelen tätä alueellista vaikuttavuutta. (D7)

tavallaan täsmällisemmän diskursiivisen puun tähän, mistä tässä on oikeastaan kyse.” (D1)

11 Meillä oli kaksi asiaa, jotka yliopiston sisällä oli äärimmäisen vaikeita: laatu ja laatujärjestelmä sekä alueellinen vaikuttavuus. Kaksi vaikeaa asiaa, jotka molemmat herätti hirveän voimakkaita tunteita ja reaktioita siihen, että mihin yliopistoa ollaan ajamassa, että ollaanko me siirtymässä johonkin sellaiseen maailmaan, joka ei olekaan tätä oikeaa perustutkimusta ja että kuka tätä rupeaa ohjaamaan ja voidaanko tätä kaikkea systematisoida. (IA5)

12 Oli huomattavan helppo (hyväksyä aihe). ---- Se, että lähdettiin katsomaan alueellista vaikuttavuutta, laskin heti sen, että menään kaikkein vaikeimpaan tehtävään yliopistolaitoksen kannalta. Myönnän, että laskin sen, että eihän yliopisto tällä tavalla häviä, kun se on noin vaikea, niin sinne kannattaa menää siitäkin syystä, että se on varma, että saa positiivista viestintä, hyväksyntää yliopistolta eikä kuitenkaan menetä mitään. Kun tehdään pioneeryöitä, niin sitten tehdään sitä ja silloin se kehitteää ajattelua. (R3)

13 Ensimmäinen ajatus käytöstä oli Kuopion yliopiston sisäisen keskustelun aikaansaaminen liittyen aluevaikuttavuuteen. Jos ajattelee, että mihin me sitä käytettiin ihan koko ajan, niin kyllä se oli se keskustelun herättäminen toisaalta alueelliseen nähkökulmaan ja toisaalta, kun meillä oli kysymyssä siihen se laatuvaikutamisesta, niin siihen, että miten me voimme itse itseään kehittää ja niillä painotuksilla, joita me… Ja kun siihen kytettiin aluenvääkälma, niin meillä se käyttö lakaisi sen, että miten Kuopion yliopisto ihan oikeasti voi tällä alueella toimia. Se lakaisi sen keskustelun, että miten meidän pitäisi omia prosesseja kehittää. (IA5)

14 ”(Arvioinnin hyödyntämisestä) tulee ensimmäiseksi mieleen se, että hyödyntämisessä itse asiassa, oikeastaanhan tehtii sen prosessin aikana, tehtiin hirveästi asennetut (puhuu itä-suomen yhteistyöstä) ------ Tämän on nosti jäsentyneenä esille tavallaan niitä kipupisteitä ja antoi henkistä tukea sille, että oli esim. rehtorina huomattavasti helpommin vastata siihen kysymykseen, että mikä on oikein ja mikä väärin ja mikä on yliopiston tehtävä ja mikä ei ole. Se ei ole vähäpätiöineen asia itse asiassa ollenkaan, jos tulosten hyödyntämistä ajatellaan. Kyllä vastaavasti sitten kun tätä kautta tiedekunnat ja niiden avainhenkilöt ovat olleet arviointiprosessissa mukana ja tietyllä tavalla kysymyksiä joutuneet pohtimaan, kyllä se vastaavasti, olen kokenut, että se on tiedekuntien pohdinnossa ja strategioissa näkynytt. Päitsi, että koko yliopiston mittakaavassa, myös sitten laitos- ja tiedekuntakuntatason asioiden käsittelyssä, vähän eri tavalla eri puolilla, ettei se ole homogeenista, mutta ei sen tarvitse ollakaan.” (R2)
"...tutkimuskentässä tuli ehkä vähän sellainen ahaa-elämyskin, että Kuopion yliopisto, joka on aika perinteisesti ollut täällä alueella toimivana yliopistona aika etäinen, kun tämä on lääketieteeseen ja tähän painottuva ja mielikuva sellainen, niin ehkä siellä syntyi sellaisia ahaa-elämystä, että meillähän onkin tällaisia ja tällaisia tutkimusryhmiä, joilla olisikin alueella olevaan elinkeinoelämään kontakteja, jos vaan osattaisiin ne löytää. Se siis ravisteli aika lailla meidän tutkimuskentää." (IA5).

Tällä hetkellä suurin hyöty on siinä, että miettimällä ja tekemällä tällaisia arviointeja, miettimällä näitä asioita suosituiin vastaamalla, niin se on sitä prosessia, mistä syntyy siitä ymmärrystä siitä, mitä alueellinen vaikuttavuus vaikuttaa voisi täällä alueella olla ja mikä voisi olla Kuopion yliopiston rooli alueellisessa vaikuttamisessa (D3).

"Tämä idea on oikeastaan aika vanha, se menee vuosiin 1993 ja 1994 ja oikeastaan OECD:n IMHE-ohjelman ja silloisen rehtorineuvoston eli CRE:n järjestämiin johtamisseminaareihin .... täällä Turussa kesällä 1994. --- Oli sellainen ajatus, että tämä olisi sellainen aihe, joka nousee esille ja siitä sitten Turun yliopistollekin voisi olla hyötyä näin meikäläisissä vanhana yliopistona, koska vaikka alueellinen rooli meillä 90-luvun alulla oli jonkunlainen, niin olihan se silloin huomattavasti epämääräisempä kuin mitä se on nyt. Siinä mielessä tämä on ollut kyllä meille hyvää aihe." (IA3)

"Nyt tässä hetkeksi muodostui selvästi tällainen ryhmä kuin akateeminen johto, joka ryhmittyi pohtimaan, että mitä pitäisi tehdä yliopiston ongelmille." Onko se vielä olemassa?" No, totta kai siitä on jäänyt jälkiä, mutta ei ehkä samalla inteseetillä kuin siinä oli hetken aikaa." (D2)

"Kyllä tässä nyt siinä mielessä kävi vanhanaikaisesti, että kyllä se aika lailla) sen jälkeen, kun tämä oli ilmestynyt, niin vähemmän siitä on puhuttu kuin olisi ollut ehkä syytä. Sehän hyvin helposti näillä käy. Puhun nyt nimenomaan siitä, että noin yleisesti yliopistoyhteisössä se ei ole kovin paljon ollut esillä. Se olisi varmaan edellytynyt yliopiston johdolta aktiivisempaa otetta sen ylläpitämiseksi, mutta se ei kyllä ollut sellaisena läpaiusevänä.” R4

"...kun tätä ruvettiin katsomaan, että onko tämä totta, että meidän toiminta onkin näin upeaa. Monet kokivat tämän asian näin. Ja monet katsoessaan omaakin tekstiä sanoivat, että luulin että tämä juttu on vaatimaton, mutta kun tuli teksti valmiaksi, niin se olikin suhteellisen hyvä. // Sitten vielä kun Goddard kertoiv meille, että hänen
käsityksensä mukaan, kun hän oli aika paljon näitä eurooppalaisia korkeakoulujen selvitellyt ja arvioinut, että alueellinen kehittämistyö meidän täykkärisä on ihan eurooppalaisa tasoa parempaa. Tämä tietenkin oli meidän porukoille... varmasti hirveän kiva viesti. Oli.” (CE2)

22 ”... lähes kaikki ihmiset jotka professoreiks valitaan, niin haluavat ensisijaisesti olla tutkijoita. Että sitten jopa tämä opettaminen on sellanen tavalla ylimääräinen tehtävä, että osalle ihmisistä sitten ne yhteydet paikallisen elinkeinoelämään on läheisempää kuin se opetus siinä mielessä että se voi useinkin liittyä heidän tutkimuksellisiin intresseihinsä ja mä sanoisin melkein että noin yksilöinä löytyy esimerkiksi meidän tiedekunnasta aikamoinen joukko ihmisiä, joilla tuntuu olevan ihan luontevaksi osana se ulkon toiminnan tai yhteydet paikalliseen elinkeinoelämään. Mut se ei oo silleen menny läpi vielä toistaseks että se olis tavallaan osa suunnittelua, että sitä miellettäis osaks yhteisön toimintaa.” (D4)

23 ”Ehkä tähän ulkoisen vaikuttavuuden arviointiin liittyen aika mielenkiintoista oli se, että vaikka se tavoitettiin näin, niin yliopistoyhteisö oli jotenkin sellaisessa tilanteessa, että haluttiin käyttää se paljon laajemmin. Eihän se siis ollutkaan leimannut sitä koko työtä se ulkoinen vaikuttavuus, vaan se oli ehkä leima, jonka alla se tehtiin. Se haluttiin tehdä sellaiseksi paljon isommaksi kokonaisarvioinniksi, jossa sitten yhtenä näkyvänä osana on tämä ulkoinen vaikuttavuus. Ehkä sitä ulkoista vaikuttavuutta pohdittiin sitten enemmän siinä koko yliopiston tasolla, mutta esim. yksittäisten tiedekuntien tasolla tietystä näissä itsearvioineissa sitä katsottiin kyllä ihan koko toiminnan näkökulmasta ja tällä ulkoisella vaikuttavuudella suoraan olisi ika pienikin merkitys. Mutta tämä oli yksi asia, se itsearvioinnin voimakas esieltuonti.” D2

24 ”... arviointia voi käyttää, jos osuu olemaan sopiva ihmiset sitte käyttämään niitä, mut arviointi itsessään ei välttämättä johda yhtään mihinkään, jos ei se niinku osu joihinkin ihmisiin, jotka ryhtyy viämään sitä eteenpäin.” (D6)

25 ”Korostan sitä, että enemmänkin on lähdetty liikkeelle, että ei me nyt kovin monessa asiassa ole päästyt kovin pitkälle vielä. Juuri se, että se on potkaisussa asioita vauhtiin, siis muutoksen vauhtiin, niin se voisi olla toinen merkitys. Ehkä sitten myöskään se, että se tavallaan kouluttaa yliopistoa vähän tällaiseen toisenlaiseen kulttuuriin.” (R4)

26 Tämänkos nosti jäsentyneenä esille tavallaan niitä kipupisteitä ja antoi henkistä tukea sille, että oli esim. rehtorina huomattavasti helpompia
How are University Evaluations used? – 181

vastata siihen kysymykseen, että mikä on oikein ja mikä väärin ja mikä on yliopiston tehtävä ja mikä ei ole. Se ei ole vähäpäätöinen asia itse asiassa ollenkaan, jos tulosten hyödyntämistä ajatellaan. (R2)

27 “Ensimmäinen ajatus käytöstä oli Kuopion yliopiston sisäisen keskustelun aikaansaaminen liittyen aluevaikuttavuuteen. Jos ajattelee, että mihin me sitä käytettiin ihan koko ajan, niin kyllä se oli keskustelun herättäminen toisaalta alueelliseen näkökulmaan ja toisaalta, kun meillä oli kytketty siiven sen laatunäkökulma, niin siihen, että miten me voisimme itse itseämme kehittää niillä painotuksilla, joita me... Ja kun siihen kytkettiin aluenäkökulma, niin meillä se käyttö laukaisi sen, että miten Kuopion yliopisto ihan oikeasti voi tällä alueella toimia. Se laukaisi sen keskustelun, että miten meidän pitäisi omia prosesseja kehittää.” (IA5)

28 Kyllä minusta tämä tällainen on enemmän koko yliopiston työkalu. Kyllä siitä tiedekunnatkin saa ehkä jotain irti ja ehkä siihen omaan arviointiinsa jotain selkärankaa ja taustaa, mutta kyllä se enemmän koko yliopiston työkalu sitten suhteessa nyt vaikka muihin yliopistoihin tai yhteistyökumppaneihin, eri ministeriöihin ja kaupunkiin, enemmän sellainen. Että on näyttää, että meitä on arvioitu ja nämä on todettu hyviksi puoliksi ja näissä meillä on kehittämisen varaa. Ja me nyt niitä kehitetään sitten tällä tavalla. (IA1; my emphasis)

29 ”Suoraan kun sanon, niin minusta siitä ei ole keskusteltu lainkaan. Eikä ole kovinkaan paljon voitua hyödyntää. OPM:n kanssa kaiken kaikkiaan tästä aluevaikuttavuusasiasta pitäisi keskustella paljon enempää ja tästä kokonaisuuden kehittämisestä ja kokonaisarvioinnista ja sen hyödyntämisestä yliopiston kehittämisessä. Hyvin vähän on keskusteltu.” (IA5)

30 ”Se (Teaching and Learning Council) on yksi asia, joka ei todellakaan perustu mihinkään ulkoisen vaikuttavuuden kysymykseen, vaan tähän yleiseen arviointiin, jota tämän saman prosessin yhteydessä käytetään. Talle kansainväliselle ryhmälle tuli selvästi kyllä esille, että meillä on koko yliopistossa hyvin heikosti panostettu opetuksen kehittämiseen ja opetuksessa ilmenevien ongelmien tarkasteluun ja sellaisten tukipalveluiden rakentamiseen, jotka edistäisivät opetuksen laatua. Sitten pienessä määrin myös tämä suhde ulkoiseen vaikuttavuuteen; ne tulevat sitä kautta, että kun tehdään opetussuunnitelmallista työtä, niin miten siinä otetaan huomioon tätä maailman kehitystä yliopiston ulkopuolella muutenkin, kuin sitä kehitystä, joka tapahtuu itse tieteessä.” (D2)
Appendix 3
Letter to the rectors of three other universities who were interviewed

Anna-Maija Liuhanen

Ref. Telephone conversation on …..

Dear … (the rectors of three Finnish universities)

As mentioned on the phone, I am preparing a doctoral thesis on the utilisation of evaluations at Finnish universities. My supervisors are Professors Seppo Hölttä and Pentti Meklin at the University of Tampere.

Little research has been undertaken on Finnish university evaluations. Existing studies focus on the impacts of the evaluations rather than on their utilisation. Thus, they do not discuss the perspectives of a user or users, which is an essential component of my study. The available reports of follow-up evaluations usually remain narrow from the utilisation point of view, as they cover only the implementation of recommendations, without considering other types of utilisation such as legitimation and knowledge building.

My aim is to create a description of how university evaluations have been used. The description consists of the purposes and functional areas of evaluation use, and to identify the users of evaluation. Further, it identifies how the evaluation process and reports have been used. In addition, factors that facilitate and obstruct the utilisation of university evaluations are considered. The description is based on interviews carried out with staff at two Finnish universities. Thus, a case study methodology has been used. In addition to university staff, I intend to interview members of peer review teams, and some Ministry of Education staff.
Types of evaluation use

When looking for an answer to the question HOW evaluation has been used, I use the types of utilisation discussed in literature. In the interviews conducted at two Finnish universities, I have recognised instrumental, conceptual, enlightening, legitimising, interactive, symbolic and ritual utilisation. However, these forms of evaluation are not clearly defined, but often overlap or are sequential in nature. Instrumental evaluation refers to decision making and measures based on the evaluation. Thus it relates to the expectation of improvement connected to evaluation. Conceptual use means knowledge building, and enlightening by providing new perspectives or understanding. The two can be characterised as learning. Legitimising utilisation, in turn, means either general PR work or using evaluation results in arguments for or against specific changes. The name “ritual” reveals the purpose of utilisation – evaluation for the sake of evaluation, or because it has been required by a higher authority. Symbolic utilisation of evaluation refers to giving an impression – of being modern or rational, for example. Finally, interactive use of evaluation means a situation where evaluation is mixed with other available information.

A description based on interviews in two universities

The picture of evaluation use that emerged from 32 interviews conducted with staff at two Finnish universities can be described briefly as follows. Overall, the utilisation of university evaluations is limited to those in positions of university leadership, such as rectors, directors of administration and various other senior staff in institutional administration. This is in line with the aim set by the Ministry for the university evaluations in early 1990s – to strengthen universities at the institutional level. I was, however, interested in whether the evaluations were used at other levels within the organisation. Hence, I kept asking whose instrument such an evaluation actually is. It seems that special measures are needed in order to distribute a university evaluation to all parts of a university. At the faculty level, the utilisation of evaluation was occasional, and there was even less of it in departments. Evaluation was utilised at these levels if the ”right kind of people” happened to be there.

➔ utilisation of university evaluation takes place mainly at the institutional level

As to the above mentioned types of evaluation use, legitimising utilisation was described almost exclusively by rectors, but not by all rectors or vice-rectors. Within a university, evaluation was used
internally to provide an argument for changes, and outside for image building. Obviously rectors also use evaluation instrumentally, for example by establishing new degree programmes or by changing the university’s strategic direction, which corresponds with the purpose of developmental evaluation.

At the faculty level conceptual and enlightening utilisation were the most usual types, while in departments the most usual type of evaluation use was ritual. The ritual use of self-evaluation meant more or less answering questions that had been asked by others. Interactive use, in turn, seems to relate to two different phases of evaluation, first to how the self-evaluation is conducted, and second to the consideration and implementation of the recommendations made by the peer review team. This is the general picture. However, individuals at each organisational level can see the utilisation of evaluation in different ways.

➔ the utilisation of evaluation is different at different organisational levels

The general view seemed to be that peer review reports are used mainly by rectors and institutional administrators, the self-evaluation report by special functions areas (such as centres for continuing education and PR units), and the evaluation process by the academic community. In spite of some exceptions, this observation seemed to be mainly right. Generally, the evaluation process in particular was considered useful, because “you learn in the process”, as one of the interviewees said. Although the self-evaluation reports were praised for being the first comprehensive descriptions of the whole university, their utilisation seemed to remain rather modest. The reports of the peer review teams, in turn, are mainly used by rectors and institutional administrators.

➔ The different parts of university evaluations have different users.

The faculties are aware of very little if anything about the self-evaluations of other faculties. Likewise, people in institutional administration and management know almost nothing about how faculties and departments conducted their self-evaluations. In some faculties, staff were disappointed when they did not get any feedback about their self-evaluations.

➔ weak interaction, no knowledge sharing

Among the factors that facilitated the utilisation of evaluation, the interviewees often mentioned the theme and timing of the evaluation,
emerging ideas and external pressures. The situation was especially opportune if the four were in line with the recommendations of the peer review team. In those cases one can speak about the pertinence of evaluation. As to the dissemination of evaluation, “right persons” were in a key role. A similar factor, though considered rather an obstacle, was the change of key personnel (such as deans, department heads). These two can be described as discontinuity. It was also suggested that the evaluation should have been “sold” to the academic community, which mainly meant describing the background and aims of the evaluation. In other words, it was necessary to answer the question why (for what purpose) the evaluation was conducted. Further, factors mentioned that either obstructed or facilitated evaluation use included the organisational nature of universities, use of a foreign language, and the expertise of the peer review team, the positive impact of the follow-up evaluation, and the challenge offered by the evaluation.

➔ pertinence, challenge and opportunity / incontinuity, random nature, discontinuity, and lack of interaction and leadership

Questions to the three rectors

1. Does the description above correspond with your experience of the evaluation at your university?

2. Is something that was clear or important in your university missing here?

3. Which of the types of utilising evaluation do you recognise as instruments for rectors’ use?

4. Should the utilisation of a university evaluation be extended to all organisational levels, or is it enough when it is used by the leadership?