Gustav Jakob Peterson
Ove Karlsson Vestman

Conceptions of Evaluation

nshu

MÄLARDALENS HÖGSKOLA

Contents.

Section 3: Case study 2 - Research.

Section 4: Case study 3 – Debates.

Section 5: Concluding discussion.

Section 6: Attachments.

Section 1: Introduction.

1. The problem.

In April 2006 Mälardalen University made an application at NSHU – the Swedish Agency for Networks and Cooperation in Higher Education – for economic means to conduct a study on evaluation. The application was granted.

The application establishes the following objectives for the intended project ("Ansökan om medel för symposium och nätverk för utvärdering", application by Ove Karlsson Vestman, Mälardalen University, addressed to Gunnel Wännman Toresson, NSHU):

1. To provide an overview of Swedish courses and programmes, which deal with evaluation, and to make a brief comparison with other European countries:
> "Ett syfte med det projekt som ansökan avser är att inventera utbudet av utbildning i utvärdering i Sverige, men vissa internationella utblickar."

2. To elaborate on differences between countries concerning evaluation research and education by discussion at a symposium in Sweden in June 2006 and at the European Evaluation Society conference in London in October 2006:
> "Det första steget i projektarbetet tas under juni 2006 med arrangerandet av ett symposium med inbjudna utvärderare och utvärderingsforskare. Förutom avsikten diskutera centrala problem inom utvärdering är syftet med symposiet att diskutera idéer till utbildning och gemensamma kurser som helst kan ges på distans. En fortsättning på symposiet kommer att äga rum vid EES:s (European Evaluation Society) konferens i oktober 2006. Där kommer en särskild panel att organiseras kring utvärdering i Europa och EU med jämförelser till USA."

3. To provide an overview of frequently discussed themes in the contemporary Swedish and European academic debates on evaluation:
> "Avsikten är också att som bakgrund till frågan om utbildning och utformning av utvärdering ge en bild av diskussionen kring aktuella problem inom utvärdering i Sverige och Europa."

4. To make a list of Swedish qualified evaluators:
> "Ett annat syfte är att inventera kvalificerade utvärderare i Sverige som kan utgöra en resurs för : utvärderingsuppdrag från olika myndigheter, kommuner etc."

In the spring of 2007 an agreement was made between Ove Karlsson Vestman, Mälardalen University, and Gunnel Wännman Toresson, NSHU, establishing that the fourth objective should, due to practical reasons, be fulfilled in connection with the Swedish Evaluation Society conference of October 2007.

In short, all the three remaining objectives aim at investigating which predominant conceptions of evaluation there are in the contemporary Swedish academic world. Thereby this report is descriptive rather than normative. More precisely, the objectives deal with conceptions of evaluation in three different spheres of the academic world; education (purpose 1 and 2), research (purpose 2) and debate (purpose 3).

Three separate case studies have been conducted. Each one of them deal with evaluation within one of the three spheres of academic work.

The first case study aims at fulfilling the first objective and the part of the second objective that deal with education. The second case study aims at fulfilling the part of the second objective that deals with evaluation research. The third case study aims at describing the

academic debate on evaluation and thereby fulfilling the third objective. In the following sections the operationalisations, the methods of analysis and the sources upon which the case studies draw will be presented.

2. Case study 1: Education.

The aim of the first case study is to present an overview of a number of Swedish higher educations, which deal with evaluation.

2.1. Operationalisations.

The following questions have been asked concerning each of the courses and programmes under study:
- Is evaluation a main theme of the course/programme?
- Which department or institute provides the course/programme?
- How many course credits does it give?
- Does it run full or part time?
- Is it a distance course/programme?
- Is it a course/programme on the foundation or on the advanced level?
- Are there any specific admission requirements?
- How can the evaluation knowledge taught be characterised?
- Is evaluation theories a theme of the course?
- Is an explicit definition of evaluation made?
- Which required readings are there?

2.2. Methods of investigation.

In order to provide a judgement on whether evaluation is a main theme of the analysed courses and programmes the following criteria have been used:
- Whether evaluation is mentioned among the goals, objectivs or topics of the course/programme.
- Whether any priority has been declared between goals, objectivs or topics that deal with evaluation and those that don't.
- Whether evaluation is in the title of the course or in the title of any of the sub courses.
- Whether it is stated that evaluation knowledge taught in some of the sub courses should be implemented in other sub courses.

The course credits of all analysed courses/programmes[1] are presented in accordance with the Bologna system. One course credit of the previous Swedish system corresponds to 1.5 course credits in the Bologna system.

Five ideal types of evaluation courses have been formed. An ideal type is a heuristic tool, which can be used in comparisons with real world objects. The properties of the ideal types will be compared to the ones of the courses and programmes under study. Thereby judgements on the character of the evaluation knowledge taught in each course/programme can be made. In the following presentation of the five ideal types it is also indicated what kind of courses can be said to resemble each ideal type:
1. A course focusing on evaluation of a certain object. One example is a course dealing with by which mechanisms different antiseptics work. Evaluation may be taught as

---

[1] Courses and programmes which deal with evaluation and are offered by universities or institutes of higher education will henceforth be referred to as evaluation courses/programmes.

one tool among others to find out the mechanisms by which the antiseptics work. The focus, however, is not on evaluation in general, neither is it on evaluation within the discipline of medicine, but on the antiseptics.

2. A course on how evaluation is or should be conducted within one discipline or field. One example of a course dealing with evaluation within one discipline is a course on how evaluation is dealt with in the discipline of pedagogik. Courses on evaluation within one field can be exemplified with a course on how evaluation of public administration is or should be performed[2].

3. A multidisciplinary evaluation course; a course comparing how evaluation is dealt with in different disciplines and teaching the basics of evaluation as conceived in the disciplines dealt with. One example is a course dealing with the criteria used in evaluations within different disciplines, for example pedagogik, sociology and economics. Another example is a course which teaches evaluation methods or theories that are frequently used in all relevant disciplines.

4. A transdisciplinary evaluation course; a course that deals with problems that are specific to evaluation and are vital no matter in which discipline an evaluation is undertaken. One example is a course which uses different philosophical theories as a starting point for a discussion on different criteria used in evaluation; which moral theories can be referred to in advocating which criteria?

5. Meta evaluation courses; a course that is focused on the consequences of evaluation for the state and for society. One example is a course dealing with the consequences of evaluation in public administration or in business enterprises.

The topics, objectives and goals of the courses/programmes under study have been compared to these five ideal types. This report mainly focuses on evaluation courses/programmes, which resemble ideal types that have been given the higher numbers above. Courses/Programmes resembling the first ideal type are only of secondary interest as such courses don't mainly deal with evaluation, but with a certain object under study.

If evaluation theories are mentioned that will be reported, but if that theme is not included in the course that will not be reported. Likewise, if evaluation is defined that will be reported, but a lack of definition will not be reported.

2.3. The cases.

First, the study has been delimited to courses and programmes at Swedish universities and institutes of higher education. The ranges of evaluation courses and programmes at all Swedish universities has been investigated. In addition we have chosen the institutes of higher education where we knew that we would or expected to find evaluation courses or programmes.

The selected universities and institutes of higher education are:
*   Ersta Sköndal University College
*   Göteborg University
*   Halmstad University
*   Karlstad University
*   Linkoping University
*   Luleå University of Technology

---

2 In this report a field is referred to as a knowledge area which in Sweden doesn't constitute a discipline, but may do so in other countries and just like a discipline has got its own scientific discourse.

- Lund University
- Mid Swedish University
- Mälardalen University
- Stockholm University
- Umeå University
- University College of Borås
- Uppsala University
- Växjö University
- Örebro University

Second, the study has been delimited to courses and programmes in the following disciplines and closely related multidisciplinary areas:
- pedagogik
- political science
- sociology
- social work
- psychology.

This delimitation of the study is justified first by the results of a pilot study that we have undertaken. The background of this pilot study is that an analysis by Ove Karlsson Vestman and Inger M. Andersson of the contents of Swedish doctoral theses on evaluation – *Pedagogisk Utvärdering som Styrning – En historia från präster till PISA* (2007) – has shown that such research "has dealt mainly with applications of evaluation within different disciplines and research areas" (p 106). Furthermore, they state that "the publications on evaluation are innumerable and its functions vary with the sector in which it is conducted, let alone over time" (p 104). As we are mainly interested in courses and programmes, which focus directly on evaluation rather than apply it, and as this previous study makes it seem likely that such courses and programmes are to be found to a greater extent in some disciplines than in others, the pilot study was undertaken; all courses and programmes offered at Uppsala University which, according to the course descriptions, deal with evaluation were analysed with the questions presented above. The choice of Uppsala University for this pilot study is justified by the fact that it is one of the greatest universities of Sweden and thereby offers courses and programmes in most academic disciplines. 17 courses were found in 12 disciplines. At Uppsala University evaluation courses and programmes, which resemble the ideal types of evaluation courses/programmes that have been given the higher numbers above were mainly found in the social sciences and closely related disciplines.

Second, this delimitation of the study is justified by our experiences as researchers; as we have conducted evaluation within the area of the social sciences and closely related disciplines we know more about how evaluation is conducted and about at which institutes evaluation courses and programmes are offered in these disciplines than in others.

Third, this delimitation of the study is justified by our ambition to achieve a certain depth of study. That is, we wanted to include a large number of evaluation courses and programmes in the examined disciplines. That would not have been possible if the study had covered all disciplines.

One more delimitation of the study has been made; not all evaluation courses and programmes in psychology at the chosen universities and institutes of higher education are

presented in this report. This is justified by the fact that most evaluation courses/programmes in psychology first and foremost resemble the first of the five ideal types of evaluation courses/programmes formed below. As stated above evaluation courses/programmes of that kind are only of secondary interest to us, as they don't primarily focus on evaluation.

2.4. The sources.

The browsers of the websites of the Swedish universities and institutes of higher education have been used to create an overview of their evaluation courses and programmes. Generally only the courses and programmes, which have the word evaluation in their course descriptions have been included in the study, as these are the ones (mostly) found when the browsers of the universities and institutes of higher education are used. In most cases the syllabus of the course/programme under study has been used to obtain further information.

However, the Internet has not always proved a sufficient source to find the syllabus and the readings of the course/programme under study. In some of these cases the institute providing the course/programme has delivered the information, in other cases a syllabus and readings from a previous semester have been used. However, some of the evaluation courses and programmes run for the first time in the autumn of 2007 and for some of these courses no syllabus and no readings have been decided upon. It will not be mentioned which sources each separate case study draws upon.

3. Case study 2: Research.

The second case study diverges from the two others in a very important way. The aims of case study 1 and 3 is to describe processes that go on without involvement of the authors of this report. By contrast, we play a most central part in the process, that is in the research project, which it is the aim of the second case study to describe.

More precisely, this second case study will present a description of how conceptions of evaluation have been brought up and evolved among the researchers that took part in the so called roots project.

3.1. Method of investigation.

The method of analysis is to define implicit or explicit conceptions of evaluation, represented by the different researchers taking part in the project under study. By discussion and elaboration conceptions of evaluation have evolved and surfaced.

3.2. The case - the "roots" project.

In 2006 a research project on contemporary Western conceptions of the origins of evaluation (to be found as well in the academic world as in the public administration) was undertaken – "Evaluation roots in the USA and in Europe – tracing traditions?". One of the aims was to describe differences as well between countries as also between Europe and the United States. That is, the method of analysis of the project was country comparisons. In this comparative perspective Sweden was one of five cases. The participating researchers were supposed to contribute with views on the origins of the development of evaluation in their respective countries. The countries under study were Sweden, the United States, Germany, the Netherlands and Italy. Differences in conceptions of evaluation as well between the academic spheres of the countries as also between their public administrations were focused upon.

In early 2006 a group of evaluation researchers interested in the development of evaluation traditions was gathered; Ove Karlsson Vestman (Mälardalen University, Sweden), Marvin Alkin (University of California, Los Angeles, USA), Frans Leeuw (the Netherlands), Nicoletta Stame (University of Roma "La Sapienza", Italy) and Reinhard Stockmann (University of the Saarland, Germany). It was decided that the group should have a panel discussion at the UKES/EES joint conference in London 4-6 October 2006 on the development of the European evaluation tradition. Before that these evaluators came together for a symposium in Stockholm, June 17-19 2006. This symposium was also attended by Jan-Eric Furubo, Evert Vedung and Gustav Jakob Petersson. Jan-Eric Furubo and Gustav Jakob Petersson also took part in the panel discussion in London. However, the Stockholm symposium was preceded by some preparatory meetings attended by a group of Swedish evaluators; Jan-Eric Furubo, Maria Bergström, Rolf Sandahl, Evert Vedung, Ove Karlsson Vestman and Gustav Jakob Petersson.

3.3. The sources.

The sources upon which the second case study draws are documents produced by the researchers in the roots project, for example minutes of discussions and drafts for papers.

4. Case study 3: Debate.

The third case study has been performed in order to map predominant conceptions of evaluation is the sphere of academic debate. Evaluation is in this case study viewed as a field of debate, defined by the arguments brought up.

4.1. Operationalisation.

As we conceive it today's academic discussion on evaluation is mainly an international one, but we would like to scrutinise that hypothesis. The discussion to a great part takes place at evaluation conferences, where evaluation researchers and practitioners get together and elaborate their thoughts. Therefore, the first purpose of this third case study is to present a picture of which themes of the Swedish debate on evaluation that are to be found more or less to the same extent in the European debate and, furthermore, which ones seem to be more frequently discussed in Sweden and Europe respectively. The second purpose is to investigate if the themes discussed in the Swedish debate also occur in the American one.

4.2. The cases.

In this case study we have focused on the latest conferences of the European Evaluation Association, the Swedish Evaluation Society and the American Evaluation Association respectively. The latest conference of the Swedish evaluation Society (SVUF), "Utvärderingssystem i Sverige – till vad och för vem?" ["Evaluation systems in Sweden -- for what purpose and for whom"], took place on October 20-21 2005 in Stockholm. The latest European evaluation Society conference was the UKES/EES joint conference "Evaluation in Society: Critical Connections", which took place on October 4-6 2006 in London. Finally, the latest conference of the American Evaluation Association, "The Consequences of Evaluation", took place on November 1-4 in Portland, Oregon.

4.3. Methods of investigation.

The themes explicitly mentioned in the titles of the papers and panel discussions of the Swedish Evaluation Society conference have been compared to the ones explicitly mentioned in the titles of the papers presented at the UKES/EES joint conference.

First, in order it to get a picture of which themes were common on both conferences the papers of the UKES/EES conference were clustered in groups consisting normally of three to twelve papers. This was done in the following way. At the conference the papers had been presented in six so-called streams, that is thematic areas. These streams were used as a starting point and each stream was divided into 2-4 clusters of papers, each consisting of papers with closely related themes. Not all paper titles could be included in this study though, as some of them were rather diffuse. Thereafter, the paper presentations and panel discussions of the SVUF conference were, as far as possible, divided into the paper clusters of the UKES/EES conference. Thereby thematic similarities could be observed.

However, some papers of the UKES/EES conference seemed to deal with themes that were not focused upon at the SVUF conference and vice versa. In order to make sure that these were not false impressions generated by the original structuring of the papers of the UKES/EES conference the lists of papers of that conference were searched for certain key words central to the themes that seemed specific to the SVUF conference.

Finally the titles of the sessions of the American Evaluation Association (AEA) conference were searched for the keywords of the main themes of the Swedish Evaluation Society conference.

Furthermore, in order to scrutinise our findings we asked our colleagues from the "roots project" for their opinions on the preliminary findings. They were asked whether they found the selected topics to be among the most prominent in the contemporary academic debate on evaluation and whether they would like to add further themes that are frequently discussed.

4.4. The sources.

The third case study draws upon the programmes of the three conferences. At the UKES/EES conference approximately 300 papers were presented. At the SVUF conference 22 paper presentations and panel discussions were held. The American Evaluation Association (AEA) conference covered more than 500 sessions.

The complete programme of the UKES/EES joint conference "Evaluation in Society: Critical Connections" can be found on http://www.profbriefings.co.uk/EISCC2006/ (2007-04-01).

The complete programme of the Swedish Evaluation Society conference is to be found at http://www.svuf.nu/konferensen2005/svuf_konferensprogram.pdf (2007-03-30).

The complete programme of the American Evaluation Association conference can be found on http://www.eval.org/search06/allschedule.asp (2007-04-01).

Section 2: Case study 1 – Education.

1. Evaluation courses and programmes in pedagogik.

1.1. Göteborg University.

1.1.1. Kvalitet och utvärdering inom utbildning.

The course "Kvalitet och utvärdering inom utbildning" has got two main themes: evaluation and quality assessment.

It is offered by the Department of Education. It gives 15 course credits and runs part-time (50 %, afternoons and evenings).

It is a course on the advanced level. The specific admission requirements include different kinds of studies in pedagogik on foundation level. It is one of the electable courses of the master programme in pedagogik.

The course has got characteristics of the second ideal type of evaluation courses; evaluation within one field or discipline - in this case education:

> "Efter avslutad kurs skall studenten uppvisa
> förmåga att kritiskt granska och förstå olika perspektiv, teorier och modeller för utvärdering och kvalitetsarbete inom utbildning"

It has also got characteristics of the third ideal type of evaluation courses - multidisciplinary evaluation courses:

> "Efter avslutad kurs skall studenten uppvisa
> [---]
> kännedom om metoder och instrument för utvärdering och kvalitetsgranskning
> förmåga att planera, leda, analysera och presentera utvärderingar och kvalitetsgranskningar
> insikter i hur utvärdering och kvalitetsgranskning kan utgöra instrument i ledning och utveckling av verksamheter"

Furthermore, the course has also got characteristics of the fourth ideal type of evaluation courses - transdisciplinary evaluation courses:

> "Kursens huvudmoment är:
> Begreppsanalys, vad innebär begreppen utvärdering och kvalitet?
> [---]"

Finally, the course also deals with topics of meta evaluation, that is with topics of the fifth ideal type of evaluation courses:

> "Kursens huvudmoment är:
> [---]
> Utvärderings- och kvalitetsfrågornas historiska och samhälleliga sammanhang
> Utvärderings- och kvalitetsarbetets roll i utbildningssystemens ideologi, organisation och verksamhet"

Evaluation theories are mentioned a number of times, however not defined:

> "Kursen ger kunskaper och färdigheter i teorier, modeller och undersökningsmetoder för utvärdering och kvalitetsgranskning."

1.2. Linkoping University.

1.2.1. Utbildning och lärande: Arbetslivets pedagogik I.

The course "Utbildning och lärande: Arbetslivets pedagogik I" (Internet-based version) has got individual and organisational learning as its main themes and evaluation is used as one tool among others to create knowledge on the prerequisites and processes of learning. That is, evaluation is not the main theme of the course:

The course is offered by the Department of Behaviour Sciences. It gives 15 course credits, is Internet-based and runs part-time (50 %).

It is a course on the foundation level. The specific admission requirements include 30 credits in behaviour sciences, social work or economics or, as an alternative, at least three years work with personnel organisation.

The course has got characteristics of the second ideal type of evaluation courses - evaluation courses that teach evaluation within one field, in this case learning (in a broad sense) and education:

> "Inom kursen behandlas dels lärande på individ- och organisationsnivå, dess förutsättningar, processer och resultat samt dels de klassiska pedagogiska frågorna "vem ska få vad, hur och varför". [---] Stort utrymme ges för teori och forskning om olika metoder och verktyg för planering, design och utvärdering av utbildning."

1.2.2. Utbildning och lärande i arbetslivet: Arbetslivets pedagogik I.

The course "Utbildning och lärande: Arbetslivets pedagogik I" has got individual and organisational learning and planning, organisation and evaluation of education as its main themes. That is, evaluation is only one theme among others.

The course is offered by the Department of Behaviour Sciences. It gives 15 course credits and runs part-time (50 %).

It is a course on the foundation level. The specific admission requirements include 30 credits in behaviour sciences, social work or economics or, as an alternative, at least three years work with personnel organisation.

The course has got characteristics of the second ideal type of evaluation courses - evaluation courses that teach how evaluation is or should be conducted evaluation within one field, in this case education and learning learning and education:

> "Kursen ger grundläggande kunskaper om utbildning och lärande inom företag och andra organisationer, samt hur man planerar, organiserar och utvärderar utbildning och andra former för kompetensutveckling."

1.3. Lund University.

1.3.1. Utrednings- och utvärderingsmetod och verksamhetsanknutet projektarbete.

Evaluation is one of the two main themes of the course "Utrednings- och utvärderingsmetod och verksamhetsanknutet projektarbete" ("Research and Evaluation in Working Life - Field Work"):

> "Kursens syfte är att ge de studerande kunskaper och färdigheter vad avser utrednings och utvärderingsmetodik som komplement och fördjupning i förhållande till tidigare grundläggande metodkurs. Ytterligare ett syfte är att ge de studerande möjlighet till praktisk tillämpning av metodkurserna i en vald organisation där de följer, dokumenterar och reflekterar över någon av de processer som pågår inom ramen för organisationens personalarbete."

The course is offered by the Department of Education (pedagogiska institutionen). It is an compulsory course of the bachelor program "kandidatprogrammet för personal- och arbetslivsfrågor". However, also other students that have finished either the courses PEDB31 and PED612 or PEDB31 and BAGA01 can attend the course. It gives 15 course credits.

The course has got characteristics of the second ideal type of evaluation courses, that is of courses dealing with evaluation within one discipline or field - in this case pedagogik:

> "Den studerande ska efter avslutad kurs kunna [---] beskriva innebörden av ett pedagogiskt perspektiv på utredning och utvärdering, innefattande villkor för lärande och utveckling i samband med dessa verksamheter"

The course has also got characteristics of the third ideal type of evaluation courses; multidisciplinary evaluation courses:

> "Kursen består av en metoddel och en verksamhetsanknuten projektdel. Inom ramen för metoddelen presenteras och diskuteras principer för hur utredningar och utvärderingar planeras och genomförs."

1.3.2. Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

The course "Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv" ("Pedagogy: Evaluation and Working for Quality in a Learning Perspective") has got four main themes of which at least two involve evaluation:

> "Kursen innehåller fyra huvudteman: 1) Teoretiska perspektiv på kvalitetsarbete och utvärdering i organisationer med fokus på pedagogiska aspekter och lärandeprocesser. 2) Vad förändringar i omvärld och i praktiserade omvärldsuppfattningar betyder för hur organisationer arbetar med att förstå och utveckla sin verksamhet. 3) Metodologiska perspektiv på kvalitetsarbete och utvärdering i relation även till FoU samt 4) Integrering och träning av förmåga att analysera, bearbeta och utveckla vissa metodiska färdigheter inom området."

The course is offered by the Department of Education (pedagogiska institutionen). It can be included in the programme "Magisterexamen med ämnesbredd i pedagogik med professionell inriktning 40 poäng" or of the master programme "Masterexamen i pedagogik 120 högskolepoäng". The specific admission requirement is that 15 course credits of the programme "Magisterexamen med ämnesbredd i pedagogik med professionell inriktning 40 poäng" have been completed. This means that also a bachelor program must have been completed. It gives 15 course credits.

The course has got characteristics of the second ideal type of evaluation courses; evaluation within one field or discipline - in this case pedagogik:

> "Med utgångspunkt i deltagarnas erfarenheter och tidigare utbildningar samt forskning inom området ska deltagarna efter avslutad kurs kunna
> • självständigt diskutera vad det innebär att ha ett pedagogiskt perspektiv på utvärdering och kvalitetsarbete, innefattande villkor för lärande och kunskapsbildning i samband med dessa verksamheter,
> • på ett kvalificerat sätt använda begreppen utvärdering och kvalitet i pedagogiska och organisatoriska sammanhang med förankring i relevant teoribildning och utifrån aktuellt forsknings- och utvecklingsarbete,
> • diskutera och problematisera hur förändringar både i omvärld och av omvärldsuppfattningar i dagens globaliserade samhälle påverkar organisationers arbete med att förstå och utveckla den egna verksamheten,
> • identifiera och problematisera pedagogiska frågeställningar i samband med kvalitetsarbete och utvärdering och att förstå konsekvenserna av olika perspektiv,
> • yrkesmässigt tillämpa teoretiska och metodiska kunskaper i kvalitetsarbete och utvärdering samt att, med hänsyn även till etiska aspekter, tolka kritiskt granska och använda resultat i verksamheten."

It is stated that it student shall learn how to anchor discussions on evaluation in relevant theories. However, it is unclear what kind of theories that are meant:

> • på ett kvalificerat sätt använda begreppen utvärdering och kvalitet i pedagogiska
> och organisatoriska sammanhang med förankring i relevant teoribildning och
> utifrån aktuellt forsknings- och utvecklingsarbete,

1.4. Mälardalen University.

1.4.1. Utvärdering B.

The course "Utvärdering B" deals entirely with evaluation. It consists of two sub courses. The first one - "Teorier och modeller för utvärdering" - deals with theoretical aspects, basic terminology and main themes of evaluation:

> "Delkursen syftar till att ge en elementär orientering om utvärderingens viktigaste begrepp, teorier, modeller, uppläggningar och problemområden. Olika sätt att se på och att genomföra utvärdering belyses genom litteratur och exempel. Exempel på utvärdering i verksamheter med inriktning på utbildning, påverkan och förändring behandlas. Centrala teman i utvärderingsläran dryftas, t.ex. interventionsteorimetoden och effektläran. Aktuella kontroverser som debatteras bland utvärderingsforskare belyses. Tyngdpunkten ligger på en internationell utvärderingsdiskurs, med exempel från svensk utvärderingsforskning."

The second sub course - "Utvärderingspraktik" - is dedicated to evaluation practices:

> "Delkursen ägnas åt utvärderingens praktik med anknytning till kursdeltagarnas egna utvärderingsuppgifter och relevant litteratur. Kursdeltagarna får i uppgift att planera en egen utvärdering. Delkursen syftar till att ge kunskaper och färdigheter så att studenterna ska kunna genomföra denna planeringsuppgift."

The course is offered by the Department of Social Sciences (Institutionen för samhälls- och beteendevetenskap). It gives 15 course credits and runs part time (50 %).

It is a course on the foundation level. It can be taken as a single subject course or as a part of "Pedagogik B".

The course has got characteristics of the third ideal type of evaluation courses - multidisciplinary evaluation courses - as it teaches how practitioners can conceive and conduct evaluations, this without distinction between disciplines:

> "Kursens syfte är att den studerande ska tillägna sig grundläggande teoretiska kunskaper i utvärdering samt praktiskt kunna planera en utvärdering. [...] Kursen innehåller studier i utvärdering inom offentliga verksamheter."

Evaluation theories are dealt with in the first sub course. Such theories are not defined though:

> "Delkursen syftar till att ge en elementär orientering om utvärderingens viktigaste begrepp, teorier, modeller, uppläggningar och problemområden."

1.4.2. Utvärdering ur samhällsvetenskapligt perspektiv D.

The course "Utvärdering ur samhällsvetenskapligt perspektiv D" focuses entirely on evaluation:

> "Kursen ger en översikt av förändringar av utvärdering under de senaste decennierna med tyngdpunkt på internationell utvärderingsdiskurs och med exempel från svensk utvärderingsforskning. Kursen behandlar centrala begrepp, teorier, modeller och olika problemområden i utvärdering."

The course "Utvärdering ur samhällsvetenskapligt perspektiv D" ("Evaluation in Social Sciences") is offered by the Department of Social Sciences (Institutionen för samhälls- och beteendevetenskap). It gives 7.5 course credits, is an Internet-based distance course and runs part time (25 %). Attendance is required twice; at the beginning and at the end of the course.

It is a course on the advanced level. The specific admission requirement is that the third semester (C-nivå) in a social science discipline has been completed.

The course has got characteristics of the third ideal type of evaluation courses - multidisciplinary evaluation courses - as it teaches perspectives, theories and models for evaluation, which are supposed to be useful in all parts of the public administration. That is, they are supposed to be useful for practitioners from all the disciplines under study:

> "Kursens mål är att den studerande skall tillägna sig
> - en överblick över förändringarna av utvärdering inom offentlig verksamhet, vd som utvärderas och varför
> - insikter i perspektiv, teorier och modeller för utvärdering inom offentliga verksamheter"

However, it has also got characteristics of a meta evaluation course, that is of the fifth ideal type of evaluation courses:

> "Kursens mål är att den studerande skall tillägna sig
> [...]
> - insikter i hur utvärdering fungerar som styrinstrument av verksamheter"

Evaluation theories are taught. The formulation "theories for evaluation" implies that prescriptive models for evaluation are taught:

> "Kursens mål är att den studerande skall tillägna sig
> [...]
> - insikter i perspektiv, teorier och modeller för utvärdering inom offentliga verksamheter"

1.5. Uppsala University.

1.5.1. Pedagogik med inriktning mot personalutveckling, baskurs A.

Evaluation is one of the themes of the course "Pedagogik med inriktning mot personalutveckling, baskurs A":

> "I kursen behandlas dels vuxenpedagogiska frågeställningar som vuxnas lärande, vuxenutbildningsinstitutioner, rekrytering till vuxenutbildning samt personalutveckling, dels planeringsprocessen för och olika former av personalutbildning och utvärdering."

It consists of four sub courses of which two, "Personalutveckling och utbildningsplanering 1" and "Personalutveckling och utbildningsplanering 2", partly deal with evaluation. The students are supposed to learn about evaluation theories and their foundations, and to conduct evaluations in practice:

> "Efter genomgången kurs förväntas studenten kunna
> [...]
> - redogöra för olika utvärderingsteorier och vilka grundprinciper dess bygger på samt föreslå en lämplig utvärderingsuppläggning i ett givet fall."

The course is offered by the Department of Education. It gives 30 course credits and runs as well part time (50 %) as a distance course as also full time.

It is a course on the foundation level. It constitutes the second semester of the two programmes "magisterprogrammet med inriktning mot personal- och arbetslivsfrågor" and "kandidatprogrammet med inriktning mot personal- och arbetslivsfrågor".

With the scarce information at hand it is difficult to judge which ideal type of evaluation courses the course resembles the most. Whether it resembles the second or third ideal type the most, that is courses dealing with evaluation within one discipline or field, or multidisciplinary evaluation courses, depends on which evaluation theories are taught and which scenario is provided for the student to suggest an evaluation design.

In the syllabus evaluation theories are mentioned. However, it remains unclear what kind of evaluation theories that are referred to:

> "Efter genomgången kurs förväntas studenten kunna
> [...]

- redogöra för olika utvärderingsteorier och vilka grundprinciper dess bygger på samt föreslå en lämplig utvärderingsuppläggning i ett givet fall."

The courses "Pedagogik med inriktning mot personalutveckling, fortsättningskurs B" and "Pedagogik, fortsättningskurs B" both consist of four sub courses of which two are chosen by the student. One of these electable sub courses, "Styrning och utvärdering", which is available for students of both courses, deals with evaluation. If this course is elected the student is supposed to learn how to analyse the consequences of different evaluation models:

"Efter avslutad kurs ska studenten kunna
[...] - analysera konsekvenser av olika organisationsmodeller för personalarbete och kompetensutveckling i organisationer alternativt analysera konsekvenser av olika styrsystem och utvärderingsmodeller för utbildningssystemet på såväl central som lokal nivå"

The courses "Pedagogik, fortsättningskurs B" and "Pedagogik med inriktning mot personalutveckling, fortsättningskurs B" are both offered by the Department of Education. They give 30 course credits and run full-time.

They are both courses on the foundation level. The specific admission requirement of "Pedagogik med inriktning mot personalutveckling, fortsättningskurs B" is that the first semester course "Pedagogik med inriktning mot personalutveckling A", or a corresponding course, has been completed. The specific admission requirement of "Pedagogik, fortsättningskurs B" is that the first semester course "Pedagogik A", or a corresponding course, has been completed.

The sub course "Styrning och utvärdering" gives 7.5 course credits. As it deals with the consequences of different evaluation models it resembles the fifth ideal type of evaluation courses; meta evaluation courses.

Evaluation and quality assessment are the two main themes of the course "Utvärdering och kvalitetsutveckling":

"Deltagarna skall förvärva såväl teoretisk som praktisk kompetens inom kunskapsområdet utvärdering samt tillägna sig ett kritiskt förhållningssätt till begreppen kvalitet och kvalitetsstyrning.
Innehåll
– Kvalitet och kvalitetsstyrning – framväxt och centrala begrepp
– Modeller och instrument för bedömning av utbildningskvalitet
– Metoder för insamling, bearbetning och analys av data
– Planering, genomförande och avrapportering av en utvärderingsstudie"

The course is offered by the Department of Studies in Education, Culture and Media (Institutionen för utbildning, kultur och medier). It gives 7.5 course credits, is a distance course and runs part time (25 %).

It is a course on the advanced level. The specific admission requirement is that the student either has completed a university programme to become a teacher, or has completed a bachelor's degree in theology or a corresponding education, or has another bachelors degree and has worked as a teacher or education organiser for a least two years and at least part time 50%.

The course resembles the second ideal type of evaluation courses; courses which deal with evaluation within one discipline or field - in this case education:

"Tyngdpunkten ligger på frågor som rör bedömning, redovisning och utveckling av utbildningskvalitet."

"Modeller och instrument för utvärdering av kvalitet i skola och utbildning ägnas särskild uppmärksamhet."

2. Evaluation courses and programmes in social work, work science and sociology.

2.1. Ersta Sköndal University College.

2.1.1. Magisterprogram i socialt arbete.

The programme "Magisterprogram i socialt arbete" consists of two compulsory courses, two electable courses and one advanced study. One of the electable courses, "Utvärdering och kvalitetsutveckling" focuses on evaluation and quality assessment:

The programme is offered by the Department of Social Work. It gives 60 course credits and normally runs part time (50 %). However, it is possible to select another tempo. The course "Utvärdering och kvalitetsutveckling", gives 7.5 course credits. Part of the content of the programme is accessible over the Internet as the programme relies on the Learning Management System LUVIT Education Center.

It is a programme on the advanced level (it corresponds to half a Masters degree).

The course on evaluation and quality assessment, "Utvärdering och kvalitetsutveckling", resembles the second ideal type of evaluation courses; courses on evaluation within one discipline or field - in this case social work:

"Efter genomgången kurs skall studenten
- ha kännedom om centrala teoretiska teman rörande utvärdering och kvalitetsutveckling inom ett kunskapsbaserat socialt arbete
- självständigt kunna planera och genomföra enklare utvärderingsuppdrag inom socialt arbete
- självständigt kunna planera och genomföra enklare kvalitetsutvecklingsuppdrag inom socialt arbete"

2.2. Halmstad University.

2.2.1. Arbetsvetenskap – Utvärdering och uppföljning i arbetslivet.

The course "Arbetsvetenskap – Utvärdering och uppföljning i arbetslivet" ("Work Science - Evaluation and follow-up processes within working life") deals with evaluation and quality assessment:

"Utvärdering, uppföljning och kvalitetssäkring är begrepp som fått stor betydelse inom arbetslivet. Innebörden hos dessa begrepp och relationerna mellan dem belyses i kursen. Modeller för utvärdering av hela sektorer eller av avgränsade program diskuteras tillsammans med olika redovisningsformer."

It consists of two sub courses which both focus on evaluation; "Utvärdering – i teori och praktik" ("Evaluation – in theory and practise") and "Process och resultat" ("Process and outcomes").

The course is offered by the School of Social and Health Sciences ("Sektionen för Hälsa och Samhälle"). It gives 15 course credits and runs part time (50 %). The sub courses give 7.5 course credits each.

It is a course on the foundation level. There are no specific admission requirements.

The first sub course, "Evaluation - in theory and practice", has got characteristics of the second ideal type of evaluation courses; evaluation within the one field or discipline - in this case work science:

> "Delkursen inleds med en orientering om fenomen som utvärdering, uppföljning, utredningar och kvalitetssäkring inom arbetslivet, hur dessa kan relateras till varandra samt om det vetenskapliga förhållningssättets betydelse."

However, it has also got characteristics of a transdisciplinary course, that is of the fourth ideal type of evaluation courses:

> "Även etiska frågor och utvärderingar ur ett jämställdhetsperspektiv tas upp."

Also the second sub course, "Process and outcomes", has got characteristics of the second ideal type of evaluation courses; evaluation within one discipline or field - in this case work science:

> "Delkursen innefattar genomförande av en utvärdering i arbetslivet, med lärarledd handledning, inom ett företag, en organisation eller en myndighet. [---] Kursen ingår i det arbetsvetenskapliga kunskapsområdet och innehåller teoretisk breddning såväl som praktisk tillämpning med inriktning på utvärdering och uppföljning i arbetslivet."

Theories are mentioned, but it is unclear what kind of theories are referred to:

> "Efter avslutad delkurs förväntas studenten kunna: beskriva och förklara centrala begrepp, teorier och problemställningar samt motivera och redogöra för olika strategier och metoder att genomföra utvärderingar och uppföljningar."

2.3. Linkoping University.

2.3.1. Socionomprogrammet.

Socionomprogrammet at Linkoping University has social work as its main theme area. Evaluation however is one of the three main themes of one of the sub courses - "Utvärdering, kvalitet och kunskapsproduktion i socialt arbete". The course constitutes half the sixth semester of the programme (15 credits). It is not said that the gained skills from this course should be implemented in other courses of the programme.

To teach evaluation is only one of several goals of the programme:

> "För socionomexamen skall studenten
> - visa sådan färdighet och förmåga som krävs för att utveckla och genomföra socialt arbete på olika nivåer i samhället i samarbete med de människor som berörs,
> - visa förmåga att tillämpa relevanta författningar i synnerhet inom det sociala området,
> - visa förmåga att förstå, utreda och analysera sociala processer och problem,
> - visa förmåga att identifiera, strukturera, utreda och utvärdera insatser på individ-, grupp- och samhällsnivå."

The programme is offered by the Department of Social and Welfare Studies (ISV). A Bachelor of Science in Social Work can be completed with 210 credits, a Master with 240. It runs daytime full-time. The specific admission requirements are that the student must have finished the foundation courses in mathematics and political science of the upper secondary school (matematik A och samhällskunskap A).

The course "Utvärdering, kvalitet och kunskapsproduktion i socialt arbete" has got characteristics of the second ideal type of evaluation courses - evaluation within the one field or discipline, in this case evaluation within the field of social work.

2.4. Lund University.

2.4.1. Socialt arbete med alkohol- och drogproblem.

The goal of the course "Socialt arbete med alkohol- och drogproblem" is that the students shall gain knowledge on social work with alcohol and drug problems:

> "Kursens mål är att deltagarna ska förvärva fördjupade kunskaper om socialt arbete med alkohol- och drogproblem."

That is, evaluation is not a main theme of the course. However, one of the tools taught to further development of knowledge on social work with alcohol and drug problems is evaluation:

> "Efter genomgången kurs ska deltagarna kunna
> identifiera och reflektera över egna förhållningssätt, värderingar och attityder till alkohol- och droganvändning
> förstå och analysera konsumtion av alkohol och droger i olika vetenskapsteoretiska perspektiv
> självständigt kunna formulera frågor, söka och kritiskt granska utvärderingar, forskning och annan relevant information rörande interventioner vid alkohol- och drogproblem
> beskriva och analysera socialarbetar- och klientrollerna i organisationers arbete med alkohol- och drogfrågor
> analysera och diskutera interventioner som organisatoriskt och samhälleligt fenomen"

This means that the course resembles the first ideal type of evaluation courses the most; courses that use evaluation as one tool among others to create knowledge on a certain object.

The course is offered by the School of Social Work (Socialhögskolan). It gives 15 course credits and runs part time (50 %). The specific admission requirement is that the degree "socionomexamen" has been completed.

2.4.2. Socialt arbete med barn och unga.

The goal of the course "Socialt arbete med barn och unga" is that the students shall gain knowledge of theories on children's development under different living conditions. That is, evaluation is not a main theme of the course. However, the course shall also provide tools for further reflection. Evaluation is one such tool:

> "Kursen avser att ge fördjupade kunskaper i teorier, svensk och internationell forskning, barns utveckling under olika livsvillkor samt om socialt arbete med barn och deras familjer på olika nivåer och i olika former. Den avser vidare att ge verktyg för reflektion kring varierande yrkesrelaterade frågeställningar."

The course consists of two sub courses of which the first deals with evaluation:

> "Moment 1: Barn, unga, familjer och socialt arbete
> Momentet innehåller en fördjupning av olika teoretiska perspektiv och forskningsbaserade kunskaper liksom träning i att söka och kritiskt granska utvärderings- och forskningsinformation. Det ger därtill träning i att formulera, besvara och kritiskt reflektera kring yrkesrelaterade problemsituationer med hjälp av denna information."

This means that the course resembles the first ideal type of evaluation courses the most; courses that use evaluation as one tool among others to create knowledge on a certain object.

The course is offered by the School of Social Work (Socialhögskolan). It gives 15 course credits and runs part time (50 %). The specific admission requirement is that the degree "socionomexamen" has been completed.

2.5. Stockholm University.

2.5.1. Socialt arbete II.

The goal of the course "Socialt arbete II" is to elaborate on the knowledge of the students on the prerequisites and limits of social work as well as the knowledge base of practical social work:

> "Kursens mål är att fördjupa kunskaperna om ramarna för socialt arbete samt om det praktiska sociala arbetets kunskapsbas."

However, as a part of this one of the purposes of the course is to provide knowledge on evaluation of social work:

> "Kursens syften är
> att ge kunskaper om skilda kunskapsteoretiska ansatser i socialt arbete
> att ge kunskaper om rättigheter och rättsliga strukturer i socialt arbete
> att ge fortsatta kunskaper om det praktiska sociala arbetets teoribildning
> att ge grundläggande kunskaper om utvärdering av socialt arbete"

Evaluation is dealt with in and is the main theme of one of the sub courses, "Utvärdering av socialt arbete", which provides 7,5 credits.

"Socialt arbete II" is offered by The Institute for Social Work. It gives 30 course credits and runs full-time.

It is a course on the foundation level. The specific admission requirement is a foundation course in social work of 30 course credits (Fristående grundkurs i socialt arbete, 20 poäng, eller Socialt arbete 1, 30 högskolepoäng).

The sub course "Utvärdering av socialt arbete" resembles the second ideal type of evaluation courses, as it deals with evaluation in one discipline - social work:

> "Krav på kvalitet, på uppföljning och utvärdering förekommer numera inom den lagstiftning som reglerar olika vårdsektorer. Olika syften för utvärderingsinsatser förekommer, olika modeller prövas och viss utvärdering av utvärdering har genomförts. Momentet innehåller en överblick över de olika utvärderingsuppdrag som kan förekomma och ramarna för dessa, exempel på hur olika utredare/forskare har löst uppgifter samt diskussioner om hur kunskap från utvärderingar kan bidra till kunskapsutveckling i socialt arbete."

2.5.2. Social utredning och dokumentation.

The course "Social utredning och dokumentation" deals with different aspects of and methods to deal with case processing and documentation, of which evaluation is one:

> "Kursen behandlar
> den rättsliga grunden för ärendehandläggning,
> offentlighet och sekretess inom privat och offentlig verksamhet,
> etiska förhållningssätt och bemötande i kontakten med de enskilda, anhöriga samt andra professioner,
> teorier och metoder för uppföljning och utvärdering i social utredning och dokumentation."

"Social utredning och dokumentation" is offered by The Institute for Social Work. It gives 7.5 course credits and runs part time (50 %).

It is a course on the foundation level. The specific admission requirement is that the student has got at least two years experience of work in the social service.

The course resembles the first ideal type of evaluation courses; evaluation is taught as one tool among others to deal with case processing and documentation.

Theories for evaluation in case processing and documentation are taught in the course (see above).

2.5.3. Socialt arbete; kunskapsbas och interventioner.

Evaluation, evidence-based research and a meta analysis together constitute one of the four main themes of the course "Socialt arbete; kunskapsbas och interventioner":

> "Kursen behandlar
> teoretiska perspektiv på det sociala arbetets utveckling i Sverige och internationellt, särskilt med avseende på systematisk uppbyggnad av ämnets kunskapsbas,

vetenskapliga strategier för att inhämta kunskaper om det sociala arbetets villkor, klienter och
interventioner,
utvärdering, evidensbasering och metaanalys som modeller för att utveckla det sociala arbetets
praktik och kunskaper om interventioners effekter,
systematiska kunskapsöversikter på det sociala arbetets centrala områden."

"Socialt arbete; kunskapsbas och interventioner" is offered by The Institute for Social Work.
It gives 7.5 course credits and runs part time (50 %).

It is a course on the advanced level and the specific admission requirement is that the bachelor
consists of at least 90 course credit in social work or in "relevant disciplines":
”Särskild behörighet: Examen på fil kand-nivå med minst 90 högskolepoäng i socialt arbete eller i
relevanta ämnen.”

The course resembles the second ideal type of possible evaluation courses, as it deals with
evaluation in one discipline - social work.
”Kursen behandlar
[---]
utvärdering, evidensbasering och metaanalys som modeller för att utveckla det sociala arbetets
praktik och kunskaper om interventioners effekter,”

2.5.4. Utvärdering och utveckling i socialt arbete.
The course "Utvärdering och utveckling i socialt arbete" is a part of the third semester of
"Socionomlinjen med inriktning mot socialpedagogik". It focuses entirely on different
aspects of evaluation:
"En orientering i utvärderings historik, internationellt och i Sverige, samt en introduktion till hur
olika utvärderings- och utvecklingsmetoder kan relateras till forskning om välfärd, politik och
socialt arbete på administrativ nivå samt i klientarbete med grupper och individer. Följande
delmoment ingår:
Utvärderingens historik. Nuvarande och framtida utveckling avseende brukarmedverkan och
intressentperspektiv.
Genomgång av modeller av och syften med utvärdering och utvecklingsarbete, med betoning på
samspel mellan olika modeller och samverkan mellan olika aktörer inom fältet.
Granskning av begrepp: Kvalitetsmätning, brukarutvärdering, evidensbaserat socialt arbete.
Planering av ett utvärderings- och/eller utvecklingsprojekt avseende syfte, urval av data, val av
lämpliga utvärderingsmetoder samt resultatets användbarhet."

"Utvärdering och utveckling i socialt arbete" is offered by The Institute for Social Work as a
part of the third semester of "Socionomlinjen med inriktning mot socialpedagogik". It gives
7.5 course credits. The specific admission requirement is that a foundation course has been
completed: "Godkänd humanistisk-samhällsvetenskaplig grundkurs."

The course resembles more than one ideal type of evaluation courses. First, the course
resembles the second ideal type of evaluation courses as it deals with evaluation in one
discipline - social work:
”Mål
[---] Kursen syftar också till att utveckla tillräckliga färdigheter för att kritiskt granska och bedöma
utvärderingar och projektplaner inom området socialt arbete.”
Second, the course resembles the third ideal type of evaluation courses, multidisciplinary
evaluation courses, as it deals with the relation between evaluation methods and themes that
are relevant for evaluators of several disciplines, for example political science, sociology,
social psychology and social work:
"samt en introduktion till hur olika utvärderings- och utvecklingsmetoder kan relateras till
forskning om välfärd, politik och socialt arbete på administrativ nivå samt i klientarbete med
grupper och individer."

Third, the course resembles the fourth ideal type of evaluation courses; transdisciplinary evaluation courses:

> "Innehåll
>
> [---]
>
> Granskning av begrepp: Kvalitetsmätning, brukarutvärdering, evidensbaserat socialt arbete."

Fourth, it deals with meta evaluation. That is, it resembles the fifth ideal type of evaluation courses:

> "Innehåll
>
> En orientering i utvärderings historik, internationellt och i Sverige,
>
> [---]
>
> Genomgång av modeller av och syften med utvärdering och utvecklingsarbete, med betoning på samspel mellan olika modeller och samverkan mellan olika aktörer inom fältet."

2.6. Umeå University.

2.6.1. Biblioteks- och informationsvetenskap – kandidatprogram.

The goal of the bachelor programme "Biblioteks- och informationsvetenskap – kandidatprogram" is that the students shall learn to analyse different ways of treating information:

> "Det övergripande målet för utbildningen är att de studerande ska utveckla förståelse för och förmåga att analysera olika former för hantering av information. Utgångspunkten är informationsförmedling och de sammanhang där information produceras, förmedlas och används"

That is, to teach evaluation is not one of the overarching goals of the programme. However, a full compulsory semester is dedicated to a course in evaluation, "Utvärdering av informationstjänster".

The programme is offered by the Department of Sociology (Sociologiska institutionen) and runs full-time. The specific admission requirement is that 90 course credits have been completed. The programme gives 90 additional course credits. The course "Utvärdering av informationstjänster" gives 30 course credits.

That course consists of two sub courses, "Utvärderingsmetodik" and "Självständigt uppsatsarbete". The goals of the course are:

> "Den studerande förväntas ha goda kunskaper och färdigheter att kartlägga och analysera informationstjänster och informationsbehov, samt att självständigt rapportera resultatet av sådant arbete."

That is, no information is supplied on how evaluation is dealt with in the course. However, the *programme* can be said to have the characteristics of the first ideal type of evaluation courses; courses that use evaluation as one tool among others to create learning about a certain object - in this case routines for dealing with information.

2.6.2. Utvärdering.

The course "Utvärdering" ("Evaluation") focuses solely on the theories and practices of evaluation:

> "Kursens övergripande mål är att ge fördjupade kunskaper i utvärderingens teori och praktik."

The course is offered by the Department of Social Work (Institutionen för Socialt Arbete) in cooperation with Umeå Centre for Evaluation Research (UCER, Centrum för utvärderingsforskning). It gives 15 course credits and runs part-time (50 %). It is a course on the advanced level.

The course has got characteristics of the fifth ideal type of evaluation courses; meta evaluation courses:

> "I kursen behandlas de sammanhang där utvärderingar genomförs, samt följder av utvärderingar."

It has also got characteristics of the third ideal type of evaluation; multidisciplinary evaluation courses:

> "Vidare studeras trender i utvärdering, olika modeller och metoder i utvärdering. Främst behandlas policy- och programutvärderingar, dvs. utvärderingar av reformer, policy, program, projekt och verksamheter."

Finally, the course has got parts which are rather of a transdisciplinary nature; that is they resemble the fourth ideal type of evaluation courses:

> "Kursen innehåller också delar om utvärderingskriterier, processer, utvärderarroller, etikfrågor, [...]."

Evaluation theories are mentioned, although not defined:

> "Kursen innehåller inslag om utvärderingsteori, metod och praktik."

2.7. Uppsala University.

2.7.1. Kandidatprogram för social omsorg om äldre och funktionshindrade.

The bachelor program "Kandidatprogram för social omsorg om äldre och funktionshindrade" deals with evaluation as a tool for investigation of social services:

> "Efter avslutad utbildning förväntas studenten
> [...]
> - ha den kunskap som krävs för att kritiskt granska den dokumentation som produceras inom området social omsorg (forskning, utredningar och utvärderingar) samt själv kunna initiera och genomföra undersökningar."

The third year includes a course in research and evaluation methods, "Påbyggnadskurs i, forsknings- och utvärderingsmetodik". The gained skills in evaluation can be implemented in the student thesis written in the third year.

The programme is offered by the Department of Sociology (Sociologiska institutionen). It gives 180 course credits and runs full-time. The main subject is sociology (105 course credits). The remaining 75 course credits are a dedicated to political science, psychology, economics and medicine. The specific admission requirement is that the first Upper Secondary School courses in mathematics and politics have been completed.

The course dealing with evaluation gives 7.5 course credits. It is offered by the Department of Sociology.

Evaluation is taught as one tool among others for supervision of social care. That is, it resembles the first ideal type of evaluation courses; courses dealing with evaluation of a certain object:

> "Sedan följer en påbyggnadskurs i forsknings- och utvärderingsmetodik. Du skall efter denna kurs ha den kunskap som krävs för att ta del och kritiskt granska den dokumentation som produceras inom området social omsorg samt själv kunna initiera undersökningar. Detta får du även möjlighet att göra inom programmets följande kurs som utgörs av ett examensarbete. Du har nu möjlighet, att utifrån ett sociologiskt perspektiv och under handledning, ytterligare fördjupa dig inom den sociala omsorgens område."

2.8. Växjö University.

2.8.1. Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik.

The course "Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik" ("Social work - Current research, Scientific theory and Research methodolgy") deals with research and evaluation methods in the social sciences, and with a focus on social work:

> "Syftet med kursen är att den studerande ska fördjupa kunskaperna om metoder för forskning och utvärdering inom det socialvetenskapliga området och om aktuell forskning med relevans för socialt arbete."

The course is offered by the School of Health Sciences and Social Work (Institutionen för vårdvetenskap och socialt arbete) and is labelled as a course in social work. It gives 15 course credits and runs full-time.

It is a course on the foundation level. The specific admission requirement is that the two first semesters of courses in social work have been taken. It is a part of the programme "socionomprogrammet 140 poäng", but it can also be taken as a single subject course.

One of the objectives of the course is to teach "strategies" for evaluation of research, a theme relevant in all the disciplines under study. That is, the course resembles the third ideal type of evaluation courses:

> "Den studerande ska vidare fördjupa kunskapen om tekniker för forskningsinformation samt utveckla förmågan att kritiskt granska, analysera och utvärdera forskningsinformation av olika slag."

2.9. Örebro University.

2.9.1. Dokumentation och utvärdering i socialt arbete.

The overarching goal of the course "Dokumentation och utvärdering i socialt arbete" is to develop forms for knowledge-based social work. In this evaluation is taught as one tool:

> "Det övergripande syftet med kursen är att utveckla formerna för ett kunskapsbaserat socialt arbete. Det sker bland annat genom att fördjupa deltagarnas kunskaper om utredning, dokumentation och utvärdering som ett instrument för att kritiskt granska och utveckla olika former av socialt arbete på individ- och gruppnivå."

The course is offered by the Department of Behavioural, Social and Legal Sciences (Institutionen för beteende-, social- och rättsvetenskap). It gives 15 course credits and runs part time (50 %).

It is a course on the advanced level. The specific admission requirements are that the student has completed one of the two degrees "Socionomexamen" or "Social omsorgsexamen", including at least 90 course credits in social work, and two years experience of work in the social services.

> "Socionomexamen eller social omsorgsexamen med minst 90 högskolepoäng/60 poäng i socialt arbete samt 2 års yrkeserfarenhet inom socialtjänsten eller liknande verksamhet."

The course has got characteristics of the fifth ideal type of evaluation courses; meta evaluation courses:

> "[Kursen] innehåller moment såsom
> - utvärderingens historik internationellt och nationellt"

However, the course first and foremost deals with evaluation within social work. That is, it resembles the second ideal type of evaluation courses; courses dealing with evaluation within one discipline or field:

> "[Kursen] innehåller moment såsom
> [---]
> - kritisk analys av utvärderingsformer i socialt arbete såsom kvalitetsmätning, brukarutvärdering och evidensbaserat socialt arbete etc.
> - dokumentationens betydelse för ett kunskapsbaserat socialt arbete
> - utvärderingsmodeller och designer som är förekommande i socialt arbete på individ, grupp- och samhällsnivå."

In this it is stressed that the student shall gain the skills needed to conduct evaluations:

> "Den studerande ska också kunna
> - genomföra professionella utredningar i socialt arbete,
> - genomföra utvärderingar på individ-, grupp- och samhällsnivå,
> - dokumentera arbetsprocessen för ett kunskapsbaserat socialt arbete,"

Evaluation theories are mentioned. However, it is not explained what is meant by evaluation theories:

> "För att uppnå lärandemål kommer kursen att bestå av olika moment som innebär att de studerande efter kursens slut fördjupat sina kunskaper om utvärderingsteorier och metoder."

3. Evaluation courses and programmes in psychology.

3.1. Mälardalen University.

3.1.1. Utvärderings- och utredningsteknik.

Evaluation and methods for investigation are the main themes of the course "Utvärderings- och utredningsteknik":

> "Kursens syfte är att den studerande skall förvärva kunskaper i utvärdering, kvalitativ dataanalys och univariat kvantitativ dataanalys av data av den typ som kan vara insamlade inom en verksamhet."

The course is offered by the Department of Social Sciences (Institutionen för samhälls- och beteendevetenskap) and is labelled as a course in psychology. It gives 7.5 course credits and runs part time (50 %).

It is a course on the advanced level. The specific admission requirement is that the student has completed at least 120 course credits, of which 90 have been completed within one field and with increasing depth of study, and of which 15 constitute a student thesis.

The course has got characteristics of the third ideal type of evaluation courses - multidisciplinary evaluation courses - as it teaches how evaluation can be done, no matter what sector:

> "Kursen tar i första hand upp utvärdering och utredning på verksamhetsnivå, i första hand, men även på grupp och individnivå."

However, it has also got characteristics of the fifth ideal type of evaluation courses; meta evaluation courses:

> "Syften med och de potentiella effekterna av utvärdering diskuteras."

3.2. Stockholm University.

3.2.1. Rättspsykologi I.

The main purpose of the course "Rättspsykologi I" ("Forensic psychology II") is to provide basic knowledge in criminal psychology:

> "förmedla grundläggande teoretiska, empiriska och metodologiska kunskaper i Rättspsykologi, dvs den del av psykologin som integrerar det psykologiska kunskapsområdet i det juridiska och kriminalpolitiska sammanhanget."

That is, evaluation is not a main theme of the course. Neither is it in the title of any of the sub courses:

> Grundläggande kunskaper i rättspsykologi, 4 poäng
> Tillämpade rättspsykologiska frågeställningar, 4 poäng
> Fördjupning, 2 poäng

However, it is discussed as one tool that is used in criminal psychology:

>"genom kliniskt och vetenskapligt underbyggd sakkunskap ge deltagarna förutsättningar för att
> förstå, bemöta och utvärdera vittnens, offers och gärningsmäns beteende i brottssammanhang."

The course is offered by the Department of Psychology. It gives 15 course credits and runs
part time (50 %).

It is a course on the foundation level. There are no specific admission requirements, but
upper secondary school courses in psychology recommended.

The course has got characteristics of the first ideal type of evaluation courses; evaluation is
taught one tool among others to be used for understanding reactions in a certain context, that
is when a crime is committed.

3.3. *Uppsala University.*

3.3.1. Beteendevetenskapligt kandidatprogram.

The fifth semester of the bachelor program "Beteendevetenskapligt kandidatprogram" is fully
dedicated to a course in evaluation, "Utvärdering B":

>"På femte terminen, före den avslutande C-kursen, läser samtliga studenter inom programmet en
> specialinrättad ämnesgemensam termin i utvärdering, Utvärdering B. Denna termin innefattar
> "perspektiv på utvärdering, metoder för utvärdering så som intervju, enkätkonstruktion, orientering
> om test samt ett avslutande projektarbete".

Furthermore, the knowledges obtained during that semester shall be implemented in the main
disciplines of the programme:

>"Det beteendevetenskapliga kandidatprogrammet syftar till att ge en bred bas inom de
> beteendevetenskapliga ämnena pedagogik, psykologi och sociologi samt ge en tillämpning av
> ämnena inom området utvärdering."

The students are supposed to learn about methods of evaluation and how to conduct
evaluations:

>"Färdighet och förmåga
> Efter avslutad utbildning från det beteendevetenskapliga kandidatprogrammet skall studenten
> [...]
> - redovisa kunskap om beteendevetenskapliga teorier och metoder för utredning och utvärdering,
> - kunna planera, leda och genomföra förändrings- och utvärderingsarbete."

The programme is offered by the Department of Psychology (Institutionen för psykologi). It
gives 180 course credits and a bachelor's degree. It runs full-time.

It's a programme on the foundation level. The specific admission requirement is that the
second secondary upper school course in mathematics and the first course in politics have
been completed.

With the information at hand it is difficult to say whether the programme resembles the
second or the third ideal type of evaluation courses the most (that is, whether it should be
classified as a transdisciplinary evaluation course or as a course that mainly deals with
evaluation within one discipline or field). That depends on whether the evaluation methods
taught are used mainly in the disciplines dealt with in the programme or not.

3.3.2. Socialt omsorgsprogram med inr mot barn och ungdom.

The last year of the four year programme "Socialt omsorgsprogram med inr mot barn och
ungdom" includes the course "Utvärdering" ("Evaluation").

The programme is offered by the Department of Psychology (Institutionen för psykologi). It gives 160 course credits. It runs full-time.

It's a programme on the foundation level. The specific admission requirement is that the second secondary upper school course in mathematics and the first course in politics have been completed.

The course "Utvärdering" gives 7.5 course credits. It is offered by the Department of Psychology and can only be taken by programme students.

4. Evaluation courses in political science, peace and development research.

4.1. Göteborg University.

4.1.1. Analys och utvärdering av fredsarbete.

The course "Analys och utvärdering av fredsarbete" has got peacekeeping work and evaluation as its two main themes. The first part of the course teaches nonviolence methods for peacekeeping work. This theme is interconnected with evaluation during the second half of the course; a field study in a conflict area:

> "Under studieresan till ett konfliktområde, som deltagare eller observatör, enskilt eller i grupp, får du möjlighet att följa och utvärdera en social aktion, vars syfte är att reducera våldet."

It is offered by the Department of Peace and Development Research. It gives 15 course credits, is Internet-based and runs on part-time (50 %).

It is a course on the foundation level, however with the admission requirement that the student has finished the course "Fred i teori och praktik (15 hp)" or a corresponding one.

The course comes closest to the second ideal type of evaluation courses as it deals with the question how evaluation is performed within one discipline, in this case peacekeeping work.

4.1.2. Utvärderingsmetoder.

The course "Utvärderingsmetoder" focuses solely on evaluation:

> "Kursens övergripande mål är att studenterna ska förvärva grundläggande kunskaper om utvärderingsområdet, med fokus på det 'smörgåsbord' av metoder som står till buds."

It is offered by the School of Public Administration. It gives 7,5 course credits and runs daytime.

It is a course on the foundation level, however only available for students of Förvaltningsprogrammet. It constitutes the fifth semester of that programme

The course mainly deals with evaluation methods, no matter in what discipline or field they are used. That is, it resembles the third ideal type of evaluation courses the most; a multidisciplinary course.

> "Denna kurs är tänkt som en intensiv introduktionskurs till det mångfacetterade och flervetenskapliga område som med en gemensam term kan kallas utvärdering. [---] Kursens övergripande mål är att studenterna ska förvärva grundläggande kunskaper om utvärderingsområdet, med fokus på det 'smörgåsbord' av metoder som står till buds. Med denna kunskap i bagaget ska de kunna analysera fördelar och nackdelar med olika utvärderingsmetoder samt kritiskt kunna granska befintliga utvärderingar av olika slag."

However, the course has also to some extent got characteristics of a transdisciplinary course, that is it to some extent resembles the fourth ideal type of evaluation courses:

> "De ska också ha en viss insikt i olika utvärderingsmetoders och -modellers vetenskapsteoretiska grundvalar."

Finally, the course has also got the character of a meta evaluation course, that is it has got characteristics of the fifth ideal type of evaluation courses as well:

> "I det första momentet *Nyckelbegrepp och modeller* behandlas utvärdering som företeelse och begrepp, dvs hur och varför utvärdering blivit så centralt i offentlig förvaltning samt på vilka sätt begreppet tolkats i svensk och internationell litteratur."

4.1.3. Implementering och utvärdering.

Evaluation is one of the three main themes of the course "Implementering och utvärdering":

> "Kursens övergripande mål är att studenterna skall få insikter i moderna och klassiska teorier inom implementerings- och utvärderingsforskning, samt i problem förknippade med granskningssamhällets framväxt."

It is offered by the School of Public Administration. It gives 15 course credits and runs daytime.

It is a foundation course and it is only available for students of Förvaltningsprogrammet. It is one of the electable courses of the seventh semester of that programme. One of the admission requirements is that the course "Utvärderingsmetoder" has been taken.

The course has got characteristics of the third ideal type of evaluation courses (multidisciplinary evaluation courses) as it teaches evaluation theories no matter in what discipline they occur:

> "Kursens övergripande mål är att studenterna skall få insikter i moderna och klassiska teorier inom implementerings- och utvärderingsforskning,"

The course also deals with the role of evaluation in society, that is with meta evaluation. In that it resembles the fifth ideal type of evaluation courses:

> "Här problematiseras också de senaste två decenniernas institutionella omvandling av den offentliga sektorn som innebär att vi på många håll i världen finner ett granskningssamhälle med stora och ökande krav på kontroll, prövning, jämförelse och bedömning. Centrala frågor är varför denna institutionella omvandling skett, vad den tar sig för konkreta uttryck och vad som kommer ut av den."

It should be noted that evaluation *theories* are mentioned:

> "Kursens övergripande mål är att studenterna skall få insikter i moderna och klassiska teorier inom implementerings- och utvärderingsforskning, samt i problem förknippade med granskningssamhällets framväxt."

4.1.4. Implementering och utvärdering.

The course "Implementering och utvärdering" mainly deals with two themes - implementation and evaluation of politics:

> "Den övergripande fråga som står i centrum i kursen är hur politik genomförs, organiseras och kontrolleras. Frågan behandlas genom att först diskutera implementeringsteori samt den forskning med andra beteckningar som också problematiserar genomförandet av offentliga välfärdsprogram. Därefter diskuteras forskning *om* utvärdering, en forskning som utmanar den traditionella bilden av utvärdering som en fråga om metodisk kompetens för att granska resultat i offentlig sektor. Kursens övergripande mål är att studenterna skall få insikter i moderna och klassiska teorier inom implementerings- och utvärderingsforskning, samt i problem förknippade med granskningssamhällets framväxt."

It is offered by the School of Public Administration. It gives 15 course credits and runs daytime.

It is a course on the advanced level and it is a part of the programme "Masterprogrammet i offentlig förvaltning", which gives 120 course credits and a master degree. That programme has got three different profiles, of which the student is to choose one. The course "Implementering och utvärdering" constitutes half the first semester of the profile "Policyanalys". The specific admission requirement is that the course "Politiska styrmedel", also a part ofthe profile "Policyanalys", has been completed.

The course has got characteristics of the fifth ideal type of evaluation courses (a meta evaluation course) as it deals with the consequences of evaluation:

> "Därefter introduceras forskning *om* utvärdering som utmanar den traditionella bilden av utvärdering som en fråga om metodisk kompetens för att granska resultat i offentlig sektor. En aspekt i denna forskning är vem och vilka värderingar som ska ligga till grund för utvärdering. En annan aspekt handlar om vad användning av utvärdering innebär, hur användning och lärande kan främjas, samt vilka icke avsedda bieffekter utvärdering kan skapa. Nära knutet till frågan om användning är vilken konstitutiv makt som ligger i såväl utvär-deringars användning som i själva utvärderingsprocessen."

Furthermore, it is said that the students shall learn how to conduct theoretically foundered evaluations. In this the course has got the character of a multidisciplinary evaluation course, that is of the third ideal type of evaluation courses, as no distinktion is made between theories used in different disciplines:

> ▫ planera och genomföra teoretiskt grundade utvärderingar, samt kommunicera utvärderingsresultat på ett sätt som främjar användning och lärande.

Evaluation theory is mentioned in the syllabus. The student is supposed to learn how to

> ▫ värdera och analysera centrala aspekter inom utvärderingsteori såsom värdering, metodologi, användning och lärande;

What is meant by evaluation theory is specified in the presentation of the content of the course:

> "Därefter introduceras forskning *om* utvärdering som utmanar den traditionella bilden av utvärdering som en fråga om metodisk kompetens för att granska resultat i offentlig sektor. En aspekt i denna forskning är vem och vilka värderingar som ska ligga till grund för utvärdering. En annan aspekt handlar om vad användning av utvärdering innebär, hur användning och lärande kan främjas, samt vilka icke avsedda bieffekter utvärdering kan skapa. Nära knutet till frågan om användning är vilken konstitutiv makt som ligger i såväl utvärderingars användning som i själva utvärderingsprocessen. Dessa och andra aspekter av utvärderingsteorin behandlas i denna kurs."

4.2. Halmstad University.

4.2.1. Utrednings- och utvärderingsmetodik.

The course "Utrednings- och utvärderingsmetodik" ("Methods of Investigation and Evaluation") deals entirely with investigation and evaluation, and mainly with investigation and evaluation in state and municipal organisation. It focuses on the practices of the public bureaucracy and their relation to public governance and learning by investigation and evaluation. Furthermore, it focuses on networks and partnerships as methods of implementation, how political decisions are made, as well as investigation and evaluation methods.

> "Kursen syftar till att utveckla kunskaper och praktiska färdigheter rörande utrednings/utvärderingsarbete i stat och kommun. Den behandlar frågor om förvaltningens arbetssätt i förhållande till politisk styrning och hur förvaltningen bidrar till politikens lärande genom utrednings- och utvärderingsinstrument. Viktiga kursmoment omfattar förståelse av politiskt beslutsfattande, arbete i nätverk/partnerskap samt utrednings/utvärderingsmetodik."

The course consists of four sub courses of which the latter three focus mainly on evaluation. They are labelled "Förvaltningspolitik i teori och praktik" ("Administrative policy in theory and practice"), "Styrning, kontroll och politiskt lärande" ("Steering, control and political learning"), "Introduktion till utrednings- och utvärderingsmetodik" (Introduction to methods of investigation and evaluation") and "Fältstudier" ("Field studies").

The course is offered by the School of Social and Health Sciences ("Sektionen för Hälsa och Samhälle"). It gives 30 course credits and runs full-time. Each of the sub courses give 7,5 course credits.

It is a course on the advanced level. Apart from a bachelor or a corresponding degree there are no specific admission requirements.

The course has got characteristics of the second ideal type of evaluation courses; evaluation within one discipline or field - in this case evaluation in public administration:
> "Kursen syftar till att utveckla kunskaper och praktiska färdigheter rörande utrednings/utvärderingsarbete i stat och kommun."

These questions are focused upon in the second of the four sub courses, "Steering, control and political learning":
> "Delkursen utgår från begreppet politiskt lärande och fokuserar intresset mot principiella problem att utreda, utvärdera och kontrollera offentlig politik i en komplex flernivåsituation. Grundfrågan handlar om hur politiska organisationer tar lärdom av utrednings- och utvärderingsarbete."

The third sub course, "Introduction to methods of investigation and evaluation", has got characteristics of a multidisciplinary evaluation course, that is of the third ideal type of evaluation courses:
> "Delkursen behandlar grundläggande perspektiv och metoder rörande kvalificerat utrednings- och utvärderingsarbete. Det gäller dels kvalitativa tekniker såsom intervjuteknik, observation, och dokumentanalys samt dels kvantitativa tekniker med hjälp av statistiska analyser."

However, this sub course has also got characteristics of the fifth ideal type of evaluation courses, that is of meta evaluation courses:
> "Vidare studeras grundläggande perspektiv och teorier rörande utvärdering av politiska reformer."

Evaluation theory is mentioned in the description of the third sub course. Meta evaluation theory seems to be the kind of evaluation theory referred to:
> "Vidare studeras grundläggande perspektiv och teorier rörande utvärdering av politiska reformer."

4.3. Lund University.

4.3.1. Utvärdering av offentlig verksamhet.

The course "Utvärdering av offentlig verksamhet" deals entirely with evaluation and development work within the public sector:
> "Kursen ger dig kunskaper om teorier och metoder kring utvärdering och förändringsarbete inom offentlig sektor för att du själv ska kunna genomföra och kritiskt granska utvärderingar."

That is, evaluation is the main theme of the course.

The course is offered by the Department of Political Science. It gives 15 course credits, is Internet-based and runs part time (50 %).

It is a course on the foundation level. There are no specific admission requirements. The target group is people that are working with evaluation and development work within the public sector:

"Den vänder sig till dem som i sin verksamhet ägnar sig åt utvärdering och därtill kopplat förändringsarbete. Kursen kan också vara lämplig för dem som befinner sig i slutet av sin grundutbildning."

The course has got characteristics of the fifth ideal type of evaluation courses; meta evaluation courses:

"Med utgångspunkt i aktuell forskning behandlas kritiskt när, hur och av vem utvärderingar genomförts."

However, it has also got characteristics of the second ideal type of evaluation courses; courses that deal with evaluation within one discipline or field - in this case public administration:

"Utvärderingarnas roll i förändringsarbete inom den offentliga sektorn diskuteras liksom kriterier och utgångspunkter för utvärderingar. Särskild uppmärksamhet ägnas den offentliga sektorns särdrag som demokrati, rättssäkerhet och politisk organisering."

It should be noted that evaluation theory is mentioned. Theories *about* evaluation, that is meta evaluation theories, seem to be meant:

"Kursen ger dig kunskaper om teorier och metoder kring utvärdering och förändringsarbete inom offentlig sektor för att du själv ska kunna genomföra och kritiskt granska utvärderingar."

4.4. Uppsala University.

4.4.1. Organisering, implementering och utvärdering i offentlig politik.

The course "Organisering, implementering och utvärdering i offentlig politik" focuses on three interrelated themes; organisation of the public sector, implementation of public policy and evaluation. It is rooted in the conception that each one of these three parts of the political system influences the other two:

"Kursen fokuserar särskilt tre moment i den offentliga politiken: organisering, implementering och utvärdering eller granskning. Den gör det utifrån en enkel grundtanke: hur den offentliga politiken är organiserad påverkar hur den genomförs och hur den kan och faktiskt utvärderas. Samtidigt är ett centralt tema att utvärdering och implementeringsprocesser i sig organiserar maktrelationer och styrningsmönster i den offentliga politiken. Lika viktigt som att förstå utvärderingens och implementeringens mer administrativa logiker är därför förståelsen av hur dessa påverkar det politiska systemet i stort."

The course is offered by the Department of Political Science. It gives 7.5 course credits and is a course on the foundation level.

The course resembles the second ideal type of evaluation courses; courses that deal with evaluation within one discipline or field - in this case governance:

"Somliga hävdar rentav att det inte längre är meningsfullt att tala om stater eller politisk styrning. I denna kurs förhåller vi oss kritiskt men intresserat till denna och liknande utsagor. Vi gör det genom att diskutera betingelserna för att organisera, implementera och utvärdera styrningsförsök."

This characteristic is further emphasised in that it also deals with evaluation as a "toolbox" for analysis of the results of implementation of public policy.

However, it has also got transdisciplinary characteristics, as it deals with which criteria should be used in evaluations within the public sector. That is, it in this resembles the fourth ideal type of evaluation courses:

"Utifrån vilka värdekriterier kan och bör offentlig verksamhet och dess resultat bedömas?"

Finally, the course has also got characteristics of a meta evaluation course. That is, it resembles the fifth ideal type of evaluation courses:

"Hur utövas makt i och genom utvärdering och vilka är konsekvenserna av utvärderingstrenden ur ett vidare perspektiv på den offentliga politiken?"

5. Selected required readings.

In the following the required readings, which are used in more than one of the evaluation courses or programmes under study, are listed. In connection with each book or article the disciplines, universities and courses/programmes in which it is used are mentioned.

Denscombe, Martyn (2000), *Forskningshandboken – för småskaliga forskningsprojekt inom samhällsvetenskapen.* Studentlitteratur, Lund.

> Pedagogik:
> Uppsala University, Utvärdering och kvalitetsutveckling.
> Social work, work science and sociology:
> Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Docherty, Peter (1996) *Läroriket: vägar och vägval i en lärande organisation.* Stockholm: Arbetslivsinstitutet.

> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program Evaluation. Alternative Approaches and Practical Guidelines* (3rd ed.). Boston: Pearson.

> Pedagogik:
> Mälardalen University, Utvärdering B.
> Social work, work science and sociology:
> Umeå University, Utvärdering.

Granberg, Otto. (2004). *Lära eller läras: om kompetens och utbildningsplanering i arbetslivet.* Lund: Studentlitteratur.

> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

Karlsson, O. (1999). *Utvärdering – mer än metod. Tankar och synsätt i utvärderingsforsk-ning: en översikt.* Àjour - en serie kunskapsöversikter från Svenska kommunförbundet, nr 3. Stockholm: Svenska kommunförbundet, Kommentus.

> Pedagogik:
> Mälardalen University, Utvärdering B.
> Mälardalen University, Utvärdering ur samhällsvetenskapligt perspektiv D.
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.
> Göteborg University, Kvalitet och utvärdering inom utbildning.
> Political science and peace and development research:
> Lund University, Utvärdering av offentlig verksamhet.

Karlsson, Ove (2003), "Evaluation politics in Europe: trends and tendencies". *Studies in Educational Policy and Educational Philosophy* 2003:1. E-tidskrift.

> Social work, work science and sociology:

Stockholm University, Utvärdering och utveckling i socialt arbete.
Political science and peace and development research:
Lund University, Utvärdering av offentlig verksamhet.

Lundmark, A. (1998). *Utbildning i arbetslivet. Utgångspunkter och principer för planering och genomförande av personalutbildning.* Lund: Studentlitteratur.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (valda delar c:a 150 s).
Lund University, Utrednings- och utvärderingsmetod och verksamhetsanknutet projektarbete.

Morén, Stefan & Blom, Björn (2003) Insatser och resultat. Om utvärdering i socialt arbete. (Rapport nr 48) Umeå: Umeå universitet, institutionen för socialt arbete.
Social work, work science and sociology:
Umeå University, Utvärdering (sid 5-116).
Örebro University, Dokumentation och utvärdering i socialt arbete.

Pommer Nilsson, Peter (2003). *En introduktion till pedagogik i arbetslivet. Begrepp, historik, nutid.* Pedagogiska institutionen, Umeå universitet, Nr 73.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).

Shaw, I & Lishman, J. (1999). *Evaluation and social work practice.* London: Sage.
Social work, work science and sociology:
Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Socialstyrelsen (2003), *För en kunskapsbaserad socialtjänst.* Redovisning av Regeringsuppdrag, [elektroniska resurs]110 s.
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Socialstyrelsen (2003), *Brukarmedverkan i socialtjänstens kunskapsutveckling,* [elektroniska resurs]*109 s.*
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Socialvetenskaplig Tidskrift (2002), *Temanummer(2-3) om socialt arbete och utvärdering.*
Social work, work science and sociology:
Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).

Socialvetenskaplig Tidskrift. Årgång 9 nr 2-3 2003 [Tidskriftsartikel]
Social work, work science and sociology:

Örebro University, Dokumentation och utvärdering i socialt arbete.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Rombach, B., & Sahlin-Andersson, K. (red.) (2002). *Från sanningssökande till styrmedel. Moderna utvärderingar i offentlig sektor* (2:a uppl.). Stockholm: Santérus.
Pedagogik:
Mälardalen University, Utvärdering B.
Social work, work science and sociology:
Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.
Umeå University, Utvärdering (sid 9-24, 209-227).
Örebro University, Dokumentation och utvärdering i socialt arbete.

Rossi, Peter, Freeman, Howard & Lipsey, Mark (1999) Evaluation. A Systematic Approach. London: Sage.
Social work, work science and sociology:
Umeå University, Utvärdering (sid 37-76,400-403, 421-422).
Political science and peace and development research:
Göteborg University, Utvärderingsmetoder.
Halmstad University, Utrednings- och utvärderingsmetodik.

Vedung, Evert (1998). *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.
Pedagogik:
Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.
Mälardalen University, Utvärdering B.
Mälardalen University, Utvärdering ur samhällsvetenskapligt perspektiv D.
Social work, work science and sociology:
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).
Örebro University, Dokumentation och utvärdering i socialt arbete.
Political science and peace and development research:
Halmstad University, Utrednings- och utvärderingsmetodik.
Lund University, Utvärdering av offentlig verksamhet.
Uppsala University, Organisering, implementering och utvärdering i offentlig politik.

Weiss Carol H. (1998), *Evaluation – Methods for studying programs and policies*(2d ed). Upper Saddle River: Prentice Hall.
Social work, work science and sociology:
Umeå University, Utvärdering.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Örtenblad, Anders. (2005). Of course organizations can learn! *Learning Organization*. Vol. 12 No. 2. p. 213-218.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).
Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs

Section 3: Case study 2 - Research.

1. Introduction.

The original idea for the project "Evaluation roots in the USA and in Europe – tracing traditions?" came up during discussions between Marvin Alkin (University of California, Los Angeles, USA) and Ove Karlsson Vestman (Mälardalen University, Sweden). The question discussed was whether it would be possible to describe the roots (origins) of the European evaluation tradition in a way comparable to Alkin's description of the development of American prescriptive evaluation models. Alkin had earlier, together with Christina Christie, published *Evaluation Roots – tracing theorist's views and influences,* a book on the development of American evaluation theorist's prescriptive models for evaluations. The two authors presented and revised their work in the article presented here as Attachment 2.1.

In early 2006 a group of European evaluators interested in the development of the European evaluation tradition was gathered; Ove Karlsson Vestman, Marvin Alkin, Frans Leeuw (the Netherlands), Nicoletta Stame (University of Roma "La Sapienza", Italy) and Reinhard Stockmann (University of the Saarland, Germany). It was decided that the group should have a panel discussion at the UKES/EES joint conference in London 4-6 October 2006 on the development of the European evaluation tradition. Before that these evaluators came together for a symposium in Stockholm, June 17-19 2006. This symposium was also attended by Jan-Eric Furubo, Evert Vedung and Gustav Jakob Petersson. Jan-Eric Furubo and Gustav Jakob Petersson also took part in the panel discussion in London.

However, the Stockholm symposium was preceded by some preparatory meetings attended by a group of Swedish evaluators; Jan-Eric Furubo, Maria Bergström, Rolf Sandahl, Evert Vedung, Ove Karlsson Vestman and Gustav Jakob Petersson. During the Swedish meetings alternative project ideas came up.

2. The Swedish research group.

Some of the ideas that were discussed within the Swedish project group are summarised in the following. The minutes of the Swedish meetings are attached (Attachment 2.2 – 2.5).

In March several different ideas were discussed. One was not to focus on prescriptive models, as Alkin and Christie had done, but on the American and European evaluation traditions. In this an important definition was made; "evaluation tradition" was henceforth defined as the sum of three objects that would all be relevant for our study; the evaluation discourse (the primarily academic discussion on how evaluations *should* be done), the evaluation practices (how evaluations are actually done) and the institutions (organisations) that do evaluations.

The intention was also to go further and discuss roots, that is, explanatory factors, of the observed differences. As well important traces in the history of thoughts as also major events, for example political turning points, were considered to have important explanatory power for understanding the developments of the European and American evaluation traditions.

Of course it was noted that the exchange of ideas between the two continents has been great. However, this was not conceived as problematic as long as the focus is kept on *differences* between evaluation traditions; the exchange of ideas can explain similarities, but hardly differences.

A problem that was considered to be more serious for a comparison between the United States and Europe is that the evaluation traditions of European countries differ sharply. A closely related problem is that international European magazines on evaluation are written in the English language. Thereby they probably don't mirror the evaluation traditions of all European countries to the same extent, as researchers of other mother tongs do not have the same opportunities as English researchers to get published. Such magazines can therefore not be used to grasp what is typically European.

Because of the differences in evaluation traditions within Europe the discussions of April 2006 came to focus increasingly on the branches of the European evaluation tree. It was suggested that the project should focus mainly on evaluation practises and institutions. Concerning the evaluation practices it was suggested that mainly evaluations of public sector activities should be dealt with.

There was broad agreement within the Swedish group on that the differences within Europe probably are at their clearest when it comes to the institutions for evaluation. It was also believed that in Europe evaluations of public sector services to a greater extent than in the USA are done by public bodies. That is, such evaluations were thought to be done to a greater extent by private companies in the USA than in Europe. If there is such a difference it could perhaps be explained by different conceptions of the role of the state in society.

More precisely, the line of reasoning was the following. State involvement in the distribution of welfare among the citizens seemed to be more controversial in the USA than in Europe. Thereby it seemed logical to assume that also the allocation to public bodies of evaluation of such distribution is more controversial. That evaluation of public sector services seems to be done to a higher degree by private institutions in the USA than in Europe seemed to support these two assumptions.

Perhaps could such a more critical attitude in the USA towards the state also explain a greater use in that country of science-based evaluation methods (evidence-based methods); there is a greater need to legitimise public sector services than in Europe.

However, it was also noted that not all the difficulties of generalising stem from regional or national differences. Differences in evaluation traditions between policy sectors were also thought to be of considerable importance, because of the diverging natures of the activities undertaken in different policy sectors.

It was briefly discussed as an alternative, or a complement, to compare evaluations on the federal level of the United States with EU evaluations.

Another project idea that came up during the Swedish meetings was to focus on prescriptive evaluation models that have been advocated as well in the USA as in several European countries. In this it could be discussed why the models have come to greater use in some countries than in others. This would be another way to focus on the roots of evaluation practices.

The development of evaluation traditions over the years was also discussed and formulated in different metaphors, for example waves of changes and advances of evaluation traditions. However, this theme was not elaborated further.

Several different methodological perspectives that might get useful in fulfilling the tasks discussed were also elaborated on.

3. The Stockholm symposium of June 17-19.

From June 17-19 2006 a symposium on the roots of the European and American evaluation traditions was held in Stockholm at the Swedish National Audit Office. The participants were Ove Karlsson Vestman, Marvin Alkin, Frans Leeuw, Nicoletta Stame, Reinhard Stockmann, Jan-Eric Furubo, Evert Vedung and Gustav Jakob Petersson. Ove Karlsson served as chairperson and Jan-Eric Furubo of the Swedish National Audit Office as host of the symposium.

Ove Karlsson opened the session by expressing hopes that it should start a fruitful cooperation on the roots of evaluation. Thereafter, as a source of inspiration, Jan-Eric Furubo presented the comparative country study that he, Rolf Sandahl and Ray Rist conducted in the *Evaluation Atlas*. The similarities between that project and the beginning one were many. First, also the Atlas-project was an attempt to map evaluation cultures. Second, attempts were made to explain the observed differences.

After this presentation the discussion mainly concerned ideas presented in two papers at hand at the beginning of the symposium. One was written by Petersson and Karlsson and the other by Alkin and Christie.

The Petersson-Karlsson paper.
As mentioned earlier, an idea that had come up during the meetings of the Swedish group was that there might be a correlation between on the one hand which actors are responsible for the distribution of welfare services in society and on the other which actors do evaluations of that distribution. It was also thought that there might be a correlation between *how* the distribution of welfare services is undertaken and *how* such activities are evaluated, for example by which standards and with which methods evaluations are done. As a preparation for the Stockholm symposium and with the discussions of the Swedish group as a starting point Petersson and Karlsson wrote a paper on this theme (here presented as Attachment 2.6). This was the first paper discussed at the Stockholm symposium.

In the paper Gösta Esping-Andersen´s distinction between three different kinds of welfare state models was referred to. In *Three worlds of welfare capitalism* (1990) he makes a distinction between state, market and family-based welfare state models, each one of them named after the institution undertaking or financing the distribution of welfare in societies characterised by that model. There was agreement within the Swedish group that Sweden, according to Esping-Andersen a typical state-based welfare state, is probably characterised by more public evaluations of the distribution of welfare in society than for example Great Britain, by Esping-Andersen defined as a typical market-based welfare state.

However, because of the strong influence of the EU on the national European evaluation traditions it was stressed that the focus must not be exclusively on national traditions.

In the paper a second theme is also elaborated; how the development of different European evaluation traditions can be explained. In this Stein Rokkan's approach in describing the political formation and nation-building processes in Europe is used. His starting points are some critical junctures of the European history, mainly the Reformation, the French

revolution and the Industrial revolution. In Rokkan's view these processes generated contrasting systems of power and opposition, which were "frozen" into some specific party alternatives when electoral democracy was introduced. The idea of the Petersson – Karlsson paper is that a similar critical juncture-approach could be used for understanding the development of different welfare state models. These developments limited the number of possible evaluation traditions.

Comments on the Petersson-Karlsson paper.

The participants of the Stockholm symposium delivered some comments on this paper.

Reinhard Stockmann gathered his and Wolfgang Mayer´s (an associate of Stockmann´s) comments on the Petersson – Karlsson paper in a paper, which is here presented as Attachment 2.8.

They believed that there are general differences between the European and American evaluation traditions. However, they also wanted to highlight national differences within Europe. In this they were mainly positive to using Gösta Esping-Andersen´s typology, although they believed the importance of national traditions to be decreasing because of the period of welfare state reforms in the 1990s and the increasing importance of EU evaluations. They were also positive towards using Stein Rokkan's approach for describing the development of European evaluation traditions.

However, they saw a problem with delivering judgements on *national* evaluation cultures. The reason for this was that day considered the differences between policy sectors to be of great importance as well. Therefore they called for studies of intra-national differences as well. In this they suggested another starting point: "By conducting an evaluation, one has always to answer the following question: What will be evaluated by whom on which criteria with which objectives? These questions can be used to structure the national evaluation field(s). Additionally, the key understanding of evaluation can be added to outline differences between sectors, nations, schools etc."

Nicoletta Stame on the other hand described the European evaluation traditions as being mainly dependent on the America one. She therefore advocated putting European names on the existing branches of Marvin Alkin's evaluation theorists tree and in a few cases add typically European branches, like Pawson and Tilley´s Realistic evaluation. Europe was described as a latecomer, which in the introduction of evaluation could reflect on the US experience. The role of the EU was also emphasised in that some countries have started evaluating their public sector services only with their EU memberships in the 1990s. Stame also pointed out major problems with the EU evaluation systems.

Stame also pointed out a number of differences between the roots of the European evaluation tradition and the ones of the American one. One such important difference that she pointed out is that the centralised state is the rule in Europe, while the United States is organised as a federal state. This is important because federalism enhances quasi-experimentalism. That welfare programmes were introduced later in Europe than in the United States was also mentioned. In this context Stame mentioned that laws in some European countries, like Italy, have played the role of programmes in the United States. This, according to Stame, is important because a new programme normally is evaluated and if it doesn't have the intended results it is abolished. Laws on the other hand are often judged without preceding

evaluations. Yet another mentioned difference between the continents is the more positive attitude in Europe than in the United States towards state interventions.

Concerning the degree of similarity between the evaluation traditions of the two continents Stame believed it to be greater in the discourse branch than in the practice branch.

Frans Leeuw basically agreed on the appropriateness of the three lines of evaluation, that is discourse, institutions and practices, which were presented by Petersson – Karlsson. He believed that a study on theorist's views would be less relevant for Europe than for the USA. He didn't want to totally neglect the importance in Europe of "intellectual heroes" though. Instead he wanted to highlight the question of why institutions for evaluation are more important in some countries than in others and why they are more important in some fields than in others.

He also suggested replacing the term evaluation practices with evaluation systems. The reason for this was that "everything is a practice", but what a "lonely hunter evaluator" is doing is not important. Furthermore, in contrast to Petersson - Karlsson, he was most sceptical to presenting differences in national culture as an explanation for differences in evaluation practices. Esping-Andersen´s approach was thereby strongly rejected. The reason for criticising this approach was that it doesn't present a mechanism. Instead Leeuw emphasized the tendency of isomorphism in the field of evaluation.

The Alkin-Christie paper.
At the Stockholm symposium also Marvin Alkin presented a paper (written by himself and Christina Christie), which dealt with modifications on their book, that is, revisions of the names on the three branches of the evaluation theorists tree – Values, Methods and Use. They too received comments from the other participants (to be found in the minutes of the symposium). One recommendation was from Frans Leeuw and concerned putting Nick Mullens on the evaluation theorists tree. Here Alkin´s and Christie's paper is presented as Attachment 1.1.

The afternoon session.
The afternoon session was mainly spent elaborating on the outlining of the panel discussion to be held at the UKES & EES joint conference in London 4-6 October. A programme for presentations was decided upon.

The minutes of the Stockholm symposium are presented below as Attachment 2.7.

Finally Nicoletta Stame's handout from the Stockholm symposium is presented as Attachment 2.9.

4. The London panel discussion.

The next time that the international roots group got together was at the UKES and EES joint conference in London, 4 - 6 October 2006. As a part of the stream "Evaluation and the European project" a panel discussion was held with the title "Evaluation roots in the USA and in Europe: tracing traditions?" The participants were Ove Karlsson Vestman, Marvin Alkin, Christina Christie, Frans Leeuw, Nicoletta Stame, Reinhard Stockmann, Jan-Eric Furubo and Gustav Jakob Petersson.

During the panel discussion the ideas which had been elaborated at the Stockholm symposium were presented. Alkin and Christie presented the latest version of their evaluation theorists tree. Stame discussed the different roots of the European and American evaluation traditions. Leeuw elaborated on the question whether the contributions of individual researchers have been of greater importance for the development of the American evaluation tradition than for the development of the European one. He also highlighted the need for empirical studies of evaluation systems. Karlsson presented the three branches of the Petersson – Karlsson evaluation tree and Petersson presented their approach for describing the roots of evaluation. Stockmann presented his and Meyer´s sectoral approach and their lead question for conducting empirical investigations: What will be evaluated by whom on which criteria with which objectives? Furubo presented the approach used in the *Evaluation Atlas*.

Thereafter the audience contributed with some opinions and suggestions. For example the question was asked whether the U.S. tree could be supplemented with institutions rather than individuals. Institutions like RAND, Abt Associates and Mathematica were suggested.

In his speech Alkin had described his tree as American centric. However, during the discussion the critique was launched that the alternative approach with its focus on institutions and the political dimensions of evaluation was just as Eurocentric as Alkin's tree was focused on American evaluation theorists. In this discussion it was suggested from the audience that the American academic approach towards evaluation might generally be more prescriptive than the European one. Alkin called for investigations on this point while Stame was most sceptical to the hypothesis.

The discussion was also extended to another continent – Africa – and evaluation of development aid was brought into the discussion.

Notes by Rahel Kahlert from the panel discussion is here presented as Attachment 2.10.

5 After London.

After the UKES and EES joint conference there have been discussions with Elliot Stern, editor of *Evaluation*, on the possibility of publishing some articles on the roots of evaluation in that magazine. The following letter has been sent to Stern:

Dear Eliot,

Thanks for your suggestion that we should submit three articles for review. Of course we want to take on this task. Our plans are outlined under the three headings below.

1. Introduction
The first one of the three articles could be Marvin Alkin´s and Tina Christie´s article on their roots book. In this letter we would like to present the second article.

In the second article we would like to present an alternative to M & C:s individualistic perspective for explaining the development of evaluation practises within organizations, that is how individuals do evaluations. We understand an evaluation practice as answers to the following questions:

- Who is doing the evaluation? Internal or external? Only one individual or a group of persons?
- What is being evaluated?
- Which methods are used in the evaluation?

That is, the development of evaluation practices within organisations is the phenomena to be explained.

This alternative approach is linked to *the neo-institutional research tradition***. The focus within this research tradition on the interaction between institutions and actors is what we find inspiring in explaining the development of evaluation practices within organizations.** This will be further elaborated below.

Under heading two we will present two different definitions of "institution", which are both frequently used within the neo-institutional research tradition. Institutions, defined in these two ways, are the two different explanatory factors that we want to highlight for explaining the development of evaluation practices.

2. The neo-institutional approach for explaining the development of evaluation practices

A. **Evaluation models** defined as **formal institutions**
The first factor for explaining the development of evalution practices within an organization is the **evaluation models** that have been adopted by the organisation under study. Evaluation models are formal rules or instructions on how evaluations *should* be done. These models do not necessarily have to correspond at all to how evaluations are actually done within that organization (the evaluation practices). However, we do think that evaluation models are important factors in forming evaluation practices.

One example of an evaluation model, which is an important factor in forming evaluation practices within organizations, is the prescriptions of the Swedish National Agency of Higher Education on how and how often evaluations should be done in the Swedish policy sector of higher education. The elaborations could go on discussing for example EU standards on how evaluations should be done, standards that probably play a part in forming evaluation practices on the European level.

B. **Evaluation codes** defined as **informal institutions**
The second factor for explaining the development of evalution practices within an organization is the **evaluation codes** that have developed within that organization. Evaluation codes are informal norms on how evaluations should be done. We believe that such evaluation codes play an important part in forming evaluation practices.

That there are such evaluation codes, means that new evaluation models can be adopted within organizations doing evaluations without the evaluation practice being modified. The models have only been adopted on the surface, but really there is business as usual and in accordance with the evaluation codes.

3. Interactions between institutions and actors
What we find inspiring about the neo-institutional discussion is its focus on the interaction between institutions and actors. On the one hand, as stated before, we believe that institutions,

defined the two ways above, are important factors in shaping evaluation practices within organisations. That is, institutions form the actions of evaluators (actors).

However, on the other hand evaluators are the ones to shape the evaluation models and to embody the evaluation codes that form evaluation practices. A clerifying of what this means in practice can be found within the important neo-institutional research field of international *isomorphism*.

Isomorphism refers to the transfer of evaluation models by international emulation/copycatting. Neo-institutional research on isomorphism presents an explanation for the transfer of evaluation models. It is stated that some nations adjust to internationally respected models for evaluation in order to gain certain advantages, like resources from international funds. Even if the fund doesn't state that certain evaluation methods should be used some methods are probably conceived as better than others. Therefore these methods will be used by organizations (for example states) applying for resources from the fund. However, in adopting these models the evaluators often have to adjust them in order to make them coherent with the specific features of their organizations. Furthermore, in their daily work the evaluators modify the evaluation codes of their organizations. That is, evaluators create the institutions, which form the evaluation practices of these very same evaluators.

This interaction between institutions and evaluators is what we would like discuss in article two.

Our suggestion is that the third article should be written from a neo-institutional perspective, but as for now we don't know about who is going to write it.

We would be very glad to have you response as soon as possible.

Section 4: Case study 3 – Debates.

1. Themes of the UKES/EES joint conference.

In the following the themes of the UKES/EES joint conference are presented. The numbers in round brackets indicate at which sessions each theme was discussed. The six streams of the conference were the following:

1. **The role of evaluation in public policy.**

Cluster 1.1. The importance of evaluation for policy-making and policy implementation.

In this stream the connections between public policy-making and evaluation were discussed. In particular it was discussed to what extent and how evaluation can play a role in reforming and improving policy-making and its implementation (1.02, 3.03, 5.03, 8.14), for example when contacted by ministerial inspectorates and national audit offices (5.02). This was discussed in different contexts, for example regional and city development (1.02, 6.01, 8.03) and public health policies (8.06, 8.07). Case studies were brought forth, for example the French state-regions contracts (2.02). It was also discussed whether it is at all possible to evaluate the sustainability of programmes. (7.01). The role of evaluation in societies marked by conflicts, like Northern Ireland, was also discussed (8.02).

To this context belongs the question in what ways evaluation can be used for securing accountability and making sure that the needs of various interest groups are fulfilled (2.01, 4.04, 2.03, 3.02), for example by the use of participatory evaluation (4.03, 7.02). The use of evaluation in control of the implementation of policies was also brought up in another context; tracing unanticipated side effects of environmental policy (2.02).

Cluster 1.2. Methods for data collection and analysis.

Finally, the always relevant question of how data should be gathered and compiled so as to become relevant to the policymakers was discussed (1.03, 1.13). One discussed problem of this field was the causality problem; what criteria should be used for deciding when causality should be assumed (5.01)? Different methodologies for programme evaluation and the use of the counterfactual method were discussed (8.01, 6.03). Also the use of multiple evaluations was discussed (6.02).

2. **Evaluation and its relationship with the professions, policy sectors and disciplines.**

Cluster 2.1. Professions.

In this stream the development of evaluation skills in different professions was discussed (2.04, 3.05), and thereby also which the professional competences of evaluators are (6.06). It was also discussed in what way evaluation can contribute to improvement of the practices in the different professions and spread learning within organisations (2.04, 2.05, 7.05).

Cluster 2.2. Policy sectors.

The question of how integrated local development projects should be evaluated was brought up also in this stream and case studies were brought forth, for example Portugal (3.06, 8.06). Which evaluation methods and systems that are appropriate for the health, social care, environmental and educational policy sectors was discussed at several occasions (4.06, 5.06, 6.06, 7.04, 8.06).

Cluster 2.3. Disciplines and discipline divides.

Evaluation of sustainability requires that discipline divides are transcended, as sustainability policies have as objectives to improve as well environmental as also social conditions. How discipline divides can be transcended in evaluation was therefore a relevant theme to discuss in this stream (3.06). The always relevant question of how and to what extent stakeholders should be involved in evaluation was also dealt with (4.05, 3.06).

3. Evaluation and the European project.

Cluster 3.1. EU evaluation and its purposes.

In this stream the evolution of EU evaluation and which purposes it has today was discussed (1.06). In that context evaluation in the domains of agriculture, rural development, drugs, justice and social inclusion policies were brought up (1.06, 1.07, 3.07).

What purpose EU evaluation can fulfil was also discussed in relation to the constitutional crisis of the EU and the policy of better regulation (1.07, 5.07). Can evaluation help creating new confidence in the European project? The closely related question of how accountability can be secured on the EU level was also discussed (2.06).

Methodological problems of conducting evaluations in the complex setting of the EU were elaborated on. And how should evaluation systems be structured in complex settings like the EU (2.07)? The Leader programme and the Structural funds were brought up as case studies (5.07, 5.08). For example, how should sectoral and transnational "added value" be assessed (6.07)? How should EU legislation be evaluated (6.08)? It was also discussed how EU partnerships should be evaluated (7.06).

And what role can national audit offices play in ensuring effective use of the EU funds (5.08)? Furthermore, which trade-offs are there between equity and effectiveness measurement in the impact assessments of the European commission (6.07)? In one presentation a parallel was drawn to policy evaluation in the federal country of Belgium (5.07). It was also discussed which difficulties and traps there are in using country examinations concerning effectiveness analysis of economic instruments and peer review (8.08).

Cluster 3.2. The Europe of nations and universities.

However, the European project was not only discussed in terms of the EU, but the evolution of national evaluation traditions was also focused upon (3.08). The European project was also discussed from yet another angle; which are the connections between the European university study programmes in evaluation (7.07)?

4. **International evaluation and evaluation in developing countries.**

In this stream the latest challenges and practices of development evaluation were discussed (1.08, 2.08).

Cluster 4.1. Institutions for evaluation.

For example, how should evaluation systems be institutionalised and evaluation organisations developed in developing countries (1.09)? The consequences of strengthening international organisations which conduct evaluations in developing countries, for example the United Nations, was also discussed (4.10). The use of national and international networks in evaluation was discussed (5.10), for example in evaluation utilisation (7.09). How can the voluntary sectors of developing countries be brought into the business of evaluation (2.09). How can evaluation be conducted from the ground up (2.13)?. How can evaluation capacity be developed in developing countries (6.10)?

Cluster 4.2. Learning and accountability.

Also in this stream (like in Stream 1) the question was brought up how learning and accountability can be enhanced by evaluation in particular in developing countries (2.09).

Cluster 4.3. Evaluation objects relevant for developing countries.

Evaluation of budget support was another theme discussed (3.09). Closely related is the theme of evaluation of poverty reduction (3.11, 5.11, 7.09). Case studies were gathered from Egypt, Indonesia and Croatia (4.09). The Tsunami Evaluation Coalition was discussed separately (4.11). The evaluation of the European commission's humanitarian aid activities was also on the agenda (6.11).

Cluster 4.4. Methodology.

Methodological questions were also focused on, like how theory based synthesis of development interventions can be made (6.09). The use of meta-analyses for developing national evaluation capacity (6.10) and the use of participatory evaluation methods (7.08) was also discussed.

5. **Evaluating institutions, programmes and systems.**

Cluster 5.1. Institutions.

Evaluation of formal institutions (organisations) was discussed and a number of case studies brought up: evaluation of a UK government delivery body (8.13), evaluation of "a deceased management service" in the UK (1.11), evaluations of non-profit organisations (3.14), evaluation of New Zealand schools (4.13). Evaluation as a tool of management was also discussed in general terms (8.11).

Also informal institutions that are vital in evaluation were discussed in this stream, for example ethics in evaluation practices (1.10). How evaluation guidelines, principles,

standards and codes of ethics shape the connections between evaluation and society was also debated (5.12).

Involvement of communities in evaluation designs and empowerment by evaluation was also discussed in this stream (6.14, 5.14).

Cluster 5.2. Programmes.

Evaluation of programme was discussed from several different angles (4.12). One theme was how large-scale public initiatives and their sustainability should be evaluated in different cultures (8.12, 7.11). And how can quality be maintained in evaluations of complex programmes (6.13)? Methodologies for evaluation of unanticipated consequences of programme action was also dealt with (3.14). A number of case studies were also brought forth; China's research and development programme, systems for evaluation of a Canadian military family service programme, systems for evaluation of social protection programmes in Slovenia (2.10), evaluation of investment (7.12), a local children's fund programme (8.13). The EU structural funds were discussed also in this stream (2.11).

Cluster 5.3. Systems.

Furthermore, systems for evaluation of university-based research (3.13) and evaluation of government driven research assessment (5.13) was debated. Evaluation of new policy instruments was another theme on the agenda (4.14). In this context the issues how stakeholder interests can be satisfied and when they should be taken into account were also dealt with (7.14).

6. **Evaluation futures**.

Cluster 6.1. Evaluation and its contexts.

In this stream it was discussed how evaluation designs should be adapted to their contexts. One theme was systems for evaluation in changing political, national and individual contexts (1.12). Also in this stream evaluation in complex governance arenas was discussed and that theme was related to the question of how the choice of evaluation design ought to be effected by cultural experiences if evaluation effectiveness is striven for (8.15). A related theme was how scientific and political demands on evaluation designs should be balanced (8.15). Also evaluation for democratic development in different policy sectors was discussed (1.12). Also factors effecting the influence of evaluations in the public sector were dealt with (5.15).

Cluster 6.2. Evaluation design.

Evaluation designs for evaluation of sustainability were discussed (3.15). Also performance monitoring for results based management (7.15) and experimental designs (5.15) were brought up. Also how contractual processes should be dealt with was discussed (7.15).

Cluster 6.3. Theories.

Finally, this stream dealt with contemporary use of theories and methods (6.15), for example program theories (2.12) and theories of different disciplines, like ethnological theories (1.12).

2. Themes of the Swedish Evaluation Society conference.

Attachment 1 consists of the programme of the paper presentations of the Swedish Evaluation Society conference. In the following it will be referred to paper presentations and panel discussions with the codes used in the programme (P is short for paper presentation and pan is short for panel discussion).

<u>Similarities between the conferences of UKES/EES and SVUF.</u>
An analysis of the titles shows that several of the most prominent themes of the UKES/EES conference were discussed also at the Swedish Evaluation Society conference. First, **the importance of evaluation for policies and implementation** (Cluster 1.1 of the UKES/EES conference) was discussed from several different angles at the Swedish Evaluation Society conference. Evaluation was discussed as a political tool, partly as a tool for decision making (P 121), but also as a tool for securing accountability and learning (P 231, P 232). Evaluation of systems for implementation of policies was discussed at several locations (pan 31, pan 42). Also the need for change in evaluation of policies of the Swedish government and in evaluation of the implementation of such policies was touched upon (P 411).

Second, **the role and identity of evaluators of different professions and disciplines** (Cluster 2.1) was discussed in different contexts (pan 12, P 331). How evaluation should be done within different policy sectors was also discussed, for example in health care (P 212), education (P 213) and defence (P 411).

Third, **evaluation of complex programmes**, that is programs that are to be implemented by a multilevel organisation structure, (Cluster 5.2) was also discussed (P 421), and in particular in the context of the EU (Cluster 3.1). In the EU context evaluation of the use of the EU funds (P 322) and evaluation of EU projects in general (P 322) was discussed. Furthermore, evaluation of a Swedish equivalent to the European multilevel setting was discussed at several occasions; the Swedish so-called policy for metropolitan areas in the 21st century. This policy was implemented by some Swedish municipalities in cooperation with the national government. Therefore evaluation of this implementation poses some challenges to evaluators similar to the ones characterising EU projects (P 132, pan 21, P 313).

Fourth, **how evaluation should be adopted to different institutions relevant for developing countries** (Cluster 4.1), like the OECD and voluntary organisations, was also discussed at the Swedish Evaluation Society conference (P 123, pan 11, pan 41).

Fifth, **evaluation of education and research** (part of Cluster 5.2) was a theme also at the Swedish Evaluation Society conference (P 112, P 213).

Sixth, **the design of evaluation systems**, the main theme of the Swedish Evaluation Society conference, was discussed in several different contexts at both conferences. At the UKES/EES conference evaluation systems for the health, social protection, university research and family service policies (Cluster 2.2, Cluster 5.2), for complex settings like the EU (Cluster 3.1), for developing countries (Cluster 4.1) and for changing political, national and individual contexts (Cluster 6.1) were discussed. Also at the Swedish evaluation Society conference was different kinds of evaluation systems discussed, for example the value of evidence-based systems (pan 43) and systems for evaluation of research (P 412). Local and regional evaluation systems were discussed in different contexts (pan 33), for example in the

context of the Swedish policy for metropolitan areas in the 21st century (P 313). Also values inherit in evaluation systems were discussed (pan 22, P 131, P 311). It was also discussed whether evaluation in some contexts has become mere routine rather than a critically examining activity (P 122, pan 13).

Differences between the conferences of UKES/EES and SVUF.
It is worth to notice that **values inherit in evaluation**, unlike at the UKES/EES conference, were relatively frequently discussed at the Swedish Evaluation Society conference. The word "value" was used in the context of values inherit in evaluation in the titles of three of the 22 paper sessions and panel discussions of the Swedish Evaluation Society conference (P 332, P 222, pan 22), but only in the title of one of the more than 100 paper sessions of the UKES/EES conference (Session 2.06, Cluster 3.1).

It should also be observed that some themes of the themes of the UKES/EES conference were not discussed at the Swedish Evaluation Society conference. First, **stakeholder involvement in evaluation** was dealt with in several different contexts at the UKES/EES conference (Cluster 2.3, Cluster 5.3, Cluster 1.1. The search words "interest" and "stake" were used.). Second, **evaluation and the sustainability of programmes** was discussed at several occasions at the UKES/EES conference (Cluster 1.1, Cluster 2.3, Cluster 5.2, Cluster 6.2. The search word "sustainability" was used.), but not at the Swedish Evaluation Society conference.

3. Themes of the American Evaluation Association conference.

Generally the scope of the American Evaluation Association conference - the consequences of evaluation - can be said to have been broader than the one of the Swedish Evaluation Society conference - the purposes of evaluation systems. On its webpage the American Evaluation Association provides examples of main themes falling within the scope of their latest conference (http://www.eval.org/eval2006/aea06.call.htm 20070328):

> "Of the possible consequences of evaluation, which should we be seeking (under what circumstances) – and why?
> - What evidence exists about the *actual* consequences of evaluation?
> - What are the unintended consequences of evaluation – for example, might certain kinds of programs and policies be put at greater risk because they are more likely to be evaluated?
> - What are consequences of past evaluations on future evaluation practice?"

First, the programme of the AEA conference was searched for the keywords of the themes discussed both at the conference of the Swedish Evaluation Society and at the UKES/EES joint conference:

1. The importance of evaluation for policies and implementation.
The programme of the AEA conference was searched for the following key words:
-- policy; found 17 times in relevant contexts, for example policy evaluation, policy change, inform policymakers, meet the needs of policymakers.
-- policies; not found.
-- implementation; not found.

2. The role and identity of evaluators of different professions and disciplines.
The programme of the AEA conference was searched for the following key words:
-- role; found two times in relevant contexts, for example professional roles and competences for evaluators.
-- identity/identities; not found.
-- profession; found six times in relevant contexts, for example professional ethics for evaluators, professional roles and competences for evaluators.
-- discipline; found once in a relevant context; multiple perspectives, disciplines and considerations.

3. Evaluation of complex programmes.
The programme of the AEA conference was searched for the following key words:
-- complex; found four times in relevant contexts; complex programmes, performance management in complex systems, complex projects, complexity theory.
-- multilevel; found three times in relevant contexts, for example multilevel evaluation, multilevel analysis.

4. Evaluation for institutions relevant for developing countries.
The programme of the AEA conference was searched for the following key words:
-- development; found five times in relevant contexts; evaluation of culturally-specific economic self-sufficiency programmes, rural development, evaluation capacity building in developing countries, poverty programmes in developing countries, international development programmes.

5. Evaluation of education and research.
The programme of the AEA conference was searched for the following key words:
-- education; found more than 35 times in contexts dealing with different forms of education.
-- research; found 19 times in the context of evaluation of research and development, that is papers dealing with evaluation as research has been excluded.

6. The design of evaluation systems.
The programme of the AEA conference was searched for the following key words:
-- system; found 17 times in relevant contexts; performance indicator systems, systems to track and evaluate research programs, how systems-based approaches can effect the consequences of evaluation, consequences of systems thinking in evaluation, designing and implementing a multi-year cross-site evaluation system, systems evaluation and evaluation systems, applications of systems thinking to evaluation.

Second, the programme of the AEA conference was searched for the keyword "value", because the theme values inherit in evaluation seems to have been more frequently discussed at the conference of the Swedish Evaluation Society than at the UKES/EES joint conference. The following four themes were found: value-driven evaluation, values and evaluations in international contexts, values-engaged evaluation, indigenus values and perspectives in evaluation.

Third, the programme of the AEA conference was searched for the keywords of the themes discussed at the UKES/EES joint conference but not at the conference of the Swedish Evaluation Society:

1. Stakeholder involvement in evaluation.
The programme of the AEA conference was searched for the following key words:
-- stake; found 10 times in relevant contexts; multiple stakeholder requirements, stakeholder relationships, establishing trust and building relationships with clients and other stakeholders, stakeholders and evaluation of schools and systems, measuring outcomes for stakeholders, promoting collaboration among diverse stakeholders, evaluator-stakeholder interaction, capacity building of stakeholder evaluation teams, stakeholder perspective in collaborative, participatory and empowerment evaluation, engaging stakeholders in interpreting their data.
-- interest; not found.

2. Evaluation and the sustainability of programs.
The programme of the AEA conference was searched for the following key words:
-- sustain; found three times in relevant contexts; consequences of providing evaluation data for sustainability purposes, from risk management to sustainability and survivability, conducting and using evaluation about program sustainability.

Section 5: Concluding discussion.

1. The objectives and the purpose of the study.

This report has had the following three objectives:

1. To provide an overview of Swedish courses and programmes, which deal with evaluation, and to make a brief comparison with the situations in other European countries.

2. To elaborate on differences between countries concerning evaluation research and education by discussion at a symposium in Sweden in June 2006 and at the European Evaluation Society conference in London in October 2006.

3. To provide an overview of frequently discussed themes in the contemporary Swedish and European academic debates on evaluation.

It was stated that these three objectives have a common purpose; to investigate which predominant conceptions of evaluation there are in the contemporary Swedish academic world. In order to fulfil the objectives three case studies have been conducted. They focus on conceptions of evaluation in one sphere each of the academic world; education (objective 1 and 2), research (objective 2) and debate (objective 3). The results of these case studies are presented and discussed in the following.

1.1. Evaluation as education.

In the following the results of Case study 1 will be presented and discussed, and an international outlook will be provided.

35 evaluation courses and programmes at the following 11 universities and institutes of higher education have been presented above:

- Ersta Sköndal University College
- Göteborg University
- Halmstad University
- Linkoping University
- Lund University
- Mälardalen University
- Stockholm University
- Umeå University
- Uppsala University
- Växjö University
- Örebro University

1.1.1. Evaluation courses and programmes of different disciplines.

In the following the evaluation courses and programmes of each discipline will be summarised.

1.1.1.1. Evaluation courses and programmes in pedagogik.

10 evaluation courses and programmes in pedagogik were found at the following 5 universities and institutes of higher education:

- Göteborg University
- Linkoping University
- Lund University
- Mälardalen University

- Uppsala University

3 of these courses/programmes give 7.5 course credits each. The remaining 7 give 15 course credits each.

1.1.1.2. Evaluation courses and programmes in social work, work science and sociology.
14 evaluation courses and programmes in social work, work science and sociology were found at the following 9 universities and institutes of higher education:
- Ersta Sköndal University College
- Halmstad University
- Linkoping University
- Lund University
- Stockholm University
- Umeå University
- Uppsala University
- Växjö University
- Örebro University

7 of these courses/programmes give 7.5 course credits each, 6 give 15 course credits each and 1 gives 30 course credits.

1.1.1.3. Evaluation courses and programmes in psychology.
4 relevant evaluation courses and programmes in psychology were found at the following 3 universities and institutes of higher education:
- Mälardalen University
- Stockholm University
- Uppsala University

2 of these courses give 7.5 course credits each, 1 gives 15 course credits and 1 gives 30 course credits.

1.1.1.4. Evaluation courses and programmes in political science and peace and development research.
7 evaluation courses and programmes in political science and peace and development research were found at the following 4 universities and institutes of higher education:
- Göteborg University
- Halmstad University
- Lund University
- Uppsala University

2 of these courses/programmes give 7.5 course credits each, 4 give 15 course credits each and 1 gives 22,5 course credits.

1.1.2. Evaluation theories in education.
5 of the 10 evaluation courses and programmes in pedagogik deal with evaluation theories. However, in none of these cases is "evaluation theory" defined. In one case the formulation "theories for evaluation" is used, which may indicate that prescriptive models for evaluation are referred to.

4 of the 14 evaluation courses and programmes in social work, work science and sociology deal with evaluation theories. However, in most cases it remains unclear what is meant by evaluation theories.

4 of the 7 evaluation courses and programmes in political science and peace and development research deal with evaluation theories. In most cases "evaluation theory" seems to refer to theories *about* evaluation, that is the theory is that deal with the contexts and consequences of evaluation (*meta* evaluation theories).

1.1.3. Evaluation on the foundation and advanced levels.

Table 1 below shows that the number of evaluation courses and programmes on the foundation level is greater than on the advanced level. All the studied disciplines follow this trend except psychology. No conclusions can be drawn from the fact that the discipline of psychology doesn't follow the trend, as the number of reported evaluation courses and programmes in that discipline is very low.

Table 1: Evaluation courses and programmes in different disciplines and on different levels.

	Pedagogik	Social work etc	Psychology	Political science etc	Total
Foundation courses	6	8	2	5	21
Advanced courses	4	6	2	2	14

Source: Table produced by the author.

1.1.4. Properties of evaluation courses and programmes in different disciplines.

Table 2 below shows the number of times that properties of the five ideal types of evaluation courses (formed in the first section of this report) have been observed among the evaluation courses and programmes presented above. It is important to note that several of the courses and programmes under study have properties of more than one of the five ideal types of evaluation courses and programmes.

The ideal type that scored the highest is the one that deals with evaluation within one discipline or field. That ideal type also scored the highest as well in the category pedagogik as also in the category social work and related disciplines. Again the number of scorers in the categoric psychology was too low for any conclusions to be drawn. The ideal type that scored the highest in the category political science is the one that deals with meta evaluation. It should also be noted that the ideal types 3, 4 and 5 together gained 75 percent of the scorers in the category political science, while they gained only ca 50 percent of the scorers in the two categories pedagogik and social work and related disciplines respectively. This indicates that political science may be the one of the studied disciplines that deals the most with evaluation topics that have scopes which reach beyond discipline divides.

Table 2: Properties of evaluation courses and programmes in different disciplines and in total.

		Pedagogik	Social work etc	Psychology	Political science etc	Total
	1	0	5	1	0	6
Character of	2	7	6	1	4	18
the courses/	3	4	3	2	4	13

programmes	4	1	3	0	2	6
	5	3	3	1	6	13

Explanations: 1 = evaluation of a certain object, 2 = evaluation within one discipline or field, 3 = multidisciplinary evaluation, 4 = transdisciplinary evaluation, 5 = meta evaluation.
Source: Table produced by the author.

1.1.5. Definitions of evaluation.

It should be noted that in none of the studied course/programme descriptions or syllabi is evaluation defined.

Verner Denvall, Per-Åke Karlsson and Shari Granlöf have conducted an investigation of educations in evaluation for social service workers. They too state that the term "evaluation" is seldom defined or explained:

> "Begreppet "utvärdering" präglas av oklarhet. Förutom att detta varit en försvårande omständighet vid kartläggningen, påverkar det givetvis den teoretiska och metodologiska inriktningen för utbildningsaktiviteterna. Vanligt är att betoningen ligger på att ge kunskaper, inklusive modeller och metoder, för att utvärdera och bedriva kvalitetsarbete som är kopplat till det egna arbetet och den egna arbetsplatsen."[3]

1.1.6. Required readings of evaluation courses and programmes.

The evaluation courses and programmes discussed above cover 194 different required readings, of which 19 are used in more than one of the courses or programmes. Above the 19 required readings, which are used in more than one of the courses or programmes, have been listed in connection to the disciplines and courses in which each reading is used. A list of all the 194 required readings can be found in Attachment 1.1.

The book that is used as required reading in the greatest number of courses is *Utvärdering i politik och förvaltning* (1998) by Evert Vedung; it is used in 8 different courses. *Utvärdering – mer än metod. Tankar och synsätt i utvärderingsforskning: en översikt* (1999) by Ove Karlsson Vestman is used as required reading in 5 different courses. *Från sanningssökande till styrmedel. Moderna utvärderingar i offentlig sektor* (2002) by B. Rombach and K. Sahlin-Andersson is used as required reading in 4 different courses. A great number of books and articles are used as required readings in 2 or 3 evaluation courses/programmes.

In our report *Evaluation as Education, Research and Debates* the required and voluntary readings are presented in connection to the presentation of each course. The required readings are also listed for each investigated discipline (with closely related areas) separately. It can be reported from that study that the evaluation courses and programmes in pedagogik cover 67 different required readings, of which 7 are used in more than one of the courses or programmes. The courses and programmes in social work, work science and sociology cover 91 different required readings, of which 9 are used in more than one of the courses or programmes. The courses and programmes in political science and peace and development research cover 43 different required readings, of which 2 are used in more than one of the courses or programmes. No corresponding investigation was made concerning required readings in psychology, as the number of found required readings was too low.[4]

It should be stressed that not all the required readings deal with evaluation. However, we have chosen to present also these titles, as they give an impression of the contexts in which

---

[3] Denvall - Karlsson - Granlöf 2003, p 8.
[4] Petersson - Karlsson 2007.

evaluation is taught. For example, some of the books and articles that are used as required readings in more than one evaluation course or programme, and which don't focus mainly on evaluation, deal with research methodology, organisational theory, organisational learning, education, continuing training and development of the social services.

In the investigation of educations in evaluation for social service workers by Verner Denvall, Per-Åke Karlsson and Shari Granlöf it is stated that such educations educations often focus on evaluation of organisational change and working life quality:

> "Det är inte självklart att fokus ligger på att ge kunskaper om insatsernas värde för dem som insatserna avser, dvs. klienterna och brukarna. Vanligt är att man fokuserar på "inre förhållanden", som organisationsförändringar och personalförhållanden."[5]

That is, organisational theory/organisational change is emphasised as an important context of evaluation as well by them as also by us.

A clear majority of the required readings are written in Swedish. This confirms findings by Verner Denvall, Per-Åke Karlsson and Shari Granlöf. In that context they stress that the evaluation discourse is an international one.[6] That indeed makes it seem remarkable that a majority of the required readings are written in Swedish.

1.1.7. International outlook.

In the following brief presentations will be made of the situations concerning evaluation courses/programmes offered in some countries, which have Master programmes in evaluation. Thereby they can be said to have come one step further than the Swedish academic world concerning education in evaluation. For further information on university-based education in evaluation in Europe we refer to an enquiry made by Wolfgang Beywl and Katja Harich in 2005. They have identified about 15 working programs. The findings will be published in *Evaluation* (forthcoming).[7]

1.1.7.1. Denmark.

Compared to Sweden evaluation as a tool of governance arrived rather late to Denmark. However, an evaluation society was founded at approximately the same time as in Sweden. At universities evaluation has been taught for more than 20 years as a more or less integrated part of different disciplines, such as pedagogy, economics, public health, development aid and social work, and in September 2006 the University of Southern Denmark moved one step further than any Swedish university, with the founding of a Master Program in Evaluation.[8]

First, the programme resembles the third ideal type of evaluation programmes (multidisciplinary evaluation programmes), as it deals with as well qualitative as also quantitative methods and as it emphasises social science thinking in general. The emphasis on political and organisational systems as social contexts of evaluation makes the programme resemble the fifth ideal type of evaluation programmes - meta evaluation programmes:

> "In addition to evaluation theory and quantitative and qualitative methods, a major emphasis is on understand political and organizational systems as social contexts for evaluation. The relevance of social science thinking in general is also emphasized."[9]

---

[5] Denvall - Karlsson - Granlöf 2003, p 8.
[6] Denvall - Karlsson - Granlöf 2003, p 8.
[7] Beywl 2006a.
[8] Dahler-Larsen 2006, p 1.
[9] Dahler-Larsen 2006, p 2.

The programme is offered by a Department for political science and public management. It is a two year program and it runs part time (50 %). It is designed for "professionals, typically in their mid-career":

> "The program is primarily targeting managers and professionals and to a smaller extent consultants."[10]

One emphasised feature of the programme is that it combines theory and practice. It is

> "university-based and research-based, and it attempts to demonstrate the value of theory and methodological training for the practice of evaluation and also for a reflexive and critical approach to evaluation as a social phenomenon."[11]

In order to reach this goal it focuses on three core areas of evaluation knowledge and skills[12]:

> - evaluation theory
> - methodology
> - knowledge of organizational and political systems

1.1.7.2. France.

In France evaluation as a tool of governance has gained grounds in the last 10 years. In the 1990s there were two driving forces behind the introduction of evaluation:

> (1) compulsory evaluation of socio-economic programmes funded under the EC regional policy, and
>
> (2) compulsory evaluation of State-Region financial plans.

Evaluation was first introduced at regional and national levels, then at intraregional levels and in NGOs. The French Evaluation Society (SFE) was founded in 1999.[13]

Evaluation was first taught in the form of on-the-job training seminars and specific courses integrated in almost all types of academic curricula. Since 1998, a few academic programmes have started to refer to evaluation in their titles. They were devoted to 50 % or more to evaluation and related methods and techniques, were "labour-market oriented, and hosted by various academic disciplines rather than strongly connected to them". At Rennes University the discipline was economics, at Tours University it was sociology, at Amiens University it was political science, and at Marseille University it was management.[14]

The Master Programme in Evaluation at the University of Lyon started in 2004. Also this programme in evaluation aims at combining theory and practice, which is not surprising as it was jointly funded by the Institute of Political Studies and a consultancy specialised in evaluation. Furthermore, it includes a 5 to 6 month internship in a public or private organisation, which has "a strong evaluation capacity". More precisely the programme has two objectives[15]:

> 1. to enable participants to be immediately effective as evaluation professionals in the public or private sector,
> 2. to provide participants with the capacity to understand the politics of evaluation with the help of social science concepts, and especially policy analysis.

It is a 2-year programme and it is open to both students and professionals in their mid-career. The topics dealt with can be described as mainly multidisciplinary, but the focus on evaluation theory in relation to policy analysis makes the programme resemble also the

---

[10] Dahler-Larsen 2006, p 2.
[11] Dahler-Larsen 2006, p 2.
[12] Dahler-Larsen 2006, p 2.
[13] Toulemonde 2006, p 1.
[14] Toulemonde 2006, p 1.
[15] Toulemonde 2006, p 2.

second ideal type of evaluation courses and programmes - courses/programmes dealing with evaluation within one discipline[16]:

- evaluation theory in relation to policy analysis
- evaluation methodology (structuring, gathering data, analysing, formulating value judgements)
- evaluation process (terms of reference, tender, methodological design, quality assurance)
- social science techniques applied to evaluation
- evaluation practices across institutions

1.1.7.3. Switzerland.

In Switzerland evaluation as a tool of governance gained grounds about 20 years ago. The 1999 amendment of the Swiss constitution, which stipulates an obligation to evaluate federal policies, gave the evaluation wave further pace. The Swiss evaluation Society (SEVAL) is the world's biggest national evaluation Society in relation to the country's population and it offers professional development workshops. There is one eight-day course at IDHAP, which is conducted in cooperation with the University of Lausanne. Its focus is mainly on policy evaluation. It is taught in French.[17]

Since 2002 the autonomous university unit for further education at Bern University offers an evaluation training programme. Originally it covered 28 training days, but in 2004 it was upgraded to 40 training days (36 ECTS credits). There is a special track of this education; a 15 ECTS credits basic course for school evaluators. In 2007 the programme will be upgraded with additional 24 ECTS credits.[18]

The programme is focused on quantitative and qualitative methods, and in that it resembles the third ideal type of evaluation programs/courses - multidisciplinary evaluation courses/programs. Also methods for analysis of the social/political/cultural contexts of the evaluated objects are taught, which makes the programme resemble the fourth ideal type of evaluation courses/programmes - transdisciplinary evaluation courses/programmes. Finally, the role of evaluation in politics and society is dealt with. In that the programme resembles the fifth ideal type of evaluation courses/programmes - meta evaluation courses/programmes.[19]

Also this programme emphasises the link between theory and practice, as it is supposed to transfer evaluation theory and methodology to the work of professionals of various fields. Most of these professionals lack a substantial training in empirical methods. As a part of the education the student takes part in an evaluation project of at least 180 hours, often more than 400 hours. Most evaluation projects take place in public arenas, such as education, health and care and social security.[20]

1.2. Evaluation as research.

The original idea for the research project "Evaluation roots in the USA and in Europe – tracing traditions?" was to describe the roots (origins) of evaluation in Europe in a way

---

[16] Toulemonde, Jacques 2006, p 2.
[17] Beywl, Wolfgang 2006b, p 1.
[18] Beywl, Wolfgang 2006b, p 1.
[19] Beywl, Wolfgang 2006b, p 1f.
[20] Beywl, Wolfgang 2006b, p 2.

comparable to Christina Christie's and Marvin Alkin's description of the development of American prescriptive evaluation models. They had described the development of such models in three different respects; methods, values and use. That is, they had described the developments of evaluation methods, of criteria to be used in evaluation and of the use of evaluation separately. The development of American prescriptive evaluation models had been described by the means of a tree model with three branches (methods, values and use), on which the names of theorists, which have come to be associated with new steps in the history of thought on evaluation, had been written. That is, the intention of the research project was to put European names on a tree model comparable to Christie's and Alkin's.

However, during discussions in the Swedish research group on this theme it seemed difficult to come up with individuals, which have advocated certain evaluation models and become associated with them. Furthermore, the discussion on the development of European evaluation did not produce a list of evaluation models. In order to grasp what we associated with the development of evaluation the term evaluation tradition was introduced. An evaluation tradition was referred to as the sum of the evaluation discourse (approximately corresponding to the discussions described by Alkin and Christie), the institutions (organisations) that perform evaluation of public sector services and the legacy of evaluations performed by these institutions. The discussions on the development of evaluation in Sweden tended to focus on the institutions that perform evaluation of public sector services and on the functions of these institutions. It was argued that in Sweden evaluating institutions mainly fulfil the functions of improving implementation of public policy, securing accountability in the chain of public representation and legitimising the measures taken by the state in the eyes of the electorate. In short; we seemed to associate the institutions and their functions, rather than evaluation theorists and their models, with the development of evaluation.

Furthermore, there was agreement within the group that the institutions, which perform evaluation of public sector services, constitute the only part of the evaluation tradition where it seemed likely that vital national or regional differences are to be found. That is, it seemed likely that different kinds of organisations perform evaluation of public sector services in different countries. This theme was further elaborated by Petersson - Karlsson in a paper for the Stockholm symposium. It was argued that the choice of welfare state model - the decision on which services the state is supposed to provide - probably plays a vital role in determining what kind of institutions that are to perform evaluation of public sector services. Such institutions can be public or private, and if public they can be national or regional, and sectorally bounded or not.

At the Stockholm symposium this perspective was contrasted to Alkin - Christie's. If not before, it became clear then that the two conceptions of evaluation were fundamentally different. Of course there are evaluation theorists also in Europe and naturally there are institutions that evaluate public sector services also in the USA, but for some reason the trains of thought concerning the development of evaluation seemed to be different among the American and the Swedish participants respectively.

In short; in Alkin's and Christie's perspective evaluation seems to be associated with prescriptive models and theorists advocating them. By contrast, the Swedish discussion within the project has been founded on a functionalistic conception of evaluation; certain institutions perform evaluation in order to fulfil some functions in relation to the welfare state. Reinhard Stockmann presented a similar, functionalistic conception of evaluation, but

emphasised that evaluation in Germany seems to be undertaken by different kinds of institutions in different policy sectors.

A broad perspective on evaluation as an instrument of welfare allocation can provide an explanation for the contrast between these two conceptions of evaluation.

In the Swedish part of the research group the impression has been brought forth on several occasions and by different researchers that evaluation in Sweden and in the USA aims at fulfilling quite different functions. It has been argued above that evaluation in the Swedish part of the research group was associated with its functions in relation to the welfare state - improving implementation of public policy, securing accountability in the chain of public representation and legitimising the measures taken by the state in the eyes of the electorate. That is, evaluation is to fulfil its main functions in relation to the hierarchical chain of public representation - the organisational system that is supposed to allocate welfare in the Swedish society.

In the United States, unlike in Sweden, the chain of public representation is not the only primary system for allocation of welfare; also markets for welfare serve that purpose. For example, an individual in need of health care in United States to a greater extent than in Sweden has to get an overview of the services offered in the market of health care services. Similarly, when a state is to introduce a programme for reform of the health care system information is needed on reforms made in other states. In Sweden by contrast reforms of the health care system are to a greater extent introduced by laws. (And at the Stockholm symposium Nicoletta Stame argued that this is a European characteristic as opposed to the American tradition of trying out programmes by quasi-experimentalism in a federal system.)

Due to the very nature of markets the functions of evaluation in American welfare distribution are different from its functions in Swedish welfare distribution. It is not to improve the implementation of public policy, to secure accountability in the chain of public representation or to legitimise the measures taken by the state in the eyes of the electorate - all important functions to fulfil in a hierarchical system, but not in a market - as much as providing "product information" in markets for welfare. Evaluation of welfare services can be ordered as well by sellers as also by buyers of such services, and it seems likely that their interests influence their choices of evaluation models and methods. It is important to note that as well individuals as also states can be buyers in such markets. Furthermore, buyers and sellers of welfare services are not the only actors, which are likely to have an interest in evaluation of welfare programmes; there is for example also the federal level, which might want to influence the choices of different states. That is, different actors demand different products.

It is our argument that due to the very nature of markets evaluation models are conceived to a greater extent in the United States than in Sweden as products offered in a market. In the American market oriented welfare distribution system it gets important to present the unique features of the offered evaluation model and present it in an easily graspable way. This may be one explanation to why evaluation models in the United States unlike in Sweden are associated with names of intellectuals - "product names".

To sum up; it was our starting point that evaluation fulfils its main functions in welfare allocation. In Sweden welfare is mainly allocated by the state. That provides an explanation to why the Swedish evaluation researchers of the project associate evaluation with institutions that perform evaluation in order to fulfil certain functions, which are relevant to hierarchical

implementation. In the USA by contrast also markets allocate welfare. That provides an explanation to why the American evaluation researchers of the project associate evaluation with evaluation theorists and "their" prescriptive evaluation models.

By this we do not mean to imply that evaluation in all countries fulfils its main functions in relation to the welfare distribution system. In this context it deserves attention that yet another conception of evaluation surfaced at the Stockholm symposium - advocated by Frans Leeuw. It focuses on the contemporary process of isomorphism, that is on the international process of copycatting, in which the participating organisations, for example states, don't aim at increasing their functionality, but at improving their reputations. Evaluation is in this view conceived as one of the reputation improving features of states and other organisations. That is, organisations evaluate in order to follow an international trend and thereby seem competent.

Evaluation has also been discussed as a "field" of debates, which cannot be grasped without knowledge of evaluations that have been undertaken, gained respect and set standards. This conception of evaluation is the foundation of the third case study.

Within the project a number of papers have been produced on these and closely related themes. They are all attached to this report.

1.3. Evaluation as debate.

In the following the result of Case study 3 will be presented, scrutinised and related to earlier research.

The analysis of the programmes of the latest conferences of the Swedish Evaluation Society, the European Evaluation Society and the American Evaluation Association supports the hypothesis that the debate on evaluation is mainly an international one. First, several of the most prominent themes of the UKES/EES conference were discussed also at the Swedish Evaluation Society conference. The observed themes discussed at both conferences were:
1. The importance of evaluation for policies and implementation.
2. The role and identity of evaluators of different professions and disciplines.
3. Evaluation of complex programmes (especially in the context of the EU).
4. Evaluation for institutions relevant for developing countries.
5. Evaluation of education and research.
6. The design of evaluation systems.

However, the two conferences didn't seem to all out deal with the same themes. Which the values inherit in evaluation are was discussed at several occasions at the conference of the Swedish Evaluation Society, but not to the same extent at the UKES/EES conference. Furthermore, stakeholder involvement in evaluation and evaluation and the sustainability of programmes was discussed at several occasions at the UKES/EES conference, but not at the Swedish Evaluation Society conference.

Finally, the extensive programme of the American Evaluation Association conference was searched for the keywords of the themes that were discussed at the Swedish Evaluation Society conference and/or at the UKES/EES joint conference. Relevant keywords of each of the observed themes discussed at the Swedish Evaluation Society conference were found at least three times in the programme of the American Evaluation Association conference. Of the themes discussed at the Swedish Evaluation Society conference two seem to have been

discussed more at the American Evaluation Association conference than the others; the importance of evaluation for policies and implementation (keywords found 17 times in relevant contexts) and evaluation of education and research (keywords found approximately 55 times in relevant contexts). Also stakeholder involvement in evaluation and evaluation and the sustainability of programs, that is the two themes discussed at the UKES/EES joint conference but not at the Swedish Evaluation Society conference, were discussed at the American Evaluation Association conference.

The general conclusion that can be drawn from this study is that the debate on evaluation seems to be internationally rather homogeneous. The three conferences, located in different parts of the western world, more or less covered the same topics, although dealt with from slightly different angles. It should be stressed that the conference of the Swedish Evaluation Society covered very much fewer sessions than the two other studied conferences. Therefore, the absence at the Swedish conference of discussion on stakeholder involvement in evaluation and evaluation and the sustainability of programs should not be interpreted as a total absence of these themes in the Swedish debate on evaluation.

In order to scrutinise our results we addressed two questions to the foreign researchers of the roots group:

1. First, we presented the themes that we found to have been the most frequently discussed ones at the three analysed conferences (without distinction between the conferences). To repeat, those themes were:
1. The importance of evaluation for policies and implementation.
2. The role and identity of evaluators of different professions and disciplines.
3. Evaluation of complex programmes (especially in the context of the EU).
4. Evaluation for institutions relevant for developing countries.
5. Evaluation of education and research.
6. The design of evaluation systems.
7. Values inherit in evaluation.
8. Stakeholder involvement in evaluation.
9. Evaluation and the sustainability of programs.

Our first question was whether they agreed that these topics seem to be among the most frequently discussed ones at international conferences on evaluation.

2. Our second question was whether they would like to make addings to the list of frequently discussed themes at international conferences on evaluation.

Marvin Alkin (University of California, Los Angeles, USA) replied to our request. He argued that the first five of the discussed topics seem to be of the greatest importance today and that two of those themes can be said to be more important than the other ones:
> "The two that strike the greatest chord are "stakeholder involvement in evaluation" and "evaluation of the sustainability of programs"."

He advocated further discussion in the roots group, focused on how these topics are dealt with in different countries.

An analysis by Ove Karlsson Vestman and Inger M. Andersson of the contents of Swedish doctoral theses on evaluation – *Pedagogisk Utvärdering som Styrning – En historia från präster till PISA* – can provide a background to some of the themes, which seem to be

frequently discussed at conferences. First, their study makes it seem likely that the theme "values inherit in performing evaluations" has become more frequently elaborated over the last decades:

> "The question of the relativity and subjectivity of evaluation is ever more focused upon. Today it is usual that evaluation researchers dissociate from the traditional view on evaluation as objective and rational and argue for a hermeneutic, interpreting approach in conducting evaluations."[21]

Second, the study makes it seem likely that also the theme "evaluation for securing accountability and contributing to empowerment" has become more frequently elaborated:

> "Evaluation is further described as a not all neutral tool for governance."[22]

Third, the study indicates that the themes "how evaluation of complex multilevel programmes should be done (especially in the context of the EU)" and "evaluation of the sustainability of programmes" have become more frequently dealt with in the last years. Furthermore, it also indicates that quantifiable data is increasingly used to tackle the complexity of such contexts:

> "At the same time the trend to strive for the measurable and the quantifiable has been revitalised; to measure and compare are explicit political demands in today's society. Observed environmental problems, which call for immediate measures, highlight the need for new evaluation models to be used in evaluations of complex interventions for sustainable development. The European Union, which assumes increasing mobility and competition on markets, requires reliable and effective tools for comparative studies."[23]

---

[21] Karlsson Vestman - Andersson 2007, p 106.
[22] Karlsson Vestman - Andersson 2007, p 106.
[23] Karlsson Vestman - Andersson 2007, p 106.

1.4. Swedish academic conceptions of evaluation.

The conceptions of evaluation that have been mapped in or have been fundaments of the three case studies will in the following be summarised and compared by the means of a distinction between internal and external conceptions of evaluation.

An internal conception of evaluation focuses on what evaluation is in itself. By contrast, an external conception of evaluation focuses on evaluation as a component in a system.

The discussion above on the required readings of the evaluation courses and programmes under study highlights an external conception of evaluation; the required readings, which don't deal with evaluation, give evaluation an educational context - that is, relate it to other areas of knowledge, in which it fulfils a function. It has been stated that some of the required readings, which don't focus mainly on evaluation, deal with research methodology, organisational theory, organisational learning, education, continuing training and development of the social services. In a broader context also the disciplines in which evaluation is taught indicate in which areas of knowledge it is considered to fulfil a vital function. As stated above, it is our impression that these disciplines are mainly pedagogik, social work, work science, sociology, psychology, political science (public administration) and peace and development research.

In Alkin's and Christie's perspective evaluation seems to be implicitly defined by its use, methods and values. Such a conception of evaluation can be labelled "internal" as it focuses on components of evaluation. The functionalistic conception of evaluation mirrored in the Swedish and parts of the European discussion can by contrast be labelled "external" as it focuses primarily on the context where evaluation is performed. The conception of evaluation as one of the reputation improving features of states and other organisations - that is, the conception of evaluation as a part of the process of isomorphism - can also be labelled external, as it focuses on evaluation as a part of a broader context.

Evaluation has also been described as a field of debate, defined by the topics debated. It has been argued that the following eight topics seem to be the most frequently debated ones in the field of evaluation:
1. The importance of evaluation for policies and implementation.
2. The role and identity of evaluators of different professions and disciplines.
3. Evaluation of complex programmes (especially in the context of the EU).
4. Evaluation for institutions relevant for developing countries.
5. Evaluation of education and research.
6. The design of evaluation systems.
7. Values inherit in evaluation.
8. Stakeholder involvement in evaluation.
9. Evaluation and the sustainability of programs.
This conception of evaluation can be labelled internal, as it doesn't place evaluation in a context, but defines vital components of it.

2. References.

Beywl, Wolfgang 2006a, "University Study Programs in Evaluation - in Search for a European Connection".

Beywl, Wolfgang 2006b, "University study programs in evaluation - in search for a European connection. Contribution from Switzerland to the interactive workshop at the EES/UKES joint conference in London, October 4-6, 2006."

Dahler-Larsen, Peter 2006, "Contribution from Denmark on "University Study Programs in Evaluation – in Search for a European Connection"." Interactive workshop at the EES/UKES joint conference in London, October 4-6, 2006-06-17.

Denvall, Verner, Per-Åke Karlsson and Shari Granlöf 2003, *Att utbilda för utvärdering - En nationell kartläggning*. Stockholm; Socialstyrelsen.

Karlsson Vestman, Ove and Inger M. Andersson 2007, *Pedagogisk Utvärdering som Styrning – En historia från präster till PISA*. Forskning i fokus, nr 35. Stockholm; Myndigheten för Skolutveckling.

Petersson, Gustav Jakob and Ove Karlsson 2007, *Evaluation as Education, Research and Debates*. Unpublished manuscript.

Toulemonde, Jacques 2006, "Contribution from France on "University Study Programs in Evaluation – in Search for a European Connection". Interactive workshop at the EES/UKES joint conference in London, October 4-6, 2006-06-17.

Section 6: Attachments.

1. Attachments of case study 1.

Attachment 1.1: Required readings of evaluation courses and programmes under study.

Ahlstrand, Roland (2001), *Norrköpingsmodellen. Ett projekt för ny sysselsättning åt personalen vid Ericsson Telecom AB i Norrköping.* Sociologiska institutionen, Lund.
> Social work, work science and sociology:
> Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Albaek, E. (1998), Knowledge, interests and the many meanings of evaluation: a developmental perspective. Scandinavian Journal of Welfare 1998:7, 94-98.
> Social work, work science and sociology:
> Stockholm University, Utvärdering och utveckling i socialt arbete.

Alvesson, Mats & Sköldberg Kaj (2004), *Tolkning och reflektion: vetenskapsfilosofi och kvalitativ metod.* Lund: Studentlitteratur
> Social work, work science and sociology:
> Örebro University, Dokumentation och utvärdering i socialt arbete.

Andersson, Inger & Karlsson, Ove (2003) "Temaintroduktion - Utvärdering, forskning och politik", Studies in Educational Policy and Educational Philosophy, vol. 1.
> Political science and peace and development research:
> Lund University, Utvärdering av offentlig verksamhet.

Anttila, S., Hagberg, L. & Mattsson, T. (2002) *Samordning av informationsförsörjning i en kunskapsbaserad socialtjänst* Socialstyrelsen.
> Social work, work science and sociology:
> Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.

Armelius, Bengt-Åke, Armelius, Kerstin (1994), *Kausalfrågor och effektfrågor vid utvärdering av behandling.* Stockholm: Statens institutionsstyrelse (SIS), 1994:4 pdf-fil [Rapport]
> Social work, work science and sociology:
> Örebro University, Dokumentation och utvärdering i socialt arbete.

Babor, Thomas (2004). *Alkohol: ingen vanlig handelsvara.* Stockholm: Statens folkhälsoinstitut.
> Social work, work science and sociology:
> Lund University, Socialt arbete med alkohol- och drogproblem.

Bee, Frances & Bee. Roland, 2003, Learning Needs Analysis and Evaluation. 2nd Edition. Chartered Institute of Personnel and Development, London.
> Pedagogik:
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

Bejerot, Eva & Hasselbladh, Hans (red.) (2002). *Kvalitet utan gränser. En kritisk belysning av kvalitetsstyrningen.* Lund: Academia Adacta.
>Pedagogik:
>Uppsala University, Utvärdering och kvalitetsutveckling.

Berglund, M., m. fl. (1996), *Dokumentation inom missbrukarvården.* Centrum för utvärdering av socialt arbete. Stockholm: Liber.
>Social work, work science and sociology:
>Stockholm University, Utvärdering av socialt arbete (Socialt arbete II) (valda delar).

Bergmark, A. & Lundström, T. (2006) "Mot en evidensbaserad praktik? – Om färdriktningen i socialt arbete" i *Socialvetenskaplig tidskrift* (under utgivning), 12 s.
>Social work, work science and sociology:
>Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Bernler, Gunnar & Johnsson, Lisbeth (1993), *Den sociala journalen. En studie i akter, deras struktur och funktion.* Institutionen för socialt arbete, Göteborgs universitet, 1993:7 [Rapport]
>Social work, work science and sociology:
>Örebro University, Dokumentation och utvärdering i socialt arbete.

Börjeson, B (2001) *Utvärdering av FOU. En studie av FOU-enheter inriktade på individ- och familjeomsorg.* Socialstyrelsen.
>Social work, work science and sociology:
>Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.

Carlström, Inge & Hagman, Lena-Pia (1995). *Metodik för utvecklingsarbete och utvärdering.* Akademiförlaget.
>Pedagogik:
>Göteborg University, Kvalitet och utvärdering inom utbildning.

Carlström, Inge (2007). *Metodik för utvecklingsarbete och utvärdering.* 5:e upp. Lund: Studentlitteratur.
>Pedagogik:
>Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Cohen, Louis, Manion, Lawrence, & Morrison, Keith (2000). *Research Methods in Education* (5th ed.).
>Pedagogik:
>Lund University, Utrednings- och utvärderingsmetod och verksamhetsanknutet projektarbete.

Cook, Thomas D. (1997) 'Lessons learned in Evaluation Over the Past 25 Years' I Chelimsky, Eleanor and Shadish, William R. (eds.) Evaluation for the 21st Century. A Handbook. London: Sage.
>Social work, work science and sociology:
>Umeå University, Utvärdering (sid 30-52).

Christianson, S. Å. (Red.). (1996). *Rättspsykologi: Den forensiska psykologin i Sverige - en kunskapsöversikt.* Stockholm: Natur och Kultur.
>> Psychology:
>> Stockholm University, Rättspsykologi I.

Christianson, S. Å., Engelberg, E., & Holmberg, U. (1998). *Avancerad förhörs- och intervjumetodik.* Stockholm: Natur och Kultur.
>> Psychology:
>> Stockholm University, Rättspsykologi I.

Christianson, S. Å., & Wentz, G. (2002). *Brott och minne.* Stockholm: Natur och Kultur.
>> Psychology:
>> Stockholm University, Rättspsykologi I.

Dahlberg, M., Vedung, E. (2001), *Demokrati och brukarutvärdering.* Lund: Studentlitteratur.
>> Social work, work science and sociology:
>> Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).

Dahler-Larsen, Peter (1999). *Den Rituelle Reflektion - om evaluering i organisationer.* Odense: Odense Universitetsforlag.
>> Pedagogik:
>> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Dahler-Larsen, Peter, 2003. Det politiske i evaluering. ur Studies in Educational Policy and Educational Philosophy, vol. 1.
>> Political science and peace and development research:
>> Lund University, Utvärdering av offentlig verksamhet.

Denscombe, Martyn (2000), *Forskningshandboken – för småskaliga forskningsprojekt inom samhällsvetenskapen.* Studentlitteratur, Lund.
>> Pedagogik:
>> Uppsala University, Utvärdering och kvalitetsutveckling.
>> Social work, work science and sociology:
>> Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Denvall, Verner & Vinnerljung, Bo (red) (2006). *Nytta eller fördärv?* Stockholm: Natur & Kultur. Utkommer tidig höst 2006.
>> Social work, work science and sociology:
>> Lund University, Socialt arbete med alkohol- och drogproblem (15 sidor läses).

Dimitrakopoulos, D & Richardson, J: "Implementing EU public policy", i Richardsson, Jeremy, red (2001): *European Union. Power and policy-making,* 2nd ed. London: Routledge.
>> Political science and peace and development research:
>> Göteborg University, Implementering och utvärdering.

Docherty, Peter (1996) *Läroriket: vägar och vägval i en lärande organisation.* Stockholm: Arbetslivsinstitutet.
>> Pedagogik:

Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I
(Internet-based version).

Donaldson, SI & Lipsey, MW (2006): "Roles for theory in contemporary evaluation practice",
I Shaw, IF m fl, red: *The Sage Handbook of Evaluation*. London. Sage.
<u>Political science and peace and development research:</u>
Göteborg University, Implementering och utvärdering.

Du Rietz, Lars, Lundgren, Ulf P., Wennås, Olof, 1987, Ansvarsfördelning och styrning på
skolområdet. DsU 1987:1. Allmänna förlaget, Stockholm.
<u>Pedagogik:</u>
Uppsala University, Pedagogik med inriktning mot personalutveckling,
fortsättningskurs B/ Pedagogik, fortsättningskurs B, Styrning och utvärdering.

Ebaugh, Helen Rose Fuchs (1988). *Becoming an ex: the process of role exit*. Chicago: Univ.
of Chicago Press.
<u>Social work, work science and sociology:</u>
Lund University, Socialt arbete med alkohol- och drogproblem.

Edvardsson, Bo, Andersson, Tobias, Sandén, Mattias och Waller, Björn (1998), *Mätning av
tjänstekvalitet i praktiken.* Studentlitteratur, Lund.
<u>Social work, work science and sociology:</u>
Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i
arbetslivet.

Edvardsson, Bo (2003) / 2., omarb. och utvidgade uppl. *Kritisk utredningsmetodik : begrepp,
principer och felkällor.* Stockholm: Liber, ISBN/ISSN: 91-47-05171-X.
<u>Social work, work science and sociology:</u>
Örebro University, Dokumentation och utvärdering i socialt arbete.

Egidius, Henry, 2003, Pedagogik för 2000-talet. (Fjärde utgåvan). Natur och Kultur,
Stockholm.
<u>Pedagogik:</u>
Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs
A (Kap 9-12).

Ejlertsson, Göran, 2005, Enkäten i praktiken. En handbok i enkätmetodik. (2:a upplagan)
Studentlitteratur, Lund.
<u>Pedagogik:</u>
Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs
A.

Ekberg, K., Eklund, J., Ellström, P-E. & Johansson, S. (red.).(2006). *Tid för utveckling?*
Lund: Studentlitteratur.
<u>Pedagogik:</u>
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Ekholm, M. & Lander, R. (1993). *Utvärderingspraktikan.* Stockholm: Liber.
<u>Pedagogik:</u>

Göteborg University, Kvalitet och utvärdering inom utbildning.

Ekman, Annalill. (2004). *Lärande organisationer i teori och praktik: Apoteket lär*. Uppsala: Acta Universitatis Upsaliensis.
>
> Pedagogik:
>
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).

Ekonomistyrningsverket (2003): *Resultatanalys för nybörjare.*
>
> Political science and peace and development research:
>
> Göteborg University, Utvärderingsmetoder.

Ekonomistyrningsverket (2006): *Effektutvärdering – att välja upplägg.*
>
> Political science and peace and development research:
>
> Göteborg University, Utvärderingsmetoder.

Eliasson, R.M. 1995, Forskningsetik och perspektivval. Lund: Studentlitteratur.
>
> Social work, work science and sociology:
>
> Växjö University, Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik.

Ellström, P-E. (1992). *Kompetens, utbildning och lärande i arbetslivet.* Stockholm: Publica.
>
> Pedagogik:
>
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Ellström, P-E., Gustavsson, M. & Svedin P-O. (1996*). Lärande i temporär organisation. En studie för ett företagsinternt utvecklingsprogram för processoperatörer*. Linköpings universitet: Institutionen för pedagogik och psykologi, Centrum för studier av människa, teknik och organisation.
>
> Pedagogik:
>
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Ellström, P-E. & Kock, H. (2004). *Nya organisationsformer - nya krav på ledarskap?* Synopsis, 5, 2004.
>
> Pedagogik:
>
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Ellström. Per-Erik. & Hultman, Glenn. (red.). (2004). *Lärande och förändring i organisationer. Om pedagogik i arbetslivet*. Lund: Studentlitteratur.
>
> Pedagogik:
>
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).

Ellström. Per-Erik, Löfberg, Arvid. & Svensson, Lennart. (2005). Pedagogik i arbetslivet. Ett historiskt perspektiv. *Pedagogisk forskning i Sverige.*
>
> Pedagogik:
>
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).

Esaiasson Peter m fl.(2002), *Metodpraktikan. Konsten att studera samhälle, individ och marknad.* Norstedts juridik, Stockholm.
> <u>Political science and peace and development research:</u>
> Halmstad University, Utrednings- och utvärderingsmetodik.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program Evaluation. Alternative Approaches and Practical Guidelines* (3rd ed.). Boston: Pearson.
> <u>Pedagogik:</u>
> Mälardalen University, Utvärdering B.
> <u>Social work, work science and sociology:</u>
> Umeå University, Utvärdering.

Fischer, Frank (1995) Evaluating Public Policy. Chicago: Nelson Hall Publishers.
> <u>Social work, work science and sociology:</u>
> Umeå University, Utvärdering (sid 1-24).

Forss, K & Urwing, M (200X): *Kvalitet i utredningsväsendet. Utveckling och tillämpning av en modell för att beöma kvaliteten på kommitéarbeten.* Regeringskanslitets förvaltningsavdelning, kommitéservice.
> <u>Political science and peace and development research:</u>
> Göteborg University, Implementering och utvärdering.

Franke-Wikberg, Sigbrit (1992), "Utvärderingens mångfald – några ledtrådar för vilsna utbildare". Stockholm: UHÄ. Projektrapport 1992:4.
> <u>Pedagogik:</u>
> Uppsala University, Utvärdering och kvalitetsutveckling.

Furubo, Jan-Eric, Rist, Ray & Sandahl, Rolf (2002) ' A Diffusion Perspective on Global Developments in Evaluation'. In International Atlas of Evaluation. Transaction Publ.
> <u>Social work, work science and sociology:</u>
> Umeå University, Utvärdering (sid 1-27, 115-128).

Gambrill, E. (1999), "Evidence-Based Practice: an Alternative to Authority-Based Practice". *Families in Society: The Journal of Contemporary Human Services.* 1999, 341-350.
> <u>Social work, work science and sociology:</u>
> Stockholm University, Utvärdering och utveckling i socialt arbete.

Giddens, Anthony (1990). *Consequences of modernity.* Cambridge: Polity.
> <u>Pedagogik:</u>
> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Giertz, Birgitta (2000). *Uppfattningar av kvalitet. En genomgång av litteratur om kvalitet och kvalitetsarbete.* Uppsala: Uppsala universitet, Rapport från enheten för utveckling och utvärdering nr 21, mars 2000.
> <u>Pedagogik:</u>
> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Goldstein, Irwin. L. & Ford, J. Kevin. (2002). *Training in Organizations. Needs Assessment, Development and Evaluation.* 4th ed. Belmont, California: Wadsworth.
> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).

Granberg, Otto. (2004). *Lära eller läras: om kompetens och utbildningsplanering i arbetslivet.* Lund: Studentlitteratur.
> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

Grinnell,Richard M. (2001), *Social work reserach and evaluation: quantitative and qualitative approaches.* Peacock Publishers.
> Social work, work science and sociology:
> Örebro University, Dokumentation och utvärdering i socialt arbete (valda delar).

Grönmo, S. 2006, Metoder i samhällsvetenskap. Malmö: Liber.
> Social work, work science and sociology:
> Växjö University, Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik.

Guba, Egon & Lincoln, Yvonna S. (1989) Fourth Generation Evaluation. Sage, Newbury Park.
> Social work, work science and sociology:
> Umeå University, Utvärdering (sid 21-49).
> Political science and peace and development research:
> Göteborg University, Analys och utvärdering av fredsarbete.

Gustavsson, M., Lindenmo, B. & Wigg, U. *Olika synsätt på lärande.* Kompendium.
> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Hall, Patrik & Karl Löfgren (2006), *Politisk styrning i praktiken.* Liber förlag, Malmö.
> Political science and peace and development research:
> Halmstad University, Utrednings- och utvärderingsmetodik (valda delar).

Hammar Chiriac, Eva & Hempel, Anders, 2005, Handbok för grupparbete – att skapa fungerande grupparbeten i undervisning. Studentlitteratur, Lund.
> Pedagogik:
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

Hanberger, A (2001): "What is the Policy Problem? Methodological Challenges in Policy Evaluation", *Evaluation* vol 7/no 1 2001 (sid 45-62).
> Political science and peace and development research:
> Göteborg University, Implementering och utvärdering.

Hanberger, Anders (2001) Policy and Program Evaluation, Civil Society and Democracy. American Journal of Evaluation Vol. 22.
> Social work, work science and sociology:
> Umeå University, Utvärdering (sid 211-228).

Hanberger, Anders (2003) 'Den dolda utvärderingspolitiken'. Studies in Policy and Educational Philosophy E-tidskrift 2003:1.
> Social work, work science and sociology:
> Umeå University, Utvärdering.

Hill, M & Hupe, PL, 2002, *Implementing public policy. Governance in theory and practice.*, SAGE, London.
> Political science and peace and development research:
> Uppsala University, Organisering, implementering och utvärdering i offentlig politik.

Hill, M (2005): *The Public Policy Process*. Harlowe: Pearson Longmans.
> Political science and peace and development research:
> Göteborg University, Implementering och utvärdering.

Holgersson, C. (2003). *Företagsledare och chefer* i SOU 2003:16, Mansdominas i förändring – om ledningsgrupper och styrelser.
> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (s. 105-122).

Holmberg, Robert (2002). *Organiserande och kvalitetsutveckling. Om sönderfall, upprätthållande och begripliggörande i arbetslivet.* Akademisk avhandling. Lunds universitet, Institutionen för psykologi.
> Pedagogik:
> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Holmblad Brunsson, Karin (2005), *Reda ut eller trassla till. Handbok för utredare.* Studentlitteratur, Lund.
> Social work, work science and sociology:
> Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Holme, Idar Magne - Solvang, Bernt Krohn, 1991. Forskningsmetodik. Om kvalitativa och kvantitativa metoder. Lund: Studentlitteratur.
> Political science and peace and development research:
> Lund University, Utvärdering av offentlig verksamhet (enligt lärares anvisningar).

Holmlund, Kerstin (red.) (2004) *Vad har skolan med kvalitet att göra?* Studentlitteratur, Lund.
> Pedagogik:
> Uppsala University, Utvärdering och kvalitetsutveckling.

House, E (1996): "The Problem of Values in Evaluation", *Evaluation Journal of Australiasia*, vol 8/no 1 1996 (sid 3-14).
> Political science and peace and development research:
> Göteborg University, Implementering och utvärdering.

House, Ernest & Howe, Kenneth (1999) Values in evaluation and social research. Thousands Oaks, CA: Sage.
> Social work, work science and sociology:
> Umeå University, Utvärdering (sid 91-103).

Humphreies, B. (2003), "What *else* counts as evidence in evidence-based social work?" *Social Work Education*, vol. 22, no. 1, 2003, 81-91.
> Social work, work science and sociology:
> Stockholm University, Utvärdering och utveckling i socialt arbete.

Hydén, Lars-Christer (1995), *Det sociala misslyckandet som berättelse.* 14 sidor, [Tidskriftsartikel]
> Social work, work science and sociology:
> Örebro University, Dokumentation och utvärdering i socialt arbete.

Hällsten, F. & Tengblad, S. (2006). *Medarbetarskap i praktiken.* Lund: Studentlitteratur.
> Pedagogik:
> Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (valda delar c:a 150 s).

Högskoleverket 2003 Socialt arbete, en nationell genomlysning av ämnet. Högskoleverkets rapportserie 2003:16.
> Social work, work science and sociology:
> Växjö University, Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik.

Jacobsson, Bengt och Sundström, Göran (2001). *Resultat utan lärande?* Score rapportserie 2001:15.
> Pedagogik:
> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Jansson, Tomas & Lennart Ljung (2004), *Projektledningsmetodik.* Studentlitteratur, Lund.
> Political science and peace and development research:
> Halmstad University, Utrednings- och utvärderingsmetodik.

Jegermalm, M. Sundh, K. (2005). *Social mobilisering I frivillig regi.* Sköndalsinstitutet.
> Social work, work science and sociology:
> Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.

Jerkedal, Åke, 2005, Utvärdering - steg för steg. Om projekt- och programbedömning. Tredje upplagan. Norstedts Juridik, Stockholm.
> Pedagogik:
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

Jess, K. (1998) *Kvalitet för brukare. Frivilligt socialt arbete ur ett brukarperspektiv.* Sköndalsinstitutet (59 sid.)
> Social work, work science and sociology:
> Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.

Johansson, S (2004): *Implementering av BBIC-projektet i socialtjänstens organisationer.* Stockholm: Socialstyrelsen.
> Political science and peace and development research:
> Göteborg University, Implementering och utvärdering.

Johnson, Björn (2005). *Metadon på liv och död: en bok om narkomanvård och narkotikapolitik i Sverige.* Lund: Studentlitteratur.
> Social work, work science and sociology:
> Lund University, Socialt arbete med alkohol- och drogproblem.

Johansson, Katarina & Wirbing, Peter (2005). *Riskbruk och missbruk. Alkohol, läkemedel, narkotika, uppmärksamma, motivera och behandla inom primärvård, socialtjänst och psykiatri.* 2a uppl. Stockholm: Natur & Kultur.
> Social work, work science and sociology:
> Lund University, Socialt arbete med alkohol- och drogproblem (Valda delar).

Johansson, S. Bergmark, Å. Lundström, T. (2006), *Tonåringar som inte kan bo hemma.* CKP, Stockholm: Rapport 2006:2, 95 s.
> Social work, work science and sociology:
> Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Jordan, B. 2005, Sex, pengar och makt. Om det kollektiva livet omvandlingar. Göteborg: Daidalos.
> Social work, work science and sociology:
> Växjö University, Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik.

Järvinen, Margaretha & Mik-Meyer, Nanna (2003). *At skabe en klient: institutionelle identiteter i socialt arbejde.* Köpenhamn: Hans Reitzel.
> Social work, work science and sociology:
> Lund University, Socialt arbete med alkohol- och drogproblem.

Karlsen, Gustav E., 2002, Utdanning, styring og marked. Universitetsforlaget, Oslo.
> Pedagogik:
> Uppsala University, Pedagogik med inriktning mot personalutveckling, fortsättningskurs B/ Pedagogik, fortsättningskurs B, Styrning och utvärdering.

Karlsson, Ove (1995) Att utvärdera - mot vad? Om kriterieproblem vid intressentutvärdering. Stockholm, HLS förlag.
> Social work, work science and sociology:
> Umeå University, Utvärdering (sid 69-120).

Karlsson, O. (1999). *Utvärdering – mer än metod. Tankar och synsätt i utvärderingsforsk-ning: en översikt.* Àjour - en serie kunskapsöversikter från Svenska kommunförbundet, nr 3. Stockholm: Svenska kommunförbundet, Kommentus.

> Pedagogik:
> Mälardalen University, Utvärdering B.
> Mälardalen University, Utvärdering ur samhällsvetenskapligt perspektiv D.
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.
> Göteborg University, Kvalitet och utvärdering inom utbildning.
> Political science and peace and development research:
> Lund University, Utvärdering av offentlig verksamhet.

Karlsson, Ove (2003), "Evaluation politics in Europe: trends and tendencies". *Studies in Educational Policy and Educational Philosophy* 2003:1. E-tidskrift.

> Social work, work science and sociology:
> Stockholm University, Utvärdering och utveckling i socialt arbete.
> Political science and peace and development research:
> Lund University, Utvärdering av offentlig verksamhet.

Karlsson, Ove (2006): "Relationen utvärdering och politik", i *Socionomen* n4 4/2006.

> Political science and peace and development research:
> Göteborg University, Utvärderingsmetoder.

Khakee, Abdul (2002) "Assessing institutional capital building in local agenda 21 process in Göteborg", Planning Theory and Practice.Vol. 3, No 1.

> Social work, work science and sociology:
> Umeå University, Utvärdering (sid 53-68).

Khakee, Abdul (2003) "The Emerging Gap between Evaluation Research and Practice" Evaluation. Vol.9(3): 338-350.

> Social work, work science and sociology:
> Umeå University, Utvärdering.

Konsten att skriva och tala. 2005. Lund: Statsvetenskapliga föreningen.

> Political science and peace and development research:
> Lund University, Utvärdering av offentlig verksamhet.

Korp, Helena, 2003, Kunskapsbedömning – hur vad och varför. Forskning i Fokus nr 13. Myndigheten för skolutveckling. www.skolutveckling.se, Stockholm.

> Pedagogik:
> Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

Kärrby, Gunni m.fl. (2003). *Pedagogisk kvalitet i skolan.* Lund: Studentlitteratur.

> Pedagogik:
> Göteborg University, Kvalitet och utvärdering inom utbildning.

Körner, Svante & Wahlgren, Lars, 2002, Praktisk statistik. Studentlitteratur, Lund.

> Pedagogik:

Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A (Kap 1-4 och 6).

Lander, R. (2005). Björkskolans modell för kvalitetsarbete och uppföljning. Arbetsrapport till timplanedelegationen. Institutionen för pedagogik och didaktik, Göteborgs universitet.
Pedagogik:
Göteborg University, Kvalitet och utvärdering inom utbildning.

Lander, R. & Granström, K. Retorik i utvärdering och uppföljning av skolan. *Spov – studier av den pedagogiska väven*, nr 26, 1995, s 40-47.
Pedagogik:
Göteborg University, Kvalitet och utvärdering inom utbildning.

Lander, R. & Granström, K. (2000). Skolinspektionen i England och Sverige. Hjälp till självhjälp eller självstyrning med betslet i munnen? *Pedagogisk forskning i Sverige.* Årg. 5 Nr 3 pp 215-234.
Pedagogik:
Göteborg University, Kvalitet och utvärdering inom utbildning.

Leeuw, Frans, Toulemonde, Jacques & Brouwers, André (1999) 'Evaluation Activities in Europe: A quick Scan of the Market in 1998' in Evaluation Vol 5
Social work, work science and sociology:
Umeå University, Utvärdering (sid 487-496).

Lejdhamre, O (2002): *Den statliga revisionen och de lokala investeringsprogrammen. En studie av hur Riksrevisionsverkets och Riksdagens revisorers utvärderingar kommit till användning*. Uppsala universitet: Statsvetenskapliga institutionen.
Political science and peace and development research:
Göteborg University, Implementering och utvärdering.

Lindensjö, Bo & Lundgren, Ulf P, 2000, Utbildningsreformer och politisk styrning. HLS Förlag, Stockholm.
Pedagogik:
Uppsala University, Pedagogik med inriktning mot personalutveckling, fortsättningskurs B/ Pedagogik, fortsättningskurs B, Styrning och utvärdering.

Lindgren, Lena (2006). *Utvärderingsmonstret: Kvalitets- och resultatmätning i den offentliga sektorn*. Lund: Studentlitteratur.
Pedagogik:
Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Littell, J. Popa, M. Forsythe, B. (2005), Multisystemic Therapy for social emotional and behavioural problems in youth aged 10-17. Nordic Cambell Center: Review 2005:1 [elektronisk resurs] 60 s.
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Lundahl, Christian och Öqvist, Oscar (2002), *Idén om en helhet. Utvärdering på systemteoretisk grund.* Studentlitteratur, Lund.

Social work, work science and sociology:
Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Lundmark, A. (1998). *Utbildning i arbetslivet. Utgångspunkter och principer för planering och genomförande av personalutbildning.* Lund: Studentlitteratur.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (valda delar c:a 150 s).
Lund University, Utrednings- och utvärderingsmetod och verksamhetsanknutet projektarbete.

Lundström, T. (2005) "Risk, individualisering och social barnavård". *Särtrycksserien* – Nr 150, Institutionen för socialt arbete. (Ur *Socionomen* 2005:6, forskningssupplement 18, s. 5-14, 14 s.
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Mager, Robert, 1988, Att formulera undervisningsmål. Studentlitteratur, Lund.
Pedagogik:
Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

McLaughlin, JA & Jordan, GB (1999): "Logic models: a tool for telling your program's performance story", *Evaluation and Program Planning*, no 22 (65-72).
Political science and peace and development research:
Göteborg University, Implementering och utvärdering.

Maher, Lisa (1997). *Sexed work: gender, race and resistance in a Brooklyn drug market.* Oxford: Clarendon.
Social work, work science and sociology:
Lund University, Socialt arbete med alkohol- och drogproblem.

Morén, Stefan & Blom, Björn (2003) Insatser och resultat. Om utvärdering i socialt arbete. (Rapport nr 48) Umeå: Umeå universitet, institutionen för socialt arbete.
Social work, work science and sociology:
Umeå University, Utvärdering (sid 5-116).
Örebro University, Dokumentation och utvärdering i socialt arbete.

Myndigheten för skolutveckling (2003): *Att granska och förbättra kvalitet.* Fritzes (eller www.skolutveckling.se).
Pedagogik:
Göteborg University, Kvalitet och utvärdering inom utbildning.

Möller, K., Gerdner, A., & Oscarsson, A. (1998), *Rällsögården. 40 års verksamhet vid ett behandlingshem för vuxna missbrukare.* SiS följer upp och utvärderar. Statens institutionsstyrelse. SiS-rapport 1/98.
Social work, work science and sociology:
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).

Mörth, U. & Kerstin Sahlin Andersson (red). (2006), *Privatoffentliga partnerskap. Styrning utan hierarkier och tvång?* SNS-förlag, Stockholm.
> Political science and peace and development research:
> Halmstad University, Utrednings- och utvärderingsmetodik.

Newman, T. et.al (2004), *Evidence-based Social Work. A Guide for the Perplexed.* Lyme Regis: Russel House Publishing, 179 s.
> Social work, work science and sociology:
> Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Norell, Margareta & Törnqvist, Claes (1995). *Berättelser om ruset: alkoholens mening för tjugoåringar.* Stockholm; Stehag: B. Östlings bokförlag Symposium.
> Social work, work science and sociology:
> Lund University, Socialt arbete med alkohol- och drogproblem.

Nyström, S., Jess, K., Soydan, H,. *Med arbete som insats. Klienteffekter och samhällsekonomisk lönsamhet i socialt arbete.* CUS. Socialstyrelsen.
> Social work, work science and sociology:
> Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.

Nytell, U. (1990). *Lots – utveckling för näringslivet och för skolan? En delrapport från SLAV-projektet.* Uppsala: Uppsala universitet, Pedagogiska institutionen.
> Pedagogik:
> Mälardalen University, Utvärdering B.

Oscarsson, Lars (2000). Den socialtjänstbaserade missbrukarvården under 1990-talet – förutsättningar, utveckling och behov. I: Szebehely, Marta (red). *Välfärd, vård och omsorg. Antologi från kommittén välfärdsbokslut.* SOU 2000:38.
> Social work, work science and sociology:
> Lund University, Socialt arbete med alkohol- och drogproblem.

O'Toole, LJ (2000): "Research on Policy Implementation: Assessment and Pros-pects", i *Journal of Public Administration Research and Theory* vol 10/2 (263-288).
> Political science and peace and development research:
> Göteborg University, Implementering och utvärdering.

Passer, M. W., & Smith, R. E. (2005/2007). *Psychology: The science of mind and behavior,* (2nd ed.). New York: McGraw-Hill.
> Psychology:
> Stockholm University, Rättspsykologi I.

Patton, Michael Quinn (1997). *Utilization-Focused Evaluation - The New Century Text.* 3rd ed. London: Sage Publications.
> Pedagogik:
> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.

Patton, Michael Quinn (2002) Feminist, Yes, but Is It Evaluation? In Feminist Evaluation. Explorations and Experiences. New Directions for Evaluation Nr 96.
> Social work, work science and sociology:

Umeå University, Utvärdering (sid 97-108).

Patton, M.Q. 2002, *Qualitative research & evaluation methods,* London, SAGE.
    Political science and peace and development research:
    Göteborg University, Analys och utvärdering av fredsarbete.

Patton, M.Q. 1982, *Practical Evaluation*. Thousand Oaks: Sage Publications.
    Political science and peace and development research:
    Göteborg University, Analys och utvärdering av fredsarbete.

Patton, M.Q. 1986, *Creative Evaluation*, 2nd edition. Thousand Oaks: Sage Publications.
    Political science and peace and development research:
    Göteborg University, Analys och utvärdering av fredsarbete.

Perrin, Burt (1998): "Effective Use and Misuse of Performance Measurement",
American Journal of Evaluation, vol 19 no 3 1998 (s 367-379).
    Political science and peace and development research:
    Göteborg University, Utvärderingsmetoder.

Pommer Nilsson, Peter (2003). *En introduktion till pedagogik i arbetslivet. Begrepp, historik, nutid.* Pedagogiska institutionen, Umeå universitet, Nr 73.
    Pedagogik:
    Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.
    Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I
    (Internet-based version).

Power, M, 1997, *The audit society*. (Valda delar), Oxford University Press, Oxford.
    Political science and peace and development research:
    Uppsala University, Organisering, implementering och utvärdering i offentlig
    politik.

Repstad, P. 1993, Närhet och distans: kvalitativa metoder i samhällsvetenskap. Lund: Studentlitteratur.
    Social work, work science and sociology:
    Växjö University, Socialt arbete – aktuell forskning, vetenskapsteori och
    forskningsmetodik.

Rhodes, R A W, 1997, Understanding Governance. Policy Networks, Governance, Reflexivity and Accountability, Open University Press, Buckingham.
    Political science and peace and development research:
    Uppsala University, Organisering, implementering och utvärdering i offentlig
    politik.

Rogers, PJ (2000): "Causal Models in Program Theory Evaluation", *New Directions for Evaluation* no 87/2000 (sid 47-57).
    Political science and peace and development research:
    Göteborg University, Implementering och utvärdering.

Rombach, B., & Sahlin-Andersson, K. (red.) (2002). *Från sanningssökande till styrmedel. Moderna utvärderingar i offentlig sektor* (2:a uppl.). Stockholm: Santérus.

        Pedagogik:
        Mälardalen University, Utvärdering B.
        Social work, work science and sociology:
        Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i
        arbetslivet.
        Umeå University, Utvärdering (sid 9-24, 209-227).
        Örebro University, Dokumentation och utvärdering i socialt arbete.

Rosenqvist, Pia (red) (2004). *Addiction and Life course*. NAD-publikation nr 44. Helsingfors: NAD.
        Social work, work science and sociology:
        Lund University, Socialt arbete med alkohol- och drogproblem (100 sidor läses).

Rossi, Peter, Freeman, Howard & Lipsey, Mark (1999) Evaluation. A Systematic Approach. London: Sage.
        Social work, work science and sociology:
        Umeå University, Utvärdering (sid 37-76,400-403, 421-422).
        Political science and peace and development research:
        Göteborg University, Utvärderingsmetoder.
        Halmstad University, Utrednings- och utvärderingsmetodik.

Rothstein, B, 1996, The social democratic state: the Swedish model and the bureaucratic problem of social reform., Univ. of Pittsburgh Press, Pittsburgh, Pa.
        Political science and peace and development research:
        Uppsala University, Organisering, implementering och utvärdering i offentlig
        politik.

Rothstein, B (1994): *Vad bör staten göra? Om välfärdsstatens moraliska och politiska logik.* Stockholm: SNS Förlag.
        Political science and peace and development research:
        Göteborg University, Implementering och utvärdering.

Rothstein, Bo (red.), 2001. Politik som organisation. Förvaltningspolitikens grundproblem. Stockholm: SNS Förlag. 3:e uppl.
        Political science and peace and development research:
        Lund University, Utvärdering av offentlig verksamhet (efter lärarens
        anvisningar).

Sanne, M (2000), *Skrivråd för statsvetare.* Studentlitteratur, Lund.
        Political science and peace and development research:
        Halmstad University, Utrednings- och utvärderingsmetodik.

Segerholm, Christina (2001) National Evaluations as Governing Instruments: How Do They Govern? I Evaluation Vol 7(4).
        Social work, work science and sociology:
        Umeå University, Utvärdering (sid 427-438).

Shaw, I. (2005), "Practitioner Research: Evidence or Critique?" In *Br. J. of Social Work.* 2005:35, pp. 1231-1248, 18 s.
        Social work, work science and sociology:

Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Shaw, I & Lishman, J. (1999). *Evaluation and social work practice*. London: Sage.
Social work, work science and sociology:
Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Sielbeck-Bowen, Kathryn et al (2002) Exploring Feminist Evaluation: The ground from which we rise. In Feminist Evaluation. Explorations and Experiences. New Directions for Evaluation Nr 96.
Social work, work science and sociology:
Umeå University, Utvärdering (sid 3-8, 109-113).

Silverman, David (2004) 2nd edition, *Doing qualitative research : a practical handbook* London : SAGE.
Social work, work science and sociology:
Örebro University, Dokumentation och utvärdering i socialt arbete.

Sjögren, J. (2003). *Forskning i praktiken – för vems bästa?* C-uppsats vid institutionen för socialt arbete, Stockholms universitet.
Social work, work science and sociology:
Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.

Skolverket (1998), *Kvalitetssäkring i skolan.* Skolverkets rapport 150. Stockholm: Liber Distribution Publikationstjänst.
Pedagogik:
Uppsala University, Utvärdering och kvalitetsutveckling.

Skolverket (1999), *Att utvärdera skolan.* Stockholm: Liber Distribution Publikationstjänst (eller www.skolverket.se).
Pedagogik:
Göteborg University, Kvalitet och utvärdering inom utbildning.

Skolverket (1999), *Skolverkets allmänna råd om kvalitetsredovisning inom skolväsendet. Allmänna råd 1999:1.* Stockholm: Liber Distribution Publikationstjänst.
Pedagogik:
Uppsala University, Utvärdering och kvalitetsutveckling.

Socialhögskolan i Lund. FORSA, Förbundet för forskning i socialt arbete.
Social work, work science and sociology:
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).

Socialstyrelsen (2000), *Nationellt stöd för kunskapsutveckling inom socialtjänsten.* SoS-rapport 2000:12. Stockholm: Socialstyrelsen.
Social work, work science and sociology:
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).

Socialstyrelsen (2003), *För en kunskapsbaserad socialtjänst.* Redovisning av Regeringsuppdrag, [elektroniska resurs]110 s.
Social work, work science and sociology:

Stockholm University, Socialt arbete; kunskapsbas och interventioner.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Socialstyrelsen (2003), *Brukarmedverkan i socialtjänstens kunskapsutveckling,* [elektroniska resurs]*109 s.*
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Socialstyrelsen (2006), *Social rapport,* [elektroniska resurs]370 s.
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Socialstyrelsen (2006), *Nationella riktlinjer för missbruks- och beroendevård.*
Remissversion 11 april 2006, [elektroniska resurs] 176 s.
Social work, work science and sociology:
Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Socialvetenskaplig Tidskrift (2002), *Temanummer(2-3) om socialt arbete och utvärdering.*
Social work, work science and sociology:
Ersta Sköndal University College, Utvärdering och kvalitetsutveckling.
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).

Socialvetenskaplig Tidskrift. Årgång 9 nr 2-3 2003 [Tidskriftsartikel]
Social work, work science and sociology:
Örebro University, Dokumentation och utvärdering i socialt arbete.
Stockholm University, Utvärdering och utveckling i socialt arbete.

SOU 2000:1, 2000. En uthållig demokrati. Politik för folkstyrelse på 2000-talet. S 14-49.
Political science and peace and development research:
Lund University, Utvärdering av offentlig verksamhet.

Statskontorets rapport (2002:21) Utvärdering och politik II – Hur använder regeringen utvärderingar?
Social work, work science and sociology:
Umeå University, Utvärdering (sid 7-79).

Stenbock-Hult, Bettina (2003), *Kritiskt förhållningssätt.* Studentlitteratur, Lund.
Social work, work science and sociology:
Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Studies in Educational Policy and Educational Philosophy 2002:2 , 2003,
http://www.upi.artisan.se/Pages/cgi-bin/PUB_Latest_Version.exe?allFrameset=1&pageId=2.
Pedagogik:
Uppsala University, Pedagogik med inriktning mot personalutveckling, fortsättningskurs B/ Pedagogik, fortsättningskurs B, Styrning och utvärdering.

Studies in Educational Policy and Educational Philosophy 2003:1-2: Temanummer: Utvärdering forskning och politik, del 1 och 2, 2003. Uppsala: Uppsala Universitet. http://www.upi.artisan.se.
> Pedagogik:
> Mälardalen University, Utvärdering ur samhällsvetenskapligt perspektiv D.

Studies in Educational Policy and Educational Philosophy 2003:3, 2003, http://www.upi.artisan.se/Pages/cgi-bin/PUB_Latest_Version.exe?allFrameset=1&pageId=2.
> Pedagogik:
> Uppsala University, Pedagogik med inriktning mot personalutveckling, fortsättningskurs B/ Pedagogik, fortsättningskurs B, Styrning och utvärdering.

Svenska språknämnden (2003), *Svenska skrivregler.* Liber.
> Social work, work science and sociology:
> Örebro University, Dokumentation och utvärdering i socialt arbete.

Thors Hugosson, Christina (red; 2003): *Värdera och utvärdera.* Pedagogiska Magasinets skriftserie, nr 2. Stockholm: Lärarförbundets förlag.
> Pedagogik:
> Göteborg University, Kvalitet och utvärdering inom utbildning.

Thurén, Torsten (2005), *Källkritik.* Stockholm: Liber, 2 rev upplaga 2005.
> Social work, work science and sociology:
> Örebro University, Dokumentation och utvärdering i socialt arbete.

Trinder, L. & Reynolds, S. (2000), *Evidence-based practice. A critical appraisal.* Oxford: Blackwell Science.
> Social work, work science and sociology:
> Stockholm University, Socialt arbete; kunskapsbas och interventioner.

Uexküll Katarina Hjördisdotter von, Lundström Lena (red) (2005), *Ett granskande samhälle. Meningar om utvärdering.* Sveriges Kommuner och Landsting. Stockholm.
> Social work, work science and sociology:
> Halmstad University, Arbetsvetenskap - Utvärdering och uppföljning i arbetslivet.

Wallén, G. 1996, Vetenskapsteori och forskningsmetodik. Lund: Studentlitteratur.
> Social work, work science and sociology:
> Växjö University, Socialt arbete – aktuell forskning, vetenskapsteori och forskningsmetodik.

Vedung, Evert (1998). *Utvärdering i politik och förvaltning*. Lund: Studentlitteratur.
> Pedagogik:
> Lund University, Pedagogik: Utvärdering och kvalitetsarbete ur ett lärandeperspektiv.
> Mälardalen University, Utvärdering B.
> Mälardalen University, Utvärdering ur samhällsvetenskapligt perspektiv D.
> Social work, work science and sociology:
> Stockholm University, Utvärdering av socialt arbete (Socialt arbete II).
> Örebro University, Dokumentation och utvärdering i socialt arbete.

Political science and peace and development research:
Halmstad University, Utrednings- och utvärderingsmetodik.
Lund University, Utvärdering av offentlig verksamhet.
Uppsala University, Organisering, implementering och utvärdering i offentlig
politik.

Vedung, Evert, Furubo, Jan-Eric & Sandahl, Rolf (2000) "Utvärdering i det svenska politiska
systemet. Åtta iaktagelser". I Nordisk Administrativ Tidsskrift Nr 2 2000.
Social work, work science and sociology:
Umeå University, Utvärdering (sid 113-130).

Vedung, Evert (2006): "Evaluation Research", i Peters, Guy & Pierre, Jon (2005): *Handbook
of Public Policy*. London: Sage.
Political science and peace and development research:
Göteborg University, Utvärderingsmetoder.

Weiss, C. H. (1998), *Evaluation.* New Jersey: Prentice Hall (2nd ed).
Social work, work science and sociology:
Stockholm University, Utvärdering av socialt arbete (Socialt arbete II) (valda
delar).

Weiss Carol H. (1998), *Evaluation – Methods for studying programs and policies*(2d ed).
Upper Saddle River: Prentice Hall.
Social work, work science and sociology:
Umeå University, Utvärdering.
Stockholm University, Utvärdering och utveckling i socialt arbete.

Willinsky, John (2001), *Only We Knew: Increasing the Public Value of Social Science
Research.* Routledge, London.
Political science and peace and development research:
Halmstad University, Utrednings- och utvärderingsmetodik.

von Otter, C. (2003:08). (red). *Ute och inne i svenskt arbetsliv. Forskare analyserar och
spekulerar om trender i framtidens arbete.* Stockholm: Arbetslivsinstitutet.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (ca 110
s, utvalda kapitel).

von Otter, C. (2004*). Aktivt arbetsliv. Om dagens behov och framtidens möjligheter.*
Nedladdningsbar: www.arbetslivsinstitutet.se/publikationer/detaljerad.asp?ID=1446
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Wolvén, L-E. (2000). *Att utveckla mänskliga resurser i organisationer*. Lund:
Studentlitteratur.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I.

Wrightsman, L. S., & Fulero, S. M. (2005). (2nd ed). *Forensic Psychology.* UK. Wadsworth.
Psychology:

Stockholm University, Rättspsykologi I.

Wächter, Rut (1998), *Utredning.* Nordstedts juridik, Stockholm, 20 sidor, i Denvall, Verner & Jacobson, Tord (red) Vardagsbegrepp i socialt arbete. Ideologi och praktik.
Social work, work science and sociology:
Örebro University, Dokumentation och utvärdering i socialt arbete.

Örtenblad, Anders. (2005). Of course organizations can learn! *Learning Organization.* Vol. 12 No. 2. p. 213-218.
Pedagogik:
Linkoping University, Utbildning och lärande: Arbetslivets pedagogik I (Internet-based version).
Uppsala University, Pedagogik med inriktning mot personalutveckling, baskurs A.

2. Attachments of case study 2.

Attachment 2.1: "Evaluation Theory Tree Re-examined", Marvin C. Alkin & Christina A. Christie.

Evaluation Theory Tree Re-examined

Marvin C. Alkin
University of California, Los Angeles

Christina A. Christie
Claremont Graduate University

Introduction

As we began thinking about this session we were reminded of a quote by Egon Guba that Marv has used many times. The situation was a meeting of the American Educational Research Association with presentations by some of his students, role playing different theorists, and indicating how they (as that theorist) would modify their point of view based upon the stimulus of ideas put forth by two Belgian philosophers. Egon, as a discussant to the student paper reflecting his views, said:

> "You see, in many ways I am not the "real" Egon Guba—at least not
>
> the one that Lindheim had in mind when she conducted her exercise.
>
> That Egon Guba was the one working and writing seven to 10 years
>
> ago….. my mind has changed about so many things that it is hard for
>
> me to recall what opinions I held then (1979, p. 139)."

This simple, but elegant, statement reminds us that theories change over time. People read things. Ideas get floated. Interpretations get modified. That is the way that evaluation theories develop and change over time.

So too with theory classification systems. The Evaluation Theory Tree that we presented in our recent evaluation book, *Evaluation Roots*, is such a system. Our view was that there were three basic elements in considering evaluation theories: use, methods, valuing. All theorists are concerned with the methods that will be employed in conducting the evaluation. All theorists recognize that evaluation is an enterprise that involves valuing (distinguishing it from most research). All theorists recognize that evaluations will be used in ways that impact upon programs. We postulated however, that theorists differed in the particular emphasis they placed on one or another of these dimensions. (We referred to these as "branches" of the Evaluation Theory Tree.) Theorists were categorized on the branch which best reflected their primary emphasis and placed on the branch in a manner that reflected some combination of history and/or influence.

As with theorists who modify their views over time, we also propose some changes to the Evaluation Theory Tree. Indeed, in the last chapter of the Roots book we already recognized changes from Chapter 2, stimulated by the entries written by various theorists. These changes were relatively minor. By and large, they were simply re-positioning sub-branches in different ways to reflect the extent to which a theorist on a particular branch tended towards the views reflected in another branch. Additionally, we had indicated two foundations for the Theory Tree: the tradition of social inquiry and the desire for accountability and control. Responding to Yvonna Lincoln's comments we changed one of the foundations to <u>social</u> accountability and fiscal control.

In this session we suggest some additional modifications of the Theory Tree shown in the Roots book (see Fig. 1). Fig. 2 shows the new tree as we now perceive it (of course our minds might change after comments made in this session). We will comment on each of these changes.

Use Branch Revisions

In its most recent incarnation, the Use branch underwent only minor revision and still closely resembles the Tree presented in Chapter 26 of the *Evaluation Roots* book. The slight modifications include repositioning Joseph Wholey to reflect an emphasis towards the Methods branch. Revisiting his material led us to consider this minor change. Also, we deleted Malcolm Provus from the branch because much of his work is now outdated, and has relatively little influence other more recent perspectives on the branch.

Methods Branch Revisions

More substantial changes were made on the Methods branch. Ralph Tyler was repositioned to a sub-branch to reflect the idea that while his theoretic point of view was in fact heavily methods related (objectives based evaluation) he was not a theoretical predecessor of those further up on the branch. His original positing was intended to reflect his influence on the field of educational evaluation (which was very significant), but upon further reflection we concluded that his overall influence on the methods branch, specifically, was less than his original position implied. Thus, we placed him on a small sub-branch near the base of the branch.

Donald Campbell continues to be the heart and soul of this branch both in the way that he directly impacted other theorists with his work on experiments, quasi-experiments, and validity, as well as the way in which his views provided a basis for counter point (particularly with respect to Lee Cronbach's view).

We have also deleted Ed Suchman from the branch. We recognize the important influence of Suchman's writing on evaluation – particularly in positioning Campbell's work prominently in the evaluation discourse. Nonetheless, despite his historical significance, it did not seem pertinent to continue to include him on the branch. Peter Rossi continues to be shown on the main portion of the branch as an influence on both Carol Weiss and, more particularly Huey Chen.

The most significant change on this branch was moving Tom Cook who previously had been positioned in a sub-branch flowing from Campbell. He was moved and is now positioned on the main branch and his sub-branch was deleted. A new sub-branch was created and we moved Bob Boruch to this new sub-branch. Boruch's sub-branch comes

directly out of Campbell and Cook. This placement is intended to reflect the steadfast influence of Campbell's random experiment (RCT) work on Boruch. The branch is offset on a sub-branch to depict and emphasize his departure from both Cook and Campbell in his (lack of) regard for quasi-experiments. This position also offsets Boruch from the others on the methods branch, all of whom have embraced quasi-experiments as an acceptable method for studying causality; although everyone on the branch would agree that the experiment is ideal if context and conditions permit.

Another major change on this branch is the addition of a new name (or should I say several new names). In the Roots book we said:

> "(T)heories included were able to be classified onto a single branch of the tree. … Whether some theories were not included due to their comprehensiveness or our conceptual inability is unclear. A particular example comes to mind: the work of Mark et al. (2000). These authors view social betterment as the ultimate objective of evaluation and present a point of view grounded in what they refer to as a 'common sense realist philosophy.'…. The very diversified nature of this perspective, while a great strength in presenting an understanding of evaluation, precludes its inclusion on the tree (p. 58-59)."

We admit that in this instance our conceptual analysis was incomplete. Our views in determining exclusion were based heavily on the (Mark et al., 2000) book. However in reflecting further on the writings of these authors we were struck by the "realist evaluation" monograph in New Directions of Evaluation (Henry et al., 1998). Clearly, it is appropriate to place Henry and Mark on the Methods branch of the tree in a location following Tom Cook.

Valuing Branch Revisions

The Valuing branch has received the most noteworthy and substantial revision. This branch has always been the most difficult to explain in relationship to its evolution. It is obvious that the branch was significantly influenced by the work of Scriven. After all, it is Scriven who proclaims that an evaluation is not evaluation without valuing—in his words, evaluation is the science of valuing (Scriven, 2001). It is work of the evaluator to make a value judgment about the object which is being evaluated. This thinking has shaped and defined the field. However, this branch also includes the work of those interested in social justice in evaluation as well as those who espouse the philosophy of subjectivity, that is, that there is no one objective reality. These theorists are also concerned with valuing, but the emphasis is on the extent to which values shape the evaluation. Questions as to whose values should shape the evaluation, why, and with what intent provide focus to their work. This is quite different from the type of valuing that concerns Scriven. With further careful examination it has become clear how these two foci differ.

With little surprise, examining the theoretical perspectives on the valuing branch within the context of the philosophy of science offers a revised understanding of how perspectives could be understood. The basic axioms of the post-positivist and constructivist paradigms offer a clearer framework to further categorize evaluation models on the valuing branch. Thus, we have elected to split the values branch in two, naming the left arm of the branch stretching toward the methods branch as valuing: post-positivist influence and the right arm, valuing: constructivist influence. It is important to stress the use of the word *influence* in our description of the newly reshaped valuing branch. That is, some perspectives

on the valuing branch are shaped more exactly by a paradigm, while with others only a paradigm's undercurrent is felt.

To clarify, positivism is generally understood to be an element of scientism and realist ontology. Views of science shifted, however, during the 20th century away from positivism into post-positivism. Post-positivists recognize that all observation is fallible and has error. Where positivists believed that the goal of science was to uncover the truth, the post-positivist believes that the goal of science is to attempt to measure truth, even though that goal cannot be obtained. Constructivism is one element of interpretativism, and ontologically takes a relativist stance. There is no single, tangible reality that can be approximated; there are only multiple, constructed realities. Epistemologically, constructivism views subjectivity as the only reality, that is, the only way the unknown can become known is through our own, individual, belief system.

Scriven's realist thinking about valuing is reflective of both the ontology and epistemology of the post-positivist paradigm. He does not argue against the idea that we should be seeking an objective truth about the object being evaluated. He in fact offers what he believes to be a comparatively unbiased method for obtaining truth about an objects' worth, and then advocates for the evaluator to make a value judgment after gathering the most credible evidence. Further, he does not reject the idea of using experiments to determine causality, but rather argues that there should be more than one method for determining causality (Donaldson and Christie, in press). Scriven's thinking pushed the field to consider valuing as a central feature of evaluation more so than anyone else. However, he has been repositioned on the valuing branch at the base of the post-positivist influence arm to reflect the ontology and epistemology of his perspective.

Stake's work respects Scriven's thinking, although he argues for using thick description to assess a program's worth via the case study method. The use of case study methods introduced the idea that value is bound to context and that context must be considered when determining value. Stake does however leave the valuing to the evaluator, and so does not seem to reject the idea that evaluation is a science of valuing (realist). It was Stake's inclusion of and emphasis on case study methods and program context that prompted a shift from realist to relativist inquiry models in evaluation. Thus, Stake sits on a bridge squarely between the split of the two emphases of the valuing branch. House's work brought significant attention to the inclusion of the values of the underrepresented to the evaluation process. Ontologically and epistemologically House's work is grounded in constructivist thinking however methodologically his work breaks from the basic axioms of the constructivist paradigm. That is, he is willing and likely to use quantitative methods when conducting evaluations. Thus, he too sits on the bridge to the right of Stake but not firmly on the constructivist branch. Eisner now is shown as an offshoot from Scriven's branch. Eisner is similar to Scriven in that he posits that the evaluator is the expert and thus determines the final value of a program. He differs from Scriven in that he believes the evaluator has the authority to judge a program merit because of one's expert understanding of the subject area (i.e., education, public health), rather than because of one's expertise as an evaluator. Because Eisner uses intense observation (which includes both quantitative and qualitative measures, but emphasizes qualitative) he is placed on an offshoot branch leaning toward the constructivist influence arm of the valuing branch. Wolf/Owens are placed on the post-positivist influence arm because their perspective does adhere to the principle that the obtainment of truth should be pursued.

Guba and Lincoln have been moved to the base of the constructivist influence arm to reflect the ontology and epistemology of their perspective which closely adheres to the basic tenets of the constructive paradigm. Their theoretical model serves as the foundation for a string of models that are influenced by the tenets of the constructivist paradigm, four of which have been added to the theory tree. First, House and Howe's Deliberative Democratic model has been included as its own model separate from the social justice work of House, which is already represented on the tree. Deliberative democratic evaluation is based on the ideals of House's earlier work on social justice in evaluation but has been influenced by the processes put forth in Guba and Lincoln's Fourth Generation Evaluation. Thus, Deliberative Democratic evaluation has been placed on the tree above Guba/Lincoln. Jennifer Greene has also been added to the tree, directly above House and Howe. Greene's value-engaged approach is rooted in deliberative democratic evaluation principles and procedures, however it places additional emphasis on framing the evaluation on stakeholders values and the use of mixed methods. Thus, her sub-branch tends in the direction of the post-positivist branch. Donna Mertens Inclusive approach is seen as a direct descendant of Guba/Lincoln but is unique in its emphasis on diversity and the inclusion of diverse groups. The evaluation perspective associated with Barry MacDonald, represented as a sub-branch on the tree, has been strongly influenced by social justice concerns and constructivist principles.

And so, the above discussion of the attached Theory Tree represents the current thinking that we have about how to classify evaluation theories. This current picture of evaluation roots will help to guide our thinking about evaluation issues till further changes are made. Which may be tomorrow, or the next day.

Evaluation Theory Tree Re-examined

Use   Methods   Valuing
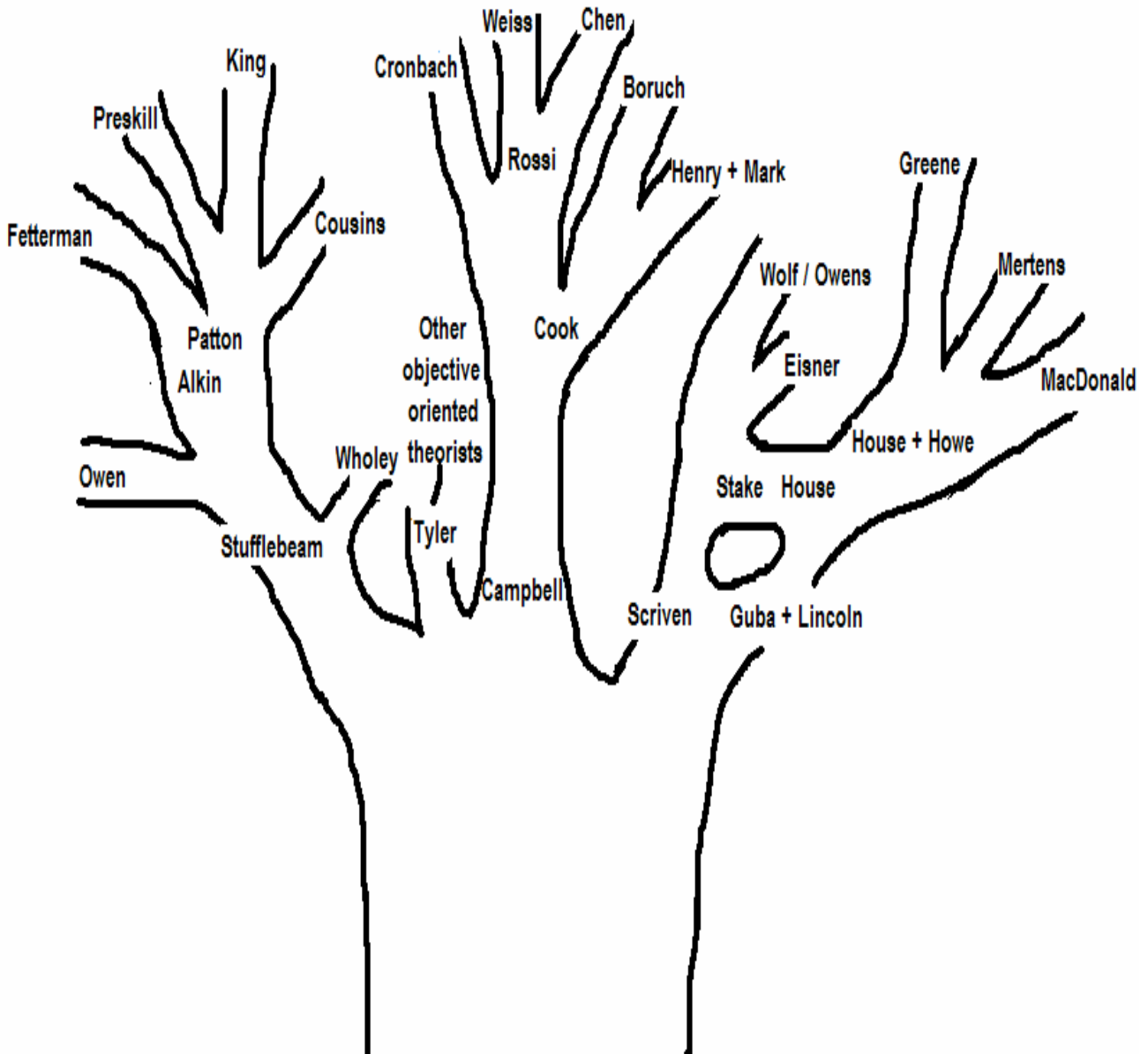
Attachment 2.2: Minutes of session March 2 2006.

Gustav Petersson 2006-03-15

Arbetsgruppen för utvärdering i Europa
Jan-Eric Furubo
Maria Bergström
Rolf Sandahl
Evert Vedung
Ove Karlsson
Gustav Petersson

The Roots of European Evaluation – *arbetsläge*.
Inom den svenska delen av projektgruppen har förslaget diskuterats att fokusera inte främst på att beskriva det europeiska utvärderingsträdets grenverk utan istället på dess rötter. Resultatet av forskningsprojektet blir i sådana fall inte en europeisk motsvarighet till Alkins arbete, eftersom dennes insats främst består i att ha presenterat utvärderingsträdets grenverk.

Jämförelsepunkt.
Den amerikanska utvärderingstraditionen har föreslagits som jämförelsepunkt till den europeiska. En svårighet som har påpekats med detta val av jämförelsepunkt är att idéutbytet mellan de båda områdena tidvis varit relativt stort. Detta gällde exempelvis under 1960-talet då Johnson-administrationens utvärderingsmetoder exporterades till Europa. Detta behöver emellertid inte betraktas som ett problem om antologin inriktas främst på skillnader mellan europeisk och amerikansk utvärderingstradition.

Ett annat alternativ är att utnyttja likheterna i utvärderingstradition och utifrån detta diskutera varför de gemensamma ansatserna och modellerna inte har varit lika framgångsrika på båda sidor om havet: Hur skiljer sig förutsättningarna?

Andra förslag?

Två typer av rötter.
Projektgruppens arbete ska presenteras i form av en antologi. Diskussionen om skillnader mellan Europa och USA har framför allt rört sig på tre olika nivåer. För det första har utvärderingstraditionerna i sig diskuterats. För det andra har två olika typer av rötter diskuterats; dels de formella institutioner som har påverkat utvecklingen av den europeiska utvärderingstraditionen, dels de idéhistoriska rötterna (de informella institutionerna). Med "utvärderingstradition" förstås här för det första utvärderingsdiskursen, för det andra den faktiska utvärderingen, för det tredje de formella offentliga institutionerna för utvärdering och för det fjärde användningen av utvärderingsresultaten. Med den faktiska utvärderingen avses sådant som val av ansatser, modeller och problem.

Två olika perspektiv.
En artikel i antologin om det europeiska utvärderingsträdets rötter skulle i princip kunna struktureras på två olika vis.

Från rötterna till kronan.

Det ena alternativet innebär att varje kapitel i antologin belyser en rot till det europeiska utvärderingsträdet, det vill säga en formell eller informell institution som har haft betydelse för utformningen av den europeiska utvärderingstraditionen. Kapitlet besvarar frågan: "*Hur har denna institution påverkat den europeiska utvärderingstraditionen?*" Enligt detta förslag börjar alltså författaren i rötterna och arbetar sig upp till kronan. Ett exempel på en informell institution av betydelse skulle kunna vara den franska revolutionens tanke om folket som principal i egenskap av kollektiv (i motsats till den amerikanska synen på det politiska systemet som baserat på en grupp av individer). Andra exempel på informella institutioner som skulle kunna tänkas ha haft betydelse för skillnader mellan utvärderingstraditionerna i de båda områdena är den pragmatisk-emiristiska idétraditionen i den anglosaxiska världen och den rationalistiska i Europa.

Ett exempel på en formell institution av betydelse skulle kunna vara EU. Vilken betydelse har valet inom EU av vissa utvärderingsmodeller haft för den europeiska utvärderingstraditionen? En annan utgångspunkt skulle kunna vara skillnaderna mellan statens samhällsrollroll i Europa och USA: Ser utvärderingstraditionerna olika ut i nattväktar- och välfärdsstater? En nära relaterad fråga är huruvida den större diversifieringen i Europa vad gäller formella politiska institutioner har skapat en rikare utvärderingstradition.

Ett speciellt fenomen som kanske skulle förtjäna uppmärksamhet som rot på det europeiska utvärderingsträdet är globaliseringen. Frågan är emellertid om det finns skäl anta att den europeiska utvärderingstraditionen påverkats annorlunda av detta fenomen än den amerikanska.

<u>Från kronan till rötterna.</u>
Det andra alternativet innebär att varje kapitel i antologin belyser något typiskt drag i den europeiska utvärderingstraditionen. Kapitlet besvarar frågan: "Vilka rötter har detta den europeiska utvärderingstraditionens kännetecken?" Enligt detta förslag börjar alltså författaren i det europeiska utvärderingsträdets krona och arbetar sig ner till rötterna.

En jämförelsepunkt som har diskuterats är inom vilka policydomäner utvärdering främst bedrivs. Det har pekats på att utbildningssektorn har varit dominant vad gäller amerikansk utvärdering, medan även socialpolitik och genom EU:s försorg regionalt utvecklingsarbete varit av stor betydelse. Hur kan dessa skillnader förklaras ur ett institutionellt eller idéhistoriskt perspektiv?

En annan tänkbar jämförelsepunkt är den mer frekventa användningen av experimentella metoder i USA, vilken skulle kunna förklaras med ett större behov att legitimera statliga interventioner än i det relativt etatistiska Europa.

Även den större tillämpningen av inomvetenskaplig diskurs i amerikansk utvärdering än i europeisk skulle kunna vara utgångspunkt för en orsaksdiskussion. Den mer frekventa användningen i USA av kvantitativa metoder inom exempelvis psykologi skulle kunna vara en förklaring.

Det skulle vidare kunna diskuteras huruvida användningen av utvärderingsresultat i den politiska processen ser annorlunda ut i Europa än i USA. Har det till exempel spelat en märkbar roll i detta avseende att den amerikanska högsta domstolen till del handhar frågor som löses på den politiska arenan i Europa?

Regionala skillnader.
En ytterligare fråga som har diskuterats är huruvida projektet även bör spegla regionala skillnader inom Europa eller begränsas till det som är gemensamt.

Metod.
En möjlig metod att fånga in vad som är typiskt för europeisk utvärdering är att gå igenom de senaste årgångarna av europeiska *Evaluation*. Det som diskuteras i denna tidskrift skulle kunna jämföras med motsvarande amerikansk tidskrift.

Alternativa metaforer.
Alternativ till Marvin Alkins trädmetafor har kommit upp under diskussionens gång.

1. Metaforen om fyra sedimenterande böljor. Som första bölja har presenterats 1950- och 60-talets från den amerikanka försvarsindustrin emanerande experimentalistiska utvärderingsvåg, vilken nådde Sverige omkring 1965. Det har emellertid påpekats att även dessförinnan fanns en utvärderingstradition bland svenska pedagoger. En andra utvärderingsbölja följde med *new public management*. Utvärdering blev i denna tradition ett medel att reglera graden av decentralisering. Den tredje utvärderingsböljan var den participatoriska som under 1990-talet följde med kritiken mot det representativa systemets funktionssätt. Den fjärde böljan är den evidensbaserade med ursprung i socialmedicinen. En viktig skillnad mellan denna bölja och den första är att roller idag blandas på ett vis som inte skett tidigare; läkaren är både forskare och vårdare. Samtliga dessa fyra böljor har lämnat spår efter sig i dagens svenska offentliga sektor.

2. Istidsmetaforen om att olika samhällsklimat tinar upp olika delar av den samlade idémängd om utvärdering som ackumulerats i ett samhälle.

3. Generationsmetafor för beskrivning av de olika vågorna av nytänkande inom utvärdering.

Kritik har av flera skäl riktats mot trädmetaforen, bland annat av det skälet att den i vissa sammanhang blir direkt missvisande eftersom inte samtliga delar av den europeiska utvärderingsdiskursen har gemensamma eller ens kompatibla rötter. Detta blir särskilt tydligt på så vis att det postmoderna tänkandet skiljer ut sig från det övriga idégodset: tankar om den ooverblickbara kontexten, skepticism mot människans allmännyttighet och att participatorisk utvärdering har ett värde även om den skulle sakna instrumentella fördelar står i skarp kontrast till 1950-talets samhällsplanerande idémiljö. Problemet kan emellertid helt enkelt lösas så att det europeiska utvärderingsträdet delas upp på flera.

Arbetsfrågor
Oavsett hur arbetsresultatet ska struktureras och avgränsas så bör följande frågor diskuteras:
1. Vilka är de typiska dragen för europeisk utvärdering?
2. Vilka rötter kan dessa fenomen tänkas ha?
3. Vilka institutionella eller idéhistoriska skillnader mellan Europa och USA skulle kunna tänkas ha genererat skillnader mellan de båda områdenas utvärderingstraditioner?
3. Kan de konstaterade typiskt europeiska dragen sägas vara starkare förankrade i vissa europeiska länder än i andra?

Attachment 2.3: Minutes of session March 20 2006.

Gustav Petersson 2006-03-22

Arbetsgruppen för utvärdering i Europa
Jan-Eric Furubo
Maria Bergström
Rolf Sandahl
Evert Vedung
Ove Karlsson
Gustav Petersson

The Roots of European Evaluation – *arbetsläge*.
På det föregående mötet (2006-03-02) strukturerades diskussionen huvudsakligen utifrån att gruppens arbete syftar till att åstadkomma en beskrivning av det europeiska utvärderingsträdets rötter. På det senare mötet (2006-03-20) rörde diskussionen till stor del grenverket. Det konstaterades att i arbetet med att åstadkomma en beskrivning av det europeiska utvärderingsträdets grenverk kan en jämförelse med USA vara fruktbar. Emellertid borde även skillnader *inom* Europa belysas.

Institutionellt fokus.
Vid det föregående mötet gjordes en distinktion mellan för det första utvärderingsdiskursen, för det andra den faktiska utvärderingen, för det tredje de formella offentliga institutionerna för utvärdering och för det fjärde användningen av utvärderingsresultaten. Med den faktiska utvärderingen avses sådant som val av ansatser, metoder, modeller och problem. På det senare mötet rådde enighet kring att några större skillnader mellan olika regioners utvärderingsdiskurser eller utvärderingspraktik förmodligen inte står att finna. Kanske ligger de största skillnaderna i hur utvärderingen organiseras. Emellertid kan även på detta område en tendens mot likriktning tänkas i och med att länder jorden runt utsätts för gemensamma *ratings*.

Jämförelse mellan amerikansk och europeisk utvärderingstradition.
Det konstaterades att den mest grundläggande skillnaden mellan amerikansk och europeisk utvärderingstradition torde vara att det i Europa finns myndigheter för utvärdering av offentlig verksamhet, medan tendensen att inrätta sådana varit svagare i USA. Där har istället företag utfört denna uppgift.

Skillnaden i var utvärderingen utförs skulle kunna förklaras med en skillnad i grundsynen på statens roll i samhället. Samma förklaring kan tänkas för en annan skillnad mellan europeisk och amerikansk utvärderingstradition som konstaterades på mötet, nämligen synen på vilket syftet är med att utvärdera offentlig verksamhet. Det framhölls att medan syftet i Europa till stor del är vara att förbättra den offentliga verksamheten så ligger betoningen i USA snarare på att kritiskt pröva den offentliga verksamhetens existens. Rekommendationerna i en utvärdering som ger en offentlig verksamhet dåligt betyg tenderar därmed att divergera; den europeiska verksamheten borde stödjas, förbättras, förses med resurser, medan den amerikanska i grunden ifrågasätts. Den informella institutionen "grundläggande syn på statens roll i samhället" diskuterades alltså återkommande som rot på de båda områdenas respektive utvärderingsträd.

Eftersom utvärderingar av offentlig verksamhet i USA i högre grad än i Europa läggs ut på företag så erbjuder en jämförelse mellan de båda områdena en möjlighet att diskutera den principiella frågan om hur utvärderingars kvalité och innehåll påverkas av vem som utför dem. Vad gäller frågan om utvärderingens kvalité konstaterades att inom olika samhällssfärer olika kriterier framhävs på en god utvärdering. På marknaden skulle utvärderingens pris kunna tänkas inta en framträdande position, medan det kritiska ifrågasättandet och forskningsanknytningen torde vara högre prioriterade värden inom akademien. Inom politiken skulle de främsta kvalitetskriterierna kunna tänkas vara i vilken grad en utvärdering kan användas i legitimerande eller informationsskapande syfte. Innebär detta att kriterierna för vad som är en god utvärdering skiljer sig åt mellan Europa och USA? Har det amerikanska beroendet av marknaden och det europeiska beroendet av utvärderande myndigheter fostrat olika dominerande kvalitetskriterier? Resonemanget knöt an till Cambridge-skolans tanke att värden kommer till uttryck redan i valet av utvärderingsdesign. Om olika implicita värden och explicita kriterier dominerar europeisk och amerikansk utvärdering ser då metod- och teorivalen olika ut?

Jämförelse mellan utvärderingstraditioner inom Europa.
Det konstaterades att en jämförelse mellan olika regioner inom Europa också skulle vara berikande eftersom skillnader finns mellan exempelvis tysk, fransk, spansk och brittisk utvärderingstradition. I samband därmed diskuterades skillnader i synen på välfärdsstaten och Joachim Voglers distinktion mellan marknads-, stats- och familjebaserad välfärdsstat. Har skillnader i välfärdsstatens uppbyggnad och omfattning skapat skillnader i synen på utvärdering av offentlig verksamhet? Vilka skillnader finns exempelvis med avseende på vem som utvärderar och varför?

Metod.
Det har tidigare konstaterats att en möjlig metod att fånga in vad som är typiskt för europeisk utvärdering är att gå igenom de senaste årgångarna av europeiska *Evaluation*. Vid det senare mötet framfördes tanken att även ratade artikelförslag som inskickats till denna tidskrift skulle kunna studeras.

Invändningen gjordes emellertid att i engelskspråkiga tidskrifter är författare från engelskspråkiga länder överrepresenterade, medan författare från exempelvis romanska och tyskspråkiga områden kan tänkas vara underrepresenterade. Gäller samma icke-representativitet på internationella konferenser? Hur ska språkproblemet hanteras?

Attachment 2.4: Minutes session April 13, 2006.

Gustav Petersson 2006-04-13

Arbetsgruppen för utvärdering i Europa
Jan-Eric Furubo
Maria Bergström
Rolf Sandahl
Evert Vedung
Ove Karlsson
Gustav Petersson

The Roots of European Evaluation – *arbetsläge*.
På det föregående mötet (2006-03-20) konstaterades att i en beskrivning av det europeiska utvärderingsträdets grenverk kan en jämförelse med USA vara fruktbar. Emellertid borde även skillnader *inom* Europa belysas. På det senaste mötet (2006-04-13) preciserades vilka delar i grenverket som är bör skildras. Det bestämdes att arbetet framöver främst skulle syfta till att beskriva skillnader mellan olika områden, såväl det europeiska som de subeuropeiska, vad gäller utvärderingens roll i samhället. Det betonades att detta inte är synonymt med att skildra utvärderarens roll. En ytterligare precisering som beslutades var att främst utvärdering av statlig verksamhet ska studeras. Sammanfattningsvis är alltså det som ska studeras dels formella och informella institutioner för utvärdering i staten, dels den verksamhet som i praktiken bedrivs inom respektive i anslutning till dessa.

Arbetsredskap.
Tänkbara redskap med vilka arbetet kan struktureras diskuterades också.

Ett diagram vars axlar beskriver individers (x-axel) respektive centralmakters (y-axel) valfrihet var det första förslaget. Exempelvis i Sovjet var medborgares möjligheter små att välja former och källa för distribution av materiell välfärd. Alltså hamnar detta exempel nära y-axeln, långt från x-axeln (högt y-värde, lågt x-värde). Ett system av reglerad valfrihet bör istället placeras närmre x-axeln. Som exempel kan ett utbildningssystem med ett inslag av friskolor nämnas. Bland de statliga kraven finns att friskolor inte får grundas på vilka värderingar som helst, men när skolorna väl grundats kan staten endast indirekt påverka medborgarnas val av utbildningsplats. Utvärderingar av utbildningsväsendet kan i dessa båda system antas fylla olika funktioner i samhället och utföras inom olika formella institutioner.

Ett annat arbetsredskap som har diskuterats vid ett tidigare möte är en typologi över olika typer av välfärdsstater. Ett möjligt val av typologi är Gösta Esping-Andersens omdiskuterade indelning av de europeiska välfärdsstaterna i socialdemokratiska, liberala och konservativa (*Three worlds of welfare capitalism*, 1990). Något förenklat kan dessa beskrivas som statsbaserade, marknadsbaserade och familje-/kyrkobaserade. Formar de olika typerna av välfärdsstater olika institutioner för utvärdering? Formar de olika utvärderingspraktik? Är syftena med utvärderingarna olika i de olika typerna av välfärdsstater?

Detta väcker frågor som till exempel de följande. Den engelska marknad där utvärderingsföretag florerar är större än motsvarande marknad i Sverige. Ska denna skillnad förstås utifrån skillnader i välfärdsstaternas strukturer? En andra inomeuropeisk skillnad som skulle kunna ha implikationer för utvärdering är att de sydeuropeiska välfärdsstaterna i högre

grad är familje- och kyrkobaserade än de skandinaviska. Innebär detta att de frågor som ställs i utvärderingar av välfärdsdistribution i samhället inte är desamma i Skandinavien som i Sydeuropa? Frågar en italiensk utvärderare efter huruvida kyrkobaserade alternativ ges rimliga möjligheter att existera vid sidan av de statliga då en svensk istället exempelvis ifrågasätter kostnadseffektiviteten i den statliga verksamheten?

Teoretiska ansatser.
Åter nämndes spänningen mellan å ena sidan amerikansk pragmatism och *Realistic Evaluation* och å den andra socialkonstruktivistiska ansatser, vilka har sitt ursprung främst i Frankrike. Temat elaborerades inte närmare.

Tidigare litteratur.
En diskussion fördes kring vilken litteratur som skrivits i liknande anda. Följande verk nämndes:
1. Furubo/Rist/Sandahl: *International Atlas of Evaluation*.
2. Helmut Vollman lär ha skrivit en antologi om utvärdering, men denna finns inte på något bibliotek i Sverige. (Detta oavsett om efternamnet stavas med ett eller två "n".)
3. Scriven/Davidson: *Journal of Multi-Disciplinary Evaluation*, 3 stycken nummer på cirka 200 sidor vardera, vilka finns att ladda ner gratis från nätet.
4. Frans Leeuw menades ha skrivit med ambitionen att övergripande beskriva skillnader mellan utvärderingstraditioner i olika områden.
5. Hemsidan för *Center of Evaluation*, USA, (Stufflebeam) lär innehålla liknande material.

Attachment 2.5: Minutes of session June 11 2006.

Den europeiska utvärderingens dimensioner – arbetsläge.
Gustav Petersson 2006-06-11

Arbetsgruppen för utvärdering i Europa
Jan-Eric Furubo
Maria Bergström
Rolf Sandahl
Evert Vedung
Ove Karlsson
Gustav Petersson

Det har tidigare konstaterats att i diskussioner som syftar till att klarlägga den europeiska utvärderingstraditionens karaktär kan en jämförelse med USA vara fruktbar. Emellertid bör även skillnader *inom* Europa belysas, eftersom det tycks troligt att det politiskt och kulturellt splittrade Europa rymmer olika utvärderingstraditioner. Tre olika relevanta aspekter av ett lands utvärderingstradition har pekats ut: utvärderingens diskurs, formella institutioner för utvärdering samt utvärderingens praktik. Att belysa skillnader mellan länder på dessa områden är inte synonymt med att skildra utvärderarens roll.

Inom gruppen har vi valt att fokusera främst på utvärdering av statlig verksamhet. Det har emellertid inte klargjorts om även EU ska studeras. En jämförelse skulle kunna göras mellan EU och den amerikanska federala nivån.

Tre aspekter av utvärdering.
De tre aspekter av utvärdering kring vilka de tidigare diskussionerna har rört sig presenteras nedan.

1. Utvärderingens diskurs.
Med utvärderingens diskurs avses för det första det Alkin behandlar i *An evaluation tree for Europe*; diskussionen om utvärdering med avseende på metoder/tekniker, kriterier och användning. Dessutom har inom gruppen lyfts fram bland annat diskussionerna om det följande. Designfrågan och effektproblemet. Vad som bör utvärderas i vilket sammanhang och i sådana fall hur; program, projekt, policy, person, produkt etc. Vilka definitioner bör användas? Vad är utvärdering? Vilka värderingar och ställningstaganden ligger i att bedriva utvärdering generellt och i att använda olika designer i synnerhet? Vilka värderingar ligger i att utvärdera ett projekts eller programs kontext? Hur bör utvärdering av nätverksbaserade eller kontraktsbaserade projekt genomföras?

2. Institutioner för utvärdering.
Med institutioner för utvärdering avses formella institutioner i staten för utvärdering. Det handlar alltså om myndigheters roll i utvärdering av offentlig verksamhet. Det bör studeras om det i olika länder är myndigheter som utför utvärderingar. Är det i sådana fall samma myndigheter som har varit ansvariga för de utvärderade projekten eller programmen? I vilka länder utförs utvärderingar av offentlig verksamhet av institutioner, vilka enbart sysslar med utvärdering? Vilken bakgrund har de personer som utför utvärderingarna?

Om inte, hur går finansieringen till? Är det den myndighet som har haft ansvarat för det projekt eller program som ska utvärderas som finansierar utvärderingen? Så är exempelvis fallet i Storbritannien och i Nederländerna, men inte i USA. I USA är alltså på så vis utrymmet för statlig styrning mindre. Är det den myndighet som har ansvarat för projektet/programmet som ordnar upphandling när utvärderingar läggs ut på entreprenörer?

Blir graden av kritiskt ifrågasättande mindre ju större inflytande den myndighet har över utvärderingen som har varit ansvarig för det utvärderade programmet/projektet? Undviks i sådana fall vissa specifika frågor?

Finns skillnader mellan länder vad gäller graden av centralisering och skalan på utvärderingssystemen?

Det bör tas ställning till huruvida enbart utvärderingar initierade av offentliga institutioner ska inkluderas i studien. Utvärderingar, i en bred bemärkelse, initieras även av andra organisationer, exempelvis fackliga organisationer, pressen och SNS.

3. Utvärderingens praktik.
Skiljer sig syftena med utvärderingarna åt mellan länder? Innebär detta att olika frågor ställs? Påverkar detta i vilken mån utvärdering utförs *ex post* eller *ex ante*? Genomförs utvärderingar i system eller individuellt?

Skiljer sig utvärderingarnas design åt mellan länder? I vilken mån används intressentmodellen? Används den i participatoriskt eller i kunskapsbefrämjande syfte? Används brukarutvärderingar i högre utsträckning i mer liberala stater än i övriga studerade? I vilken mån används evidensbaserad utvärdering? Används samma metoder eller finns det skillnader mellan länder med avseende på användningen av kvantitativa/kvalitativa metoder? Är enkätmetoder vanligare i liberala stater än i övriga studerade?

Används utvärderingsresultaten? I vilken utsträckning och av vem?

Utförs metautvärderingar? Fokuserar dessa på de tidigare utvärderingarnas metodologi eller resultat?

Förklarande variabler.
Skillnader mellan länder eller regioner med avseende på de tre presenterade aspekterna av utvärderingstraditioner skulle kunna förklaras med skillnader på nedanstående områden. Anta exempelvis att en skillnad mellan USA och Sverige i synen på statens roll i samhället anses förklara någon skillnad mellan de båda ländernas utvärderingstraditioner inom en eller flera sektorer. Då utgör den svenska synen på statens roll i samhället en relevant rot på det svenska utvärderingsträdet, medan den amerikanska synen är en relevant rot på det amerikanska trädet.

1. Vilken som är den dominerande synen på statens roll i samhället. Om policyprocessen omfattar implementeringsfrågor så kan det tyda på låg legitimitet i samhället för statliga interventioner. Det torde påverka bland annat vilka syften utvärderingar har och kanske vilka frågor som ställs.
2. Huruvida befolkningen upplever sig som relativt homogen eller inte. Om olika identiteter finns representerade så påverkar detta uppfattningen om vad staten är, vilket torde påverka vad som utvärderas.

3. Huruvida det utvärderade politikområdet är kontroversiellt. Detta torde exempelvis påverka hur ifrågasättande utvärderingarna på området är. Anta att i det ena av de studerade länderna är befintligheten av friskolor kontroversiell, medan den inte är det i det andra landet. Syftet med fler utvärderingar i det första landet än i det andra torde då vara att i grunden ifrågasätta verksamhetens befintlighet.

4. Hur staten enligt konstitutionen ska organiseras, exempelvis som federal eller unitär stat med eller utan författningsdomstol osv. Exempelvis kan tänkas att en författningsdomstol fyller vissa av utvärderingens syften. Färre egentliga utvärderingar torde då utföras och med mer begränsade syften. Den höga graden av decentralisering i federala stater torde exempelvis motivera utvärderingar med syftet att kontrollera hur stort inflytande den federala organisationen i praktiken utövar på delstaterna.

5. I vilken mån politiska beslut fattas i konsensus. Används exempelvis intressentmodellen vid parlamentets utskottsremittering? Den svenska SOU-modellen är här av relevans. Vid stor konsensus kan antalet utvärderingar tänkas vara mindre än vid mindre konsensus.

6. Omfattningen av utredningar och försöksverksamhet som föregår politiska program och projekt. Om denna omfattning är stor så kan antalet utvärderingar tänkas vara mindre. Försöksverksamheten har då redan skapat den legitimitet som en utvärdering skulle kunna ge.

7. Vilken relationen är mellan forskning, politik och administration. Tillfrågas experter vid beslutsfattande? I sådana fall erhålls infallsvinklar och kanske synpunkter på vad som bör utvärderas och hur, som annars inte kommer fram.

8. Förvaltningstjänstemäns bakgrund. Akademisk utbildning? Partibok? "Fälterfarenhet" från det område som styrs från den aktuella myndigheten? Det påverkar vad som är värt att utvärdera i efterhand, eftersom implementeringen av politiken kan tänkas ha olika svagheter beroende på vem som har implementerat.

9. Verkställighetssystem. I vilken utsträckning implementeras beslut av en klassisk hierarkiskt organiserad byråkrati? Genomförs de medelst kontrakt med privata aktörer? Används kontrakt mellan statliga och kommunala/delstatliga organ? Görs försök att integrera självorganiserade nätverk i implementeringen? Även detta kan tänkas påverka vilka svagheter implementeringen har och alltså vad som är värt att utvärdera.

10. Styrfilosofier. Vilka styrmedel den politiska nivån använder för att vinna kontroll över implementeringen påverkar vad som bör utvärderas.

11. Huruvida det utvärderade politikområdet har haft politisk prioritet i historien.

12. Ländernas ekonomiska utveckling under senare tid och hur de har reagerat på de krav på resultats mätbarhet som följt med *new public management*.

13. Alkins paradfaktor: utvärderingsforskares prestationer, vilka har format utvärderingstraditioner.

Tankemodell.
I detta stycke presenteras en tankemodell som kan vara fruktbar för den som gör empiriska studier till belysande av den svenska utvärderingstraditionens karaktär.

Eftersom utvärderingstraditionerna kan skilja sig åt mellan olika policysektorer så bör en författare i ett första steg avgränsa sin studie genom att välja en sådan sektor. Utvärderingstraditionen inom denna sektor kan undersökas med avseende på någon eller några av de presenterade aspekterna av utvärdering. Motsvarande undersökning kan genomföras för samma sektor i något annat land och därefter kan jämförelser som klarlägger skillnader göras. I sådana fall kan eventuella skillnader kanske förklaras med hjälp av de förklarande variabler som har presenterats ovan.

Tankemodellen illustreras av nedanstående diagram. I den översta raden presenteras valen av länder och policysektor. Under denna rad presenteras i tre olika kolumnhuvuden de tre olika aspekterna av utvärdering. I dessa tre kolumner presenteras observerade skillnader.

	Skillnader i utvärderingstradition inom sektor x mellan länderna 1 och 2		
Jämförelse-aspekter För-klarande faktorer	Skillnad i diskurs (Vad som sägs)	Skillnad i praktik (Vad som görs)	Skillnader mellan formella institutioner (Vem som gör)
Förklaringsfaktor 1			
Förklaringsfaktor 2			
etc.			

Alternativa jämförelser har också diskuterats. Istället för att jämföra utvärderingstraditionerna i två länder inom samma sektor skulle utvärderingstraditionen i en sektor i Sverige vid olika tidpunkter kunna jämföras. Ett annat alternativ vore att jämföra utvärderingstraditionerna i två sektorer i Sverige idag. Ett sådant upplägg har nackdelen att observerade skillnader kan förklaras med att verksamheterna är olika till sin karaktär. Dessa jämförelser klargör inte den svenska utvärderingstraditionens karaktär. Den senare skulle däremot kunna ge belägg för tesen att vissa sektorer på grund av verksamhetens karaktär i högre grad än andra är utvärderingsdrivande. Ett exempel på en sektor vars verksamhets karaktär skulle kunna tänkas vara utvärderingsdrivande är utbildningssektorn.

Slutligen skulle också, i mån av generaliserbarhet, sektorsavgränsade jämförelser mellan Europa och USA kunna göras.

Metod.
Undersökningar skulle kunna göras genom intervjuer med forskare. Vad har varit typiskt för diskussionen om utvärdering av specificerade sektorer i deras respektive länder? Hur går dessa forskare till väga vid utvärderingar av dessa sektorer? Vilka institutioner för utvärdering i staten finns inom dessa sektorer i de olika länderna?

Tidigare litteratur.
Lästips som kan stimulera diskussionen:

1. Furubo/Rist/Sandahl: *International Atlas of Evaluation*. Furubo/Sandahl konstaterar att det finns en positiv korrelation mellan utgifter för utvärdering och graden av välfärd. I detta verk även Derlien, Hans-Ulrich och Ray C. Rist, "Policy Evaluation in International Comparison". De hävdar att huvudsyftet med policyutvärdering har skiftat från att informera om hur implementeringens slutprodukter kan förbättras till att vara ett redskap för reallokering av resurser mellan program.

2. Helmut Vollman: "Evaluation and evaluation research". Ska publiceras i Fisch, Frank, Gerald Miller och Mara Sidney (Eds.), *Handbook of Public Policy Analysis. Theory, Politics and Methods*. New York, Marcel Dekker Publisher. Diskuterar vad det innebär att utvärdera samt utvärderingens definitionsflora och historia.

3.Scriven/Davidson: *Journal of Multi-Disciplinary Evaluation*, tre nummer på cirka 200 sidor vardera, vilka finns att ladda ner gratis från nätet.

I det första av dessa nummer redogörs för Elliot Sterns uppfattning att utvärdering i Europas stater har formats av fyra faktorer (ss. 69 f):
- huruvida det i det aktuella landet finns en stark nationell identitet eller befolkningen snarare är splittrad i grupper med olika identiteter.
- graden av kulturell diversifiering och solidaritet.
- graden av politisk decentralisering.
- tilltagande komplexitet och nationalstatens problem att möta denna.

I det första numret redovisas också en studie över vilka som är de dominerande temana i *American Journal of Evaluation* (Wingate, Lori A., 2004, "What's Happening in AJE (2003-2004)", ss. 103-113.) EN motsvarande redovisning finns också för *Evaluation* (Schröter, Daniela C., 2004, "Evaluation: The International Journal of Theory, Research and Practice (2003-2004), ss. 114-120.)

4. Frans Leeuw hävdar i "Evaluation in Europe" (i Stockmann, Reinhard (Ed.), 2004, *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Andra upplagan. Opladen; Leske + Budrich) att marknaden för utvärdering växer fortare inom EU och på nationalstatsnivå än regionalt.

5. Leeuw, Frans, 2003, "Trends, Topics and Theories" i The Danish Evaluation Institute (eds.), 2003, *Educational Evaluation around the world*. Köpenhamn.

6. Hemsidan för *Center of Evaluation*, USA, (Stufflebeam).

7. Evert Vedung: "Utvärdering som megatrend, gigatrend och fyra böljor." Publicerad?

8. Nicoletta Stame: "Evaluation and the policy context: the European experience", article presented at the 2003 Conference of the Australasian Evaluation Society. Författaren hävdar att EU:s utvärderingssystem fokuserar i högre grad på hur mycket pengar som läggs på utvärdering än på hur effektivt de används. Detta leder till att EU:s utvärderingsprojekt inte fungerar kunskapsbefrämjande. De tjänar inte heller ett ansvarsutkrävande syfte eftersom kopplingen mellan beställare (EU) och utförare är för lös. I utvärderingarna varken generaliseras utvärderingsresultaten eller erbjuds förslag på förbättringar. (s. 6 ff) Påverkas de nationella utvärderingstraditionerna i samma riktning?

9. Stockmann, Reinhard (Ed.), 2004, *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Andra upplagan. Opladen; Leske + Budrich)

10. Stockmann, Reinhard, 2006, *Evaluationsforschung und Qualitätsentwicklung – Eine Grundlage für wirkungorientiertes Qualitätsmanagement*. Münster; Waxmann Verlag.

11. Furubo, Jan-Eric, opublicerat manuskript, "Policy Analysis and Evaluation in Sweden – discovering the limits of the rationalistic paradigm". Chapter in Frank Fischer, Gerald Miller and Mara Sidney, Rutgers University (eds.) *Handbook of public Policy Analysis: Theory, politics and methods (forthcoming)*.

Om projektet ska fokusera både på utvärderingar initierade av offentliga institutioner och sådana som initierats av andra organisationer så bör Kommunförbundets rapport "Utvärderande aktörer" beaktas.

Attachment 2.6: "An Evaluation Tree for Europe", Gustav Jakob Petersson & Ove Karlsson Vestman.

2006-05-30

<div align="center">

Mälardalen Evaluation Academy

An Evaluation Tree for Europe

Gustav Jakob Petersson and Ove Karlsson Vestman

</div>

Abstract.
The purpose of this modified paper is to present to you an approach for elaborations on the differences between European and North American evaluation. We started with Marvin C. Alkin's evaluation tree as it is presented in **Evaluation Roots – Tracing Theorists' Views and Influences (2004***). We don't find it enough to put European names on his tree. We think that not only researchers' work should be discussed, but also the formal institutions for evaluation and the evaluation practices. These three branches we label the European evaluation tradition. However, there are major differences in evaluation traditions between European countries and regions. In order to explain these differences, as well as differences between Europe and the USA, we believe that the societal and historical national or regional contexts must be considered. This theme is elaborated and some suggestions made as to how the roots of the different European evaluation traditions could be described. The paper concludes with some questions, which will serve as discussion points at our June meeting in Stockholm.***

The Initiative.

With his book *Evaluation Roots – Tracing Theorists' Views and Influences* (2004) Marvin C. Alkin presented his perspective on evaluation. His interesting elaborations inspired us to take on a similar task. However, the objective of the ongoing research project *An evaluation tree for Europe* is not quite the same as Alkin's. Instead of wanting to describe evaluation in general we want to present a European perspective, or perhaps some European perspectives, on European evaluation in contrast to North American evaluation. On the EES conference of coming October the results of the discussions on this theme are to be presented in the form of a panel discussion. The group preparing this panel discussion consists of the panel participants – Marvin C. Alkin, Nicoletta Stame, Frans Leeuw, Reinhard Stockmann and Ove Karlsson Vestman – and additional Swedish researchers taking part in the discussions preceding the London EES conference.

The purpose.

Our purpose with this paper is to present to you an approach for describing the European evaluation tree and in particular how it differs from the North American one. Alkin's tree metaphor has undoubtedly got pedagogical strengths. However, we believe that if his tree branches were used our evaluation tree would look very much like the North American one. The reason for this belief is that the academic traditions, from which evaluation has originated, today are mainly international. For example, the empiricist tradition can today be

found in most Western countries. Therefore we want to give the branches a partly different content.

<u>Our branches</u>.
In our version the European evaluation tree has got three branches. The first is the discussion on evaluation, the evaluation discourse. This branch contains among other things what Alkin has split up on his three branches. The second branch describes the formal institutions for evaluation in Europe. The third mirrors evaluation practices in Europe. Together we label our three branches the European evaluation tradition. We do not believe that the major differences in evaluation traditions across the globe are to be found in the field of the evaluation discourses, but rather in the institutions for evaluation and how evaluations are launched in practice.

<u>The European evaluation trees</u>.
We don't think that the European evaluation tradition should be described with only one evaluation tree. During discussions within the Swedish part of the project group we have come to believe, that if a fair description shall be made, a comparison with the North American one is not enough. Indeed, one of the most striking characteristics of the European evaluation tradition seems to be the regional differences between formal institutions for evaluation. As Europe is politically and culturally diversified we also believe that there could be important differences in evaluation practices. Therefore we rather want to describe the European evaluation trees than the European evaluation tree.

Of course an evaluation tradition is formed by many different factors. First, there is consensus about some basic attributes of evaluation. That is to say, that what hasn't got these characteristics is not considered to be evaluation. For example; when you evaluate you gather facts and say something about them. We also know that there are international trends in evaluation. We also believe, however, that national history, political frameworks and other contextual factors play an important part in forming a national or regional evaluation tradition. This belief is starting point for the elaborations below. That is not to say that the tools suggested are considered the only possible ones.

A possible structure of an evaluation tree for Europe.
One contextual factor that we believe to be most important in forming a national or regional evaluation tradition is to what extent it is considered legitimate for the state to intervene in and regulate the activities of the population. That will influence how evaluation of public sector activities is undertaken. Therefore we think that the legitimacy of state intervention will form the evaluation tradition. Our elaborations on the branches of the European evaluation trees are based on this assumption and they exemplify our method to register differences between evaluation traditions. After these elaborations our thoughts on how the roots of the European evaluation traditions could be described are presented.

Of course we have considered your answers to the questions that were sent to you as the project began. These answers, as well as the articles that you most kindly have sent to us have been of great help in outlining our proposal. The regular meetings of the Swedish part of the project group have also been an important source for inspiration.

<u>The branches of the European evaluation tree</u>.

In our suggestion the structure for the European evaluation tree has got three branches and each of these branches has got its own sprigs and roots. Feel free to come up with suggestions for yet more branches.

The evaluation discourse   Institutions for evaluation   Evaluation practises

down
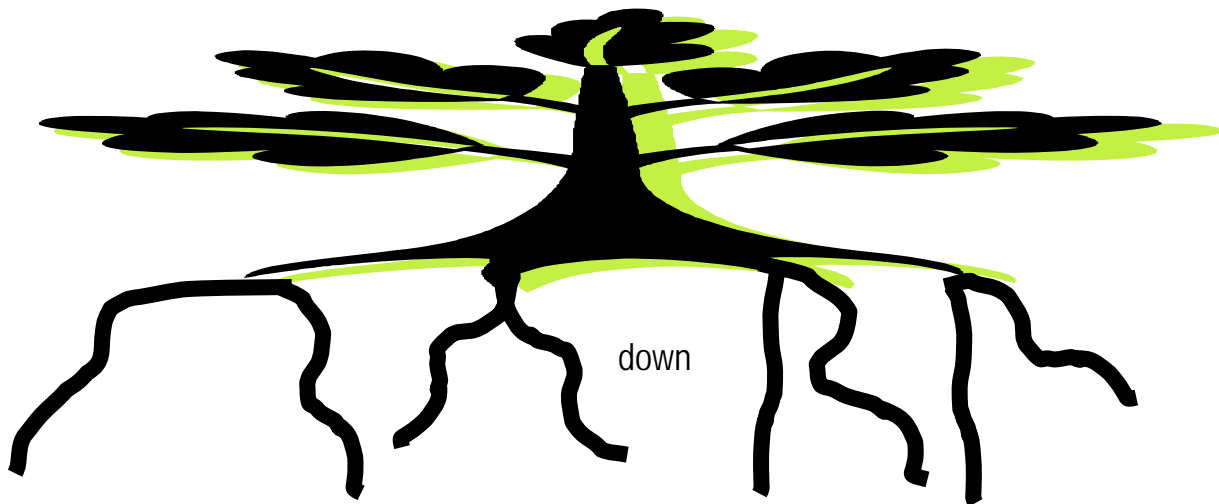
1. The evaluation discourse
The first branch is the evaluation discourse, that is, the ongoing discussion in Europe on evaluation. We refer mainly to discussions in academic circles, but of course relevant discussions can also take place in other parts of society. The questions discussed are for example the following ones.

What should be evaluated? Policies, projects, programs, organizations, management systems, persons, products? Which design should be used? Which content should an evaluation have, that is which definitions, criteria and models should be used under which circumstances? Which are the general advantages and disadvantages of, for example, evidence based evaluation or case studies? How should the effect problem be dealt with? How should data be collected? Who should carry out evaluations?

How could and should evaluation results be utilized? As an integrated part of results based management, client-oriented management or process-oriented management? Does evaluation facilitate organizational change? Can it be used for policy integration in networks?

And, of course, what is evaluation? For what values do you take side when choosing for example certain methods, criteria and objects for your evaluation?

We haven't been able to point out any major differences between the sprigs on this branch and the corresponding ones on the North American evaluation tree. The discussion on evaluation seems to be mainly an international one. In some aspects the discussion even seems to be dominated by North American ideas. The Europeans have in this sense rather been importers exporters of ideas.

We did, however, notice some evaluation discourse sprigs, which have perhaps grown larger in Europe than in the USA and which are of a genuinely European origin. For example, is it also your impression that realistic evaluation has grown stronger in Europe than in the USA? You are welcome to add to this not very long list. The remaining two branches seem to contain a greater number of genuinely European features of evaluation.

So far Europe has only been discussed as a number of states. But the EU should of course not be neglected. One important characteristic of the implementation of EU policies is that it shall be implemented in partnerships. That is, networks undertake the implementation. Thereby EU projects connect private enterprises and intergovernmental structures. Is it also your impression that more has been written in Europe, than in the USA, on network implementation and network evaluation of public policy? If so, could the EU demands on partnership implementation of EU policies explain such a difference? That is, do the demands on network implementation create a need for literature on network implementation and network evaluation?

2. Formal institutions for evaluation

The second branch concerns the formal institutions in which evaluation is being carried out. While the European states are characterised by public institutions for evaluation public bodies in the USA tend to out-source evaluations to private enterprises. Even though this trend towards involving private business in evaluation is presently growing stronger, which not the least has to do with North American evaluation enterprises presently becoming international, we still believe it to be stronger in the USA than in Europe.

In searching the causes for this major difference between the continents the marked contrast between North American and European views on the role of the state in society seems to be of great importance; while American private enterprises are supposed to critically examine public sector activities European evaluations are supposed to help to improve (or perhaps legitimate?) such projects.

However, we also want to take this theme further. When evaluations are undertaken by public institutions, is then the evaluating institution the same as the one, which has undertaken the evaluated project/program etc.? And when evaluations are not undertaken by public institutions, what influence has the institution, which implemented the evaluated project/program etc., got on the evaluation? Does it finance the evaluation? Does it order the evaluation? Are there differences between Europe and the USA in this respect?

A related question is whether it could be generally stated that European evaluations of public sector activities are more centralised than corresponding North American ones?

It is also your impression that European researchers to a greater extent than North American ones get involved in high-level public evaluations of public sector activities? Perhaps in particular in Scandinavia, this has been a marked feature of the organizing of evaluations, but it is our impression that it could be generally stated that European researchers are getting more involved with this kind of evaluations. Feel free to disagree.

The EU, again, should be discussed as a very special phenomenon as it is neither a federation nor just an international organization. How should the EU formal institutions for evaluation be described? How do they differ from federal institutions for evaluation in the USA?

3. Evaluation practises

The third branch consists of the different evaluation practises — how we carry out evaluations in practice. Note that this is not synonymous with the evaluation discourse, which may not at all correspond with how evaluations are being carried out in practice. The most basic question in this field is perhaps to what extent evaluations are being launched in different policy sectors in different countries.

A question that's been discussed in this context is if European evaluation projects tend to be of a greater scale than the ones of the USA. It is our impression that major evaluation programs in the United States are mainly associated with the Johnson years while they in Europe have been more common in recent years. Is this also your impression? Which are the major differences between North American and European evaluation systems?

The scale of the EU evaluation systems deserves attention in this context. For example evaluations of the Leader project or projects financed by the Structural Funds are immense undertakings, as "lessons learnt" shall be reported from smaller regions to national centres and from there to Brussels.

A related question is, whether the fact, that many different stakeholders take part in the network implementation of EU policies, creates demands on network evaluation, stakeholder evaluation and participatory evaluation. In this context we define network evaluation as evaluation based on criteria of the actors which have taken part in the implementation of the evaluated network implemented program. And are comparable network implemented projects undertaken in the USA?

A closely related question is whether EU instructions on evaluation of EU policy create a more homogenous European evaluation practice than the one which has been at hand before? And if so, how does this practice differ from the evaluation practices of the federal level of the USA? Nicoletta Stame has stated that the EU evaluation systems tend to focus more on evaluation of how much money is spent on different projects than the effectiveness of its use. ("Evaluation and the policy context: the European experience", article presented at the 2003 Conference of the Australasian Evaluation Society, p. 6.)

Which other similarities and, perhaps more interestingly, differences are to be pointed out between EU evaluation practices and US multi governmental evaluation?

It could also be discussed whether international organisations for foreign aid influence the evaluation practices of Europe, and if this influence somehow differs from how North American evaluation practices are influenced.

The ideological differences mentioned above concerning the role of the state in society may also help to explain major differences in the practical use of evaluation results. When North American and European evaluators detect implementation failures they tend to make different recommendations. While the European evaluator often recommends strengthening the public institution responsible for the evaluated project, the American evaluator would perhaps rather consider the results of his or her study a reason to recommend a decrease in the size of the public sector. A related question is whether the use of ex-ante evaluation and formative evaluation compared to the use of ex-post evaluation is greater in states where the legitimacy of public interventions is low, than in states where such legitimacy is high? Does the legitimacy of public interventions affect the extent to which meta-evaluation is used?

In that ideological context it is also interesting to discuss whether European and North American evaluators as a result of their often diverging views on the state tend to ask different questions. Is it perhaps so that American Evaluators to a greater extent than European ones tend to question the need for the evaluated activities?

We do also believe, that the greater use of quantitative methods in the Anglo-Saxon world (especially in the USA) than in Europe can be explained with a greater need in those countries to scientifically "prove" the effectiveness and efficiency of planned public undertakings.

Last but not the least it should be discussed for whom we evaluate. In what respects is the chosen evaluation design adjusted to the needs of the institution that's supposed to use the evaluation results?

Three versions of the European evaluation tree.
The structure presented above covers a wider perspective on evaluation than Alkin's version, but it doesn't mirror the great plurality concerning institutions and practices within Europe.

It is our belief that differences within Europe concerning formal institutions for evaluation and evaluation practices could be explained by considering the differences in how welfare services are distributed in different European countries and how this should effect evaluation. We believe that Gösta Esping-Andersen's description of three different capitalistic welfare state models can be a tool of great help in hypothesizing such differences.

Gösta Esping-Andersen has been one of the contributors to debates on the changing nature of welfare states in Western societies. In *Three Worlds of Welfare Capitalism* (1990) he launches an analysis of the character and role of welfare states in the functioning of contemporary Western societies. Esping-Andersen distinguishes three major types of welfare states, connecting these with variations in the historical development of different Western countries. The three types presented are the liberal, conservative, and social democratic welfare states. In a somewhat simplified form these types can be described as depending respectively on markets, families/churches and the state in the distribution of welfare services. As his classification of European countries has been criticised we want to leave this question open for discussion.

Now, the question is how the choice of a welfare state model should influence the organisation and practices of evaluation. Could it be so that state centred welfare societies tend to create institutions for evaluation within the state, while market centred welfare states prefer outsourcing evaluations to private enterprises and the family/church centred ones use a mix of the two organisation models? The question raised above about which influence the public institutions, that implemented the evaluated program/project etc., have on evaluations of their work should be elaborated; who undertake, order and finance evaluations in the different kinds of welfare states? Do differences in this respect influence which questions are asked? And do the evaluation practises differ? That is, for example, are different questions asked and different evaluation models used? For example, it doesn't seem unlikely, that in state centred welfare states questions as to how public welfare distribution can be bettered are stressed, while evaluations of public welfare distribution programs in market centred welfare states rather tend to question the very being of such programs? And do evaluations in family/church centred welfare states rather ask for under what circumstances family or church based alternatives are allowed to compete with state welfare services in for example child care?

A fact, which perhaps can be related to Esping-Andersen's typology, is that the Scandinavian countries relatively early began undertaking evaluation of policy instruments in the field of educational policy. Could it be stated that the incitements to make the use of such instruments more effective are as strongest in state centred welfare states? Quasi experiments were also

undertaken relatively early and have come to great use in Sweden. This can surely be understood in relation to the state centred welfare model; social engineering on the state level was, at least initially, a prime characteristic of that model. Stakeholder models are of very frequent use in Scandinavia and perhaps can also this fact be related to the state centred model.

The question if new public management has influenced evaluation practices and institutions more in market based welfare states, than in the other ones, should also be asked. After all, this tradition has got much of its origin in Great Britain; the leading European market based the welfare state. On this point your special knowledge of your respective home country will be most valuable.

The roots of the European evaluation trees.
In explaining regional differences between the European evaluation traditions we think that a wider societal and historical context than the discussion between scientists must be considered. We see two main ways to approach this task. They are presented below.

It should be noted, that when suggesting mapping the "roots" or explanatory variables of specific marks of evaluation in specific regions, we don't, in a strict sense, mean causally explaining these characteristics. It should rather be understood as an intention of coming up with factors, which make regional differences seem understandable.

It should also not be overseen that our three branches are not formed independently from each other. A sprig on one branch could be a root of a sprig on another branch. It's of course so that the evaluation discourse effects the evaluation practices, that the combination of formal institutions for evaluation effect which problems are discussed among academics and so on. Below we will not discuss such interdependencies, but focus on factors external to evaluation traditions.

The long roots.
The first way we see to map the roots of the European evaluation traditions is to tell the long history of the foundations of evaluation. Differences in origin explain differences in contemporary evaluation traditions.

Perhaps Stein Rokkan's approach in describing the political formation and nation-building processes in Europe could be of help in taking on this task. His starting points are some critical junctures of the European history. From them he explains structures of the European party systems. Concerning the critical junctures his main concern was with the reformation, the French revolution and the industrial revolution. In Rokkan's view these processes had generated contrasting systems of power and opposition, which were "frozen" into some specific party alternatives when electoral democracy was introduced.

Now, the question is which events of the European history should be considerared as especially important for the development of evaluation traditions? For example, in Great Britain and Germany the state played different roles in the industrialization process. Is there a reason to believe that this difference is today mirrored in differences as to how industrial policies are evaluated?

Another possible approach in explaining differences in formal institutions for evaluation or in evaluation practices would be to describe the history of ideas rather than the history of policies. Harald Swedner has made an attempt to roughly structure the European history of ideas since the 16th century. The academics he finds to be the most important ones are presented chronologically and divided in traditions. (Harald Swedner, 1983, *Socialt arbete, En tankeram*. Lund; Liber Förlag, p. 161.)

The short roots.
The second way we see to map the roots of the European evaluation traditions is to describe the political and societal contexts of our own time.

As stated above, we do not mean, that Esping-Andersen's typology is the only tool, which could be of use in hypothesising demarcation lines between different European evaluation traditions. We do however think that the role of the state in society today should be elaborated in explaining regional differences between evaluation traditions.

The constitutional foundations of the state should also be discussed. Is it so, that evaluations are organized and launched differently in federal and unitary states?

Do the routines of political decision making influence how evaluations are undertaken? Are there differences between on the one hand countries in which several stakeholders are included in the process of political decision making (for example consensual democracies) and on the other countries in which the political majority has got greater possibilities to influence the decisions? Is the stake-holder-model for evaluation more frequently used in one of the groups than in the other? Are different questions asked in the two groups?

Do the routines of policy implementation matter for evaluation? If private sector stakeholders take part in the implementation, do they then normally take part in the evaluations?

Does a successful application of a specific evaluation method influence how widely it will be used in that country or region in the future, no matter if it's appropriateness in other sectors could be questioned or not? That coincidentally different methods were introduced early in different countries could explain national and regional differences in evaluation practices today. Frans Leeuw has shown that in some countries specific evaluation methods have won such a degree of status that their appropriate application to evaluations of education is no longer questioned (Frans Leeuw "Trends, Topics and Theories" in *Educational Evaluation around the World, An international Anthology*. Copenhagen; The Danish Evaluation Institute, p. 22.).

These are only some suggestions for questions, which could be considered in mapping the roots of different versions of the European evaluation tree. Feel free to come up with other suggestions.

Questions for further elaboration.
The following questions concern the main themes of this paper. We would like to discuss them with you at the June meeting in Stockholm.
1. Is there a typically European evaluation tradition? Are the differences within Europe so pronounced that a generalizing European evaluation tree would also mirror the Evaluation tradition of the United States?

2. If Alkin's tree structure is used, which European names would you like to place on the three branches?

3. If you compare the formal institutions for evaluation and the evaluation practices in the EU with such institutions and practices in your respective home countries, which differences do you see?

4. Do you agree with us on that an evaluation tradition is bound to it's historical and societal context?

Attachment 2.7: Minutes of symposium June 17-19 2006.

2006-09-21 (Revised 2007-01-27.)

<div align="center">

Mälaradalen Evaluation Academy
www.mdh.se/isb/mea

Minutes of
"Evaluation roots in the USA and in Europe: tracing traditions?"
Stockholm, June 17-19, 2006

</div>

Gustav Jakob Petersson
gustav.petersson@mdh.se

1. Introduction.

This is a résumé of the Stockholm symposium of June 18, held at the Swedish National Audit Office, Stockholm. It consists of two main parts.

The first part is the program for the panel discussion "Evaluation roots in the USA and in Europe: tracing traditions?" at the UKES & EES joint conference in London, 4-6 October. This program was decided on in the afternoon of the session of June 18.

The second part is a reconstruction of the presentations held between 9 and 12.30 am. The sources used were mainly the recordings made at the session and the papers that Nicoletta Stame and Reinhard Stockmann most kindly delivered and in which the contents of their respective presentations are recapitulated. An exception is the presentation by Marvin Alkin, which is a transcription of a recording. Some comments that were made on June 18 on the presentations have been included below. The ones that were made in connection to a presentation are also presented in that context here. The afternoon comments are presented separately. A few additional comments have been made by the author of the these minutes in order relate the presentations to each other.

The presentation by Ove Karlsson Vestman has not been included, because it mainly concerned organizational details on the Stockholm symposium and because the rest of it was a brief presentation of the paper by Petersson – Karlsson.[24]

Neither has the presentation by Reinhard Stockmann been included. The reason for this is that he presented a paper consisting of a number of most thoroughly discussed critique points on the paper by Petersson – Karlsson. This paper, written by Stockmann and Wolfgang Meyer.[25] To also have included Stockmann's presentation of June 18 would have been presenting the same material twice. The paper also contains several suggestions on how a research project on the European evaluation discourse could be outlined.

---

[24] Here presented as Attachment 2.1.
[25] Here presented as Attachment 2.3.

Present on the session of June 18:
Ove Karlsson Vestman, chair person
Frans Leeuw
Nicoletta Stame
Reinhard Stockmann
Marvin Alkin
Jan-Eric Furubo
Evert Vedung
Gustav Jakob Petersson

2. Program for "Evaluation roots in the USA and in Europe: tracing traditions?".

The program below is for the panel session "Evaluation roots in the USA and in Europe: tracing traditions?", which will be held on October 5 2006 at the EES conference in London. This program was outlined in Stockholm in the afternoon of June 18. The initials tell who is responsible for each presentation.

Presentations:
1. Introduction		3 min.
2. The American theorists tree	M.A.	10 min.
3. Why theorists are less appropriate for Europe	N.S + F.L.	15 min.
4. An alternative tree for Europe	O.K. + G.P.	10 min.
5. a. A sectoral perspective	R.S.	15 min.
b. Systems	F.L. + J.E.F.	10 min.
6. Discussion, opinions from the audience		25 min.

3. Presentations held on June 18.

The presentations are here presented in the same order as they were held on June 18. The recordings of the relevant parts of the session are to be found on two tracks, here referred to as 1a, 2a and 2b. The times presented in relation to the presentations mark the starting position of the citation on the respective track.

3.1 Jan-Eric Furubo.

Furubo made a presentation of some parts of the work that was made by himself, Ray Rist and Rolf Sandahl on the *Evaluation Atlas*. This was meant as inspiration for the London discussion as there are obvious parallels between that project and the project (perhaps) starting now. First, also the Atlas-project was an attempt to map evaluation cultures. Second, attempts were made to explain the observed differences. (2a 24 min)

There were two purposes of doing the atlas. The first was to describe the evaluation situation globally, though with a certain stress on Europe. The second purpose was to explain the observed pattern. (2a 17 min)

In the descriptive part of the Evaluation Atlas 21 countries were studied by using nine different variables. Each country was for each variable given a score between 0 and 2. 2 indicated a high "evaluation maturity" and 0 a low one. In this way the "evaluation culture" of each country was described. (Furubo – Rist – Sandahl 2002, p. 14- 16) The nine variables were (Furubo – Rist – Sandahl 2002, p. 14- 16):
1. Evaluation takes place in many policy domains
2. Supply of domestic evaluators in different disciplines
3. National discourse concerning evaluation
4. Professional organizations
5. Degree of institutionalization – Government
6. Degree of institutionalization – Parliament
7. Pluralism of institutions or evaluators performing evaluations within each policy domain
8. Evaluation within the Supreme Audit Institution
9. Proportion of outcome evaluations in relation output and process evaluations

The pattern observed to some extent supported results that Hans-Ulrich Derlien had got 15 years earlier and which he then described with the waves-metaphor. However, the new results pointed not to voluntary diffusion, but rather to forced "pushes", a halted diffusion to a group of countries in the mid-80s. Then it stopped. After that the spreading of an evaluation culture to yet more countries seems to have been dependent on external pressure from organizations like the OECD or the WB. (2a 19min)

In the Atlas-project also the predisposition for an evaluation culture was studied. For example it was studied whether statistical methods were used in the public administration. This showed for example that countries with a legalistic tradition adopted evaluation relatively late. It was also shown that it seems to be relatively likely that welfare states shall adopt evaluation. This was explained partly with that there are more state activities to evaluate in a welfare state than in other ones, but it was also explained with that the rationalistic belief is a basic fundament of the welfare state. This belief in the human ability to observe causal relationships is a necessity for evaluation. (2a 22min)

<u>3.2 Marvin Alkin</u>.

[*The notes of this presentation only are a transcript based on a recording. The paper and the picture referred to are to be found in Attachment 2.1.*]

Ok…I perhaps will be repeating what you already know because I am really going to briefly talk about the tree, which I am best able to have comments on as I am less familiar with the European situation than all of you of course. It was never my intent that you would try to put Europeans on the tree that I had structed. I view the tree as something that I had put together based on my perspective. If it were valuable to you to use the same branches that would be fine. If you wanted to (conceive) perceive them in a different way that would of course also be fine. Let me comment on the background of the book first.

First of all the….I just…. before I left I saw a review of the book in the American Journal of Evaluation by Jane Davidson, which was I think pretty favourable. I e-mailed her before I left and commented that while I was going to incriple over a few of the points she made I thought that generally the comments were correct. I point out that the title of the book is Evaluation Roots, Tracing *Theorists* Roots and Influence. That is we are talking about Theorists and that theories basically are housed in the way that individuals, theorists, think and so the criticism is not of … well you know … this isn't really theories, its theorists well, that's part of the title.

The theorists of course are talking about prescriptive theory not descriptive theory…. prescriptive theory namely an individual is prescribing the way in which they believe the evaluation should be conducted. It's not descriptive in the sense that we have done an empirical study about the way in which evaluation should be conducted. Then theorists we felt the need to talk about who is a theorist or who we would identify for being eligible for being on the tree and we said an individual with a particular evaluation point of view or approach that is identifiable…people would say oh yes….that's an evaluation like Scriven or that's a Stake way of doing it and we then exempted individuals who are primarily methodologists there might be methodologists involved in evaluation which is important evaluation methodoly but not theorists. We identified people who were evaluation issue analysts that is who wrote about specific topics one at a time but not a total evaluation perspective. It's interesting, what comes to mind there is Ernie House…early Ernie House ….because at the time I was director of the center I had Ernie visit the center for a term and that's when he wrote this book The Logic Of Evaluative Argument which was a setup of study evaluation monograph which basically at that time Ernie had not really come to thinking in a holistic term, a holistic way about evaluation, he was dealing issue by issue and that book pointed it out really well because that book was a book where …

I was stimulated by that book as a matter of fact to do some other things because he had been heavily referenced to Belgian philosophers Pearlman and Obrexta (?) and so I think I may have even mentioned this in the roots book and what I did was … I was so influenced by that, that I normally have my students in my comparative evaluation theory course do a lot of simulations where they role play theorists and have to get into the mind set of the theorists and not just regurgitate "he said he said", they have got to be the theorists and I asked tough questions when I interviewed them. But what I did was I had them write a paper on why we went through these Belgian philosophers philosophy book and I had them write a paper on my view on evaluation having read Obrex and Pearlman by Robert Stake by Michael Skriven by whoever so that…but anyhow…

I go off on a tangible which I am very prone to do So, that finally not those who are primarily practioners of evaluation of writers about evaluation so someone might you know be a

practioner and do a lot of evaluation or even write very standard textbook on evaluation….standard textbook but which doesn't really doesn't put forth a particular theoretical approach. These are the exemptions that I engaged in here. Then, one of the comments I do want to make. This was the early tree that you found in the second to last chapter of the book. So, this was in there and you know I've been criticized and rightly so for the fact that it's too American and that there are a lot of others who could have been included and that's probably correct but then again in putting together a book you have to go with the people you know and these are the people I know and these are the people who I knew I should get to write chapters because all of them are friends, all the them are people that I know very well so that's part of it. Also there are probably too many people on here from education and that is probably a correct criticism but again I think that falls within the same pattern.

I was really quite amazed at the fact that I was able to get the participation of all these people. The one person I didn't get participation from who had agreed to write was someone that I didn't know personally, Barry Mcdonald, but Stake helped to arrange his participation and he ended up agreeing but not producing. So it validated my original conception that mainly you go with the people you know because you can put pressure on them to get it done. So…that anyhow … that's the case.

So the notion then is that all evaluators are concerned about how their evaluation will be used, all evaluators are concerned about the kinds of methods that they will use and all evaluators recognize that there is some kind of valuing process involved. Originally I had written about that many years ago in the International Encyclopedia of Educational Research and Practice … Torsten Husein and Neville Posilltwaight were editors of so this conception goes back a long ways well not the tree but the conception and then subsequently Ernie House wrote with me in the evaluation of the four volume encyclopedia of educational research, which I was executive editor of, where we expanded the notion of these three concepts and subsequently the notion of trying to get the categories to think of them as more than just categories of people but relationships within those categories. Relationships that evolve, relationships where there were connections between people that seemed to evolve.

So, for example, this Use branch and I'm sorry if I am just repeating the stuff that you already know but its' what I know. This youths branch, the origin of that branch was really about evaluation and decision making, the purpose of evaluation was to serve the decision making and we saw that in Daniel Stufflebeam's work very nicely and we also saw that in Joe Wholy's work but in a different perspective. Wholy came out of … Dan was out of education, actually a psychologist, Stufflebeam was a policy guy who filtered in and out of government depending upon whether or not it was democrats in charge or republicans in charge…he was on the democrat side of course. So his perspective was a federal government perspective of how can we perform evaluations within a government setting. The notion of evaluatability assessment, that there are some situations where it's really not…

[Leeuw: That's Wholy?]

Yeah…Wholy

That evolved into Michael Patten saying; look you know, it's not just the decision maker, if we really want to get news to take place it's got to be someone who is in a position to use the information and who cares about using so that the notion of primary intended user. I'm not

going to go through each of these branches but I just wanted to show you the way in which a branch evolves. So the notion of a primary intended user, Patten and then from that Cousins, Prestel, King, folks who said; look it's not just the primary intended user, but someone using it, there's the notion of wanting to have an organization involved and having organizational learning that goes beyond just the user. So there is a relationship on a branch between people. That's evolutionary and we see the same thing in the Methods branch. I originally had Tyler, Ralph Tyler, at the beginning of the branch because that's worth it, in education at least, to date it back into the 1930's which was really very important work and really to my mind was some of the first discussion of evaluation as evaluation in the field but then sort of got lost for 20 or 30 years to be revitalized in the 1960's. Campbell of course is the primary inspiration for the whole branch. The work of Campbell and so what we have then is Cook and Ed Sushmin, a sociologist who really just published that one book Evaluative Research which was a really a very good book but really never did very much more with evaluation other than that. But he was put on the tree primarily because of his influence on Campbell, getting Campbell to understand that he was not only just doing research, talking about research, but that he was he was really doing evaluation so I put Sushmin on for that reason. Barouch, Rossi comes out of that stream. Krombach and Weis come off of that stream but it's interesting because Krombach coming off the Campbell stream is primarily Krombach heading views counter to Campbell but inspired by the debate with Campell. And Carol Weiss, who one might think of as on the Use branch, but I didn't because I think that she is out of that stream, but I reflect that she is Use oriented by the way in which her branch curves. I know that's … by the way which her branch curves somewhat in that direction … laughs … so we did a lot with this curving business … laughs …

Now, the valuing branch of course Scriven is the inspiration for that. That is the essence of evaluation in Scriven's mind; an evaluator sends values, he says this is good, this is bad which Scriven easily does. Stake of course is very prominent in this branch. Eisner; many of you probably don't know very much about Eisner because he's in education but he's basically definitely a valuing kind of guy … so he … we'll come back, we don't need to talk about it … House we had off in that direction, Wolf and Owen which was like a one shot deal… They wrote on adversary evaluation … trial by jury that, you know, you would come up with … what you do is you come up with a protein and a contein the program is good, program is lousy they each gather their data and they actually have a trial setting. So it is a valuing thing they did a one shot … well they were different. Wolf wrote and Owens wrote so there were two separate theorists here. And there was one big example of that evaluation which was a Hawaii project where they actually did an evaluation for the state of Hawaii in that format. I never found it a terribly feasible thing to do…

[Karlsson:] I noticed that in television they have used that concept many times. I saw I think Ernie had a program and they tried to have these pro and con juries presenting their opinions. I mean. the one who finally made the judgement is the audience and we have that latest in Sweden when it comes to the voting about the curteousy for the Euro…that the parties was invited to have a panel for and against and they were inviting speakers that were having that kind of "I'm for it" and then they were challenged by the other side and this jury also was the audience. But the idea of again…but the solution of how you collect all the information is left forward to the audience who looked at the program. But the idea of how to turn for and against is still working. At least in television format it works very well.

[Alkin:] So what I decided to do was to I am continually making changes and I talk in the book about the process of change, the theorists change and the whole. I've done a lot of some

research and some investigation about the kinds of things that influenced theorists to change their views over time; contacts with other people who theorists placement on a category system. Nico had it yesterday, someone reacts and said you put me there…what is there in my writings that would make him think that I should be in that position. And so these kinds of reflections get people to think about their own theories and change them. So we, I mean Tina Christine and I, Tina was a former doctoral student of mine now a professor at Claremont Graduate University. So Tina and I are continually thinking about these things. We just love getting together and talking theory which sounds boring perhaps to lots of people but we get quite a kick out of it. So we've done some restructuring on this tree….the Methods branch we made a few changes. We moved Tyler off the sub branch recognizing that while he is recognizing that while he is influential in the early stages he really wasn't really influential in terms of influencing the rest of that whole branch. So we moved him off to a little…a little sub branch. Now Campbell of course continues to be the heart and soul of this branch, both in the way that he directly impacted other theorists with his work on quasi experiments and validity and as a basis for counter point to Krombach as I mentioned earlier. I deleted Sushmin from the branch even though he was influential because it was really just basically that one influence on Campbell to simplify the branch. Rossi continues to be shown on the main portion of the branch both as an influence on Carol Weiss and certainly on Huey Chen. And there were a significant change as we sort of moved Cook off to the side…and as on the main branch….and I sort of moved Bob Barousch to a sub branch, because Bob is really a little different from some of the other people on this branch in that his regard or rather lack of regard for quasi experiments. Bob is totally committed to the necessity for experimental design and in turn has been one of those who's heavily influenced the US government in that direction, unfortunately, expressing my own mind. Earlier we had mentioned in the book Gary Henry Melmark and George Gilmiss and I had said theories included are able to be classified into a single branch. These authors view social betterment as the ultimate objective of evaluation and present a point of view grounded in what they refer to as Common Sense Realist Philosophy. Their very diversified nature of this perspective while a great strength precludes its inclusion on the tree. We admit this our conceptual analysis was incomplete. We relied too much on the mark et al 2000 book and not enough on their realist evaluation monograph, which we then became persuaded, led them to have a sufficiently unique position to be represented on the Methods branch. This is a strictly American…we've deleted any of the McDonald, John Owen, and said ok let's just stick with strictly an American tree here. They basically comment that mechanisms…they comment that emergent realist evaluation is an evaluation methodology that gives priority to the study of generated mechanisms, where mechanisms are defined as the underlying causes of the changes that occur which puts them very cleary cleary on this branch.

On the Use branch we had very very few changes. Basically we deleted Malcolm Provis who had been on this other chart. Well, you'd remember it. Who had been on this other chart and who we felt his work was no longer timely and really hadn't influenced anyone. He's the guy who wrote Discrepancy Evaluation and he was a director of research and evaluation in the Pittsburgh public schools and it was for a while a kind of influential book but its really faded from existence. So that's basically that branch.

Now the branch where we've really made major changes is the Valuing branch. It's obvious that the work that the Valuing y branch was greatly influenced by the work of Skriven after all he proclaims that evaluation is not evaluation without evaluing. In his words evaluation is the science of evaluing. Ah…but the branch was always a bit of a problem because in addition to people like Skriven it included those interested in social justice those who had spoused the

philosophy of subjectivity that there is no objective reality. So, it…. the branch then….while everyone was interested in valuing, there were so many different ways in which that concern for evaluing took place. So what we have done is there were….ah….we looked at the theoretical perspective of the valuing branch within the context of philosophy science and felt that this offered a kind of revised understanding of how the perspectives could be understood. We have the basic axioms of post positivists and constructivist paradines ….I am so dry today….Um, and we basically then diverted and talk about a post positive influence and a constructivist influence as a branching off on that valuing branch. As a way to clarify the differences that we are not very apparent I think in the tree that we present in the book. So the post positivists all observation is fallible and has error but the goal of science is to attempt the measure truth although that goal cannot be fully obtained. Constructivism recognizes no single tangible reality, multiple constructive realities. So, Skriven's realist thinking about valuing is reflective of both the ontology and epistemology of the post positivist paradigm. It offers what he believes to be comparatively unbiased method pertaining truth about an honest worth and then advocates for the evaluator to make a valued judgement after gathering what he believes to be the most credible evidence. He does not reject the idea of using experiments to determine cause-laity, although he doesn't do it very often, ah, but argues that there should be more than one method for determining cause-laity. Stake's work respects Skriven's thinking, although he argues for using thick description. Thick description to access the programs worth primary by a case study method. So the use of a case study method introduces that value is bound to context and that context must be considered in determining value. Stake however leaves the valuing to the the evaluator. Ah, he does not give up the role of valuing. We saw….I mean Stake is very quite controlled in that respect. I had an article with another of my students on stakeholders and the way in which different theorists perceive the role of stakeholders. And Stake very firmly while he believes in Stake holders does not give up a lot of the evaluative decision making to Stake holders he retains it himself. So Stakes inclusion of an emphasis on case study methods and analysis to him prompted a shift from realist to relativists inquiry models none the less Stake sits on this bridge between the split of the two empathies but on the side of the post positivists. House's work brought significant attention to the inclusion of the values of the under represented to the evaluation process. While his work is grounded in constructivist thinking, methodologically his work breaks from the basic acts against, from of the constructivist paradine. He is willing and likely to use quanitated methods when conducting evaluations. Thus he too…meaning this House here sits on the bridge. What we've done….and I'll come back to that. Let's look at this left branch here, Eisner to pick as an offshoot of the Skriven's branch because he obviously points to the evaluator as the expert that determines the final value of the program although in a vastly different way then Skriven does. And Wolf and Owens are like wise placed on that branch. Now returning to this other branch here, now what we have is Guba & Lincoln, they've been moved to the base of the constructivist influence arm to reflect the epistemology of their perspective which closely adheres to the basic tenants of the constructivist paradigm. Their theoretical model serves as the foundation…..I'm sorry, I have some handouts of this on which I'll give you….I should have done it beforehand…I saw someone taking notes….sorry about that…their model serves as the foundation for a string of models influenced by constructivist thinking. Now, House and Howe as different from earlier House alone, the liberative democratic model has been included as its own model separate from the social justice work of House which is already represented on the tree. Deliberative democratic is based on the ideals of House's earlier work on social justice but has been influenced by the prophecies put forth in Guba & Lincoln, fourth generation evaluation….ah House and Howe believe that evaluator should accept authority but not power. An evaluator…ah…. so anyhow. An evaluator who caters to those with power perpetuates

inequality which is why inclusion is the first criteria of deliberative democratic evaluation, dialogue is the next criteria and deliberations is the final criteria. So House and Howe's approach is value engaged, fact and value are not mutually exclusive they exist on a continuum where a middle ground exists between root fact and bare value….um…that's quotes from them. I'm adding Jennifer Green also. I am quite impressed with her work….has also been added to the tree, directly above House and Howe. Green's value engaged approach is rooted in House's work, clearly it uses deliberative democratic evaluation, however it differs in that she strongly stresses stakeholder involvement to a…..which House does not. In a manner which closely resembles that of participatory evaluation approaches up in the upper left corner. She also explicitly emphasizes the use of mixed method designs and field work within her evaluation. So she's up there and if there were some way for me to bend that over to the far left which there isn't, I would, that's where a screen might come in handy. Donna Mertin's inclusive approach, ah, could be but I think Jennifer Green definitely belongings in this branch but I …she really…I would love to just sort of arch her over to the Use branch as well. Donna Mertin's inclusive approach is also a direct descendant of Guba & Lincoln but it's unique in its emphasis on diversity and inclusion of diverse groups. I've got more to say about her but I think we'll just let it go at that. I have some other comments about how monumental the task is to try to expand this beyond an American perspective and particularly to get into all of the work that you have proposed on institutions and practice and I think I will reserve those comments written down for further later discussion cause I've already used too much time.

Let me just give….maybe to show….you can have a….

Let me get over here and pull some papers out for just a…..

Sorry about that….

[Karlsson] What I'm going to say now maybe could be misunderstood like to put it myself is a very important rule but let me just give you an idea of what it could be like when you try to understand Ernie House. How House changed…I spent almost a year with House in 1990 and then I came to Colorado University after been working with an stakeholder model in a rather large project for about 5…4 or 5 years you see…..

What I did I think I introduced to Ernie the practical complications of working with these ideas of social justice when it was put into reality working together with stakeholders and how you balance the different views from different stakeholders. That was my problem in my dissertation….that I tried to see if there is a way of have a general statement of values or if it was just left to each of the stakeholder groups and it was like a market if one could find any kind of general values on what basis could you argue for those general values. That was what I had thought Ernie should, could give me some answers because I have read him and I thought he had gone into John Rawls's new views and theories like that. We came out with a model that should be more of a dialogue and I wrote an article together with Tomas Schwandt and some other writers….a number of Evaluation….and then Ernie and Howe who were partners in that office decided to write a book together and Ken Howe is a philosopher and liked to have his chapter challenging Skriven especially and Ernie was writing his idea about his dialogue version. And then I visited them and I went through the book and give some criticism of the book because I thought it was two books actually; Ken Howe´s and Ernie House's book. So they made it more of a, maybe they succeeded to, I think still when I read it I recognize very well two different debates. So Ernie moved suddenly because when I met

him first he was very much against Gubin-Lincoln this constructivist theory but then he was more or less convinced that he was not a positivist but…so I think still you have to ask Ernie where he stands here when it comes to these two branches if he really is the old Ernie or if he is together with Ken Howe and maybe influenced by my work that you can put him of there. So that is just one example of how difficult it is to…..

[Alkin] You see…that's why we have that first House, the first House is not on that branch just one sort of in that middle location…I know these conception things are kind of silly but I get a kick out of it…

[Karlsson] I think it's a clever observation that Ernie has two identities here but I just want to express...I think from my position...I think one of the reasons why is that he has gone in company with some other guys here….

[Alkin] Yeah….

[Karlsson] And then maybe then he has moved in his writing….and I don't know exactly if he changed in his writing…..

Laughs…

[Alkin] Well you know your comment is really ah very important in terms of the notion of discourse that you mention in your paper because the whole business of collegial discourse and the way in which it changes theorists views was pointed out very nicely by your presence in that situation…

[Karlsson] Yeah…it's an exchange very much on exchange over this…people come to the United States and the opposite. I mean Ernie has been in Sweden many times and in Denmark, so he has been influenced by the ideas I think from researchers…maybe more interested in the welfare states, and a view of the welfare states than you meet traditionally in the American concept.

[Alkin] Yeah…

3.3 Nicoletta Stame.
Nicoletta Stame delivered some thoughts on what a European discourse tree, comparable to Alkins American one, would look like.
"With Evaluation Roots, Marvin Alkin has drawn a wonderful picture of an evaluation theorist's tree in the US, creating branches and locating authors on the branches. If we had to compare that tree to a similar European tree we would not find many differences in the labels of branches, for the simple reason that the European debate is dependent on the international debate in which US authors are prominent. We could therefore put European names on the existing branches, and in a few cases we could add some branches with respective names (e.g. in the case of realistic evaluation by Pawson and Tilley, that could be a new sub-branch in the methods (?) branch[26])." (Stame 2006, p. 1)

---

[26] Pawson and Tilley would object to being located on the methods branch, because they think of realist evaluation as belonging to the theory-driven approaches that have superseded the methodological debates of the '70s and '80s; but also other theory-driven theorists, such as Rossi, Chen and Weiss are on that same branch in Alkin's tree.

Stame mainly focused on the respective roots of American and European evaluation practices. Her reasoning was presented as a response to the paper by Petersson – Karlsson.

"In a review of Evaluation Roots E.J. Davidson (2006) recalls that Alkin had put labels also on the trunk (accountability and systematic social enquiry), and proposes to consider them roots, instead[27]. This takes us to answer Karlsson and Petersson's invitation to think about European evaluation roots[28]. My contention is that the US roots that favoured the emergence of evaluation were, first, different from the European ones, and, second, responsible for some special characteristics of the evaluation approaches that are enlisted in the branches of the US tree. And I want to inquire into the legacy of European roots on the European evaluation outlook. My reasoning will proceed by comparing the European experience to the US one." (Stame 2006, p. 1)

First the roots of the American evaluation tradition were discussed.
"According to a well established tradition, the US roots at the origin of evaluation, were:
a.   federalism that allows to consider single states as laboratories for social experiments (Manski and Garfinkel, 1992),
b.   pragmatism, that looks for truth and values in research, not as external given
c.   the invention of programs (of the Great Society, but even before in education, international aid, R&D, etc.): programs are public interventions that have to prove to be effective in a political culture that is market-oriented.
d.   the development of social science research methods.
These roots favoured a pluralistic culture of evaluation[29], open to methodological debates and to a blossoming of different schools, as are reflected in Alkin's tree branches. In the US, constructivist approaches can compete with positivist ones, theory-based approaches can contest the terrain to method oriented approaches, and the recent prescription of randomized control trials by evidence-based policies was met by a strong statement by the American Evaluation Association that RCT are not always the best recipe, to say the least ….[30]" (Stame 2006, p. 2)

Thereafter the roots of European evaluation were discussed. The assumption of the paper by Petersson – Karlsson that the structure of the welfare state influences the evaluation tradition was not used. Instead a distinction was made between the Europe of the Roman law and the countries of the common law tradition.
"Now, would the same roots as in the US be present [in Europe]?
a. federalism? In Europe, centralized states are the rule, but there are a few federalist exceptions (Germany). As for the European Union, it is a neither a single state nor a federal state, but a new organism, working alternatively as an inter-governmental system and a multi-level governance." (Stame 2006, p. 3)

To this point Stame added a most interesting point on the role of a constitutional court in some European countries. She exemplifies with Italy were attempts have been made to introduce the American kind of welfare programs only in some parts of the country. This

---

[27] To continue with the tree metaphor, Alkin has depicted the *branches* of the evaluation *tree* by way of collecting material on the *roots* of his authors' theories.

[28] They propose to think in terms of long roots (events in European history) and short roots (political and societal context of our time).

[29] Wildavsky (1979: 218) notes that phenomena like these, that are frequent in the US political culture, could not exist where a Weberian bureaucracy asks for the right way of doing things .(ctr.)

[30] Karlsson and Petersson say that in order to "prove" the effectiveness of public interventions in the US there is greater use of quantitative methods. I think that seen in the combination of all the roots outlined above, that goal could be attained by a plurality of approaches and methodologies.

however was by parts of the population not only considered unethical because it gives only some parts of the population benefits, but also unlawful. Experiments as such are thereby impeded. (2a 98 min)

b. "pragmatism? Not really! Strong ideological convictions, instead.
c. Programs? They are a recent discovery, mainly introduced by the EU. What prevailed for a long time was a culture of structural reforms and new laws that were expanding the state power (the welfare state is a European creature) to more and more fields. The political culture is generally favourable to state intervention, that is seen to redress market failures. But laws do not require to be evaluated, only to be implemented. And if the opinion changes about a law, a new law will be passed.
d. Research methods? Yes, there is a long tradition." (Stame 2006, p. 3)

The typically European roots were discussed separately from the ones common with the USA and divided into the categories "specific events" and "cultural features". This answers a question launched by Petersson – Karlsson on whether the "long roots" of evaluation should be discussed mainly in terms of intellectual development or major political-societal changes. The implicit answer seems to be: both. (Petersson – Karlsson 2006, p. 7 f) However, Stame also discusses "short roots", for example the influence of the EU today on the national evaluation practices. This corresponds to the second alternative suggested by Petersson – Karlsson. (Petersson – Karlsson 2006, p. 8) Stame´s discussion could thereby be interpreted as a combination of the different alternatives proposed by Petersson - Karlsson.

"As for specific events, one has to account for the evolution of the evaluation tradition and practice internationally. Even in the US, those roots were crucial at the beginning, but then something else occurred, that changed the landscape, and evaluation approaches. And Europe, as a latecomer, could reflect on the US experience. Here we can refer to the evaluation waves as described by Derlien (1990):
1. what we have said above refers to the first wave, that was sprung by the Great Society policies
2. then came the second wave, resulted from a disillusionment with state intervention, the reduction in public expenditure owing to the fiscal crisis of the state, and the emergence of New Public Management principles. During this wave evaluation expanded to other industrialized countries
3. then came the third wave, when evaluation was brought from outside: from such international agencies such as the EU into European states beneficiaries of territorial re-equilibrium programs; and from donor agencies (WB, UNDP, OECD, etc.) into developing countries beneficiaries of international aid.
Thanks to this chronology, we can see that while in the US evaluation moved from the first to the second wave, in Europe evaluation started in the second wave, but not everywhere alike: some states (UK, Northern states, Germany) adopted evaluation during the second wave while other ones (France, Italy, Spain, Greece) adopted evaluation as a result of the third wave. By the way, this means that whereas evaluation in US had the original imprinting of a period of optimism, in Europe, at least for the national states that had adopted it in the second wave, it was originally linked to the need for rethinking the role of state intervention that public opinion had always considered good. And it was other public institutions that did it.
    Then, let's look at European cultural features.
    First feature, the positive attitude toward the state. At the beginning this made the need for evaluation inexistent: why to assess the value of something that is already considered good? Then, when the state failures became evident, evaluation was seen as a way for

improving the state interventions, not for reducing it; at least so long as the trend toward outsourcing and privatization did not get such momentum as it enjoys today.

Second, laws not programs. Laws that are passed are there to stay, not to be experimented in a given time frame. If something goes wrong, it is generally argued that the law was good, the implementation did not follow. But in the last resort, even laws can be changed, and it can happen with no evaluation done. This way of thinking is at the origin of the many evaluation commissions that are actually supposed to inquire into the working of laws, and whose resemblance to an evaluation actor is only in the name." (Stame 2006, p. 3 f)

Stame argues that when laws play the role of programs there is generally less evaluation and exemplifies with the Italian fight for a law reform of the health system. The law was made in 1978 and it was the first big universalistic law of the welfare state of Italy. Then it was decided that it should be replaced by a more managerial system of running the health system and in 1992 another law was made. However, this second decision did not rest on evaluation; between 1978 and 1992 no evaluation of the Italian health system was made. Only with this entering of new public management did evaluations enter this sector. This was said as a contrast to the American experience of the 1960s with programs that were evaluated. (2a 98 min)

"Third, the role of public administration. This point introduces a consideration about the differences between political cultures. In Europe there is a basic distinction between countries of the Roman law tradition, and countries of the common law tradition. Only in the former (e.g. France, Italy, Spain, but also Germany) the government and public administration enjoy a special status that is regulated by administrative law. This allows for a hierarchical system that has its own rules. Nowhere this system is better articulated than in France, where the Napoleonic code still inspires a centralized state, and where the public administrator is trained to become a state functionnaire (not to serve society, as in the common law tradition of the civil service). This centralistic, Napoleonic imprinting has had a great influence on evaluation, that comes to be seen as a further inspection whether public administration works as established (legitimacy), rather than as a way of assessing programs' effectiveness in addressing social problems and producing changes. But one could also think at the Weberian bureaucracy (modelled on the German state) with its characteristics of impersonality and expertise.

All these features explain why at the beginning evaluation met a strong resistance; at the same time, they may be responsible for the way evaluation is conceived and practiced nowadays that it has indeed become an integral part of many laws and government bodies (Ministries, agencies, etc.). The words that better characterize it are rigidity, conformism, bureaucratic style. Some of the consequences of the above could be:
- evaluation guidelines are not seen as suggestions of what could be done, but as prescriptions of what should be done [31].
- Ex ante evaluation is more widespread than ex post evaluation.
- Evaluation approaches differ mainly according to sectors and policies than to evaluation questions."
(Stame 2006, p. 4 f)

---

[31] I refer to the perception by operationals about such guides, which is often warranted by the way the Guides themselves are written. There are good exceptions, most notably the "Guide to Evaluation of Socio-economic Development", issued by the DG Regio (www.evalsed.info): however, it is not as much skilfully utilized as one would have expected.

In her presentation Stame made a clear distinction between on the one hand the Europe of nations and on the other the EU. The EU evaluation tradition was dealt with as a separate unit of analysis.

"So far for the national states, mainly those of a public administration tradition. However, this sketch would not be complete if we did not address the EU. As we have seen, a great spurt to evaluation in Europe came in the third wave, from the European Commission requirements for evaluating Structural Funds programs for social cohesion and territorial re-equilibrium[32]. The EC has established its own evaluation system which has the following characteristics:

- evaluations are commissioned by the EC or by the national states according to rigid terms of reference, that ask to evaluate everything, and meticulously state the methodology that should be used (how many interviews, the report index, etc.). This has produced the strengthening of big research companies that become expert of monitoring sets of indicators established in advance, and that concentrate on assessing money spent rather than results obtained.
- there are of course different approaches in different policy arenas. The most common are logical frameworks in aid to cooperation, peer review in R&D, and everything that can be associated with goal-oriented approaches and linear causality in other fields.

    Under these conditions, EC evaluation suffers from being almost exclusively an exercise in accountability, with little space for learning. One reason for this could be found in the nature of the commissioning authority. The EC administration has been modelled on the French one, with the aggravating fact that its legitimacy is not as well established as that of the prototype, and that it always acts in order to show that it is indeed doing things right. Thus EU evaluation is perceived by beneficiaries as an obligation for getting the money, and by the EU functionnaires as a proof that they acted correctly. This leaves little space of manoeuvre for the emergent realities that appear during an evaluation, for the learning opportunity that come out of it, and for the search for suitable evaluation approaches." (Stame 2006, p. 4-6)

The question about how there can be similarities between the European and the American evaluation discourses when the roots are not the same Stame dealt with separately.
"Is it simply a matter of an academic debate, with few links to practice?
We could offer a few tentative answers:
   a. the evaluation discourse branch is something that refers to what theorists believe evaluation is about, and how it should be done; hence, it is normal that these ideas travel to other countries and are received, re-elaborated and integrated into different contexts.
   b. The picture we have drawn can be considered the context in which European theorists elaborate their own approaches in order to tackle institutional predicaments (or even opportunities) offered by the European roots. Realist evaluations, theory of change evaluations and other approaches used with complex socio-economic programs are a way of tackling issues that cannot be dealt with by the linear models that continue to be presented in most evaluation guidelines.
   c. New approaches could emerge with less institutionalized policy arenas. For instance it could be interesting to see what evaluation designs and methodologies have been developed in the peer reviews that have been done as a tool for the Open Method of Coordination, a way of implementing EC policies in fields where the autonomy of states remains undisputed, as in employment policies.

---

[32] The first such programs were called "Poverty", reminiscent of the War on Poverty, that initiated both programs and evaluation. only in some parts of the country.

     d. More freedom for original approaches to evaluation could be found in less institutionalized policy arenas, or - this is the hope - in cases when new policy styles prevail thanks to special experiences with innovative programs, integrated policies and other ways that defy well established practices." (Stame 2006, p. 6 f)

3.4 Frans Leeuw.
Frans Leeuw recommended Marvin Alkin to put Nick Mullens on the evaluation theorist's tree, this because of the work *Theory and theory groups in sociology*. This was a network analysis in which it was evaluated which researchers were cited by different researchers. Alkin was positive to this idea. (2b 45 min)

Leeuw basically agreed to the three lines of evaluation, that is discourse, institutions and practices, which were presented by Petersson – Karlsson. He believed that a study on theorist's views would be less relevant for Europe than for the USA. He didn't want to totally neglect the importance of "intellectual heroes" though. This was exemplified first with Karl Mannheim who in 1934 in held a plea for program theory in Principia Media. A second example was launched with the British criminologist Flight (?) Sherman who in the 1950s did experimental probation research. In Leeuw's judgement the group should engage also in this theme, but not put it up front. It was suggested that the members that the former and contemporary members of the European Evaluation journal, now 15 years old, could be studied, what they have done and which other researchers they have cited. (2b 50 min)

But basically Leeuw agreed with Petersson – Karlsson on that the state institutions generally are the most important entities for understanding evaluation;
"I fully agree with you on that institutions in this part of the world are so much more important. National audit offices have in the 80s played a very important role in trying to get evaluation going within the executive. But if you ask who are the intellectual heroes are within these National audit offices only the people of the inner world will know, but there is no Chelimsky. The institutions have been more important than the people. The institutions really have in a number of European countries, though not all of them, played an important role in getting evaluation going. Do you know why? The asked the question what do the governments know about the NGOs, about their effectiveness. The answer was: We don't know. And then the National audit offices said: Thy should know. They put pressure and they had of course intellectual power. They could go to parliament, they could blame, lame and shame you, and that's the power of a number of institutions." (2b 53 min)

Other examples supporting this line of reasoning were also presented and a question for further research was presented:
"Then of course the question is why these institutions are more important in some countries than in others and why they are more important in some fields than in others. That should be on the agenda to find out: What makes these institutions leaders instead of a Peter Rossi?" (2b 55 min)

Leeuw also suggested that the approach should be underpinned with neo-institutional theory, as this mixture of sociology, economics and history could be helpful in explaining why the [state?] institutions sometimes fall apart, but sometimes get very influential. (2b 57 min)

As to the part of the Petersson – Karlsson paper on evaluation practises Leeuw recommended that it should be replaced with a part on evaluation systems. The reason for this was that "everything is a practice", but what a "lonely hunter evaluator" is doing is not important.

Leeuw informed that in an ongoing project he and Furubo are working on bringing forth a typology consisting of 6, 7 or 8 systems of evaluation. Perhaps this could serve as inspiration, so the suggestion. (2b 60 min)

Leeuw also launched some critique. In contrast to Petersson - Karlsson, he was most sceptical to presenting differences in national culture as an explanation for differences in evaluation practices. In the paper in question different welfare regimes were dealt with as explanations for differences in national evaluation practices and long roots of the kind described by Stein-Rokkan were recommended for describing differences between countries of the same regimes. The reason for criticising this approach was that it doesn't present a mechanism. Instead Leeuw emphasized the tendency of isomorphism, thereby once more using the neo-institutional approach. In that he cited a hypothesis developed within a research project in which he himself had taken part: (2b 60 min)
"Certain evaluation methods develop their own lives and achieve such a degree of status in certain countries that their appropriate application to a given educational sector is not questioned." (The Danish Evaluation Institute, 2003, p. 19.)

The study showed that the summative evaluations and the stake-holder model were used more and more in 11 of the 12 countries studied. That is, there was no variation depending on regimes. That means, so Leeuw, that a path dependency could be observed, which was the same in almost all the countries studied. The introduction of evaluation models in new areas is not motivated by regime appropriateness or the arguments of intellectual heroes; the path dependency has "a reality of it's own", Leeuw argued. (2b 63 min)

This was exemplified with the introduction of new evaluation methods in the Eastern European transition states after 1989. These states have since then copied the evaluation traditions of the western countries. There was also another empirical critique against the welfare regime-line of arguing; in the Netherlands evaluation since the 1970s was embraced by as well socialist governments as also by liberal ones. However, it was also said that the purposes for which evaluations were launched differed with the changes of governments [which however supports the line arguing of Petersson – Karlsson]. Other examples were also presented on the universal use of certain methods at different times. One was the spread of the methods advocated by the Campbell corporation. (2b 66 min)

3.5 Afternoon comments.
Evert Vedung made a comment on the paper of Petersson – Karlsson. His remark concerned an unclarity, which seems to have been somewhat confusing. The term "formal institution" is used without a definition. Vedung meant that a distinction should be made between organisations, which launch evaluations, and the systems of evaluation, which these organisations adopt. (1a 6 min)

Jan-Eric Furubo launched a hypothesis on the importance of constitutional courts in forming the evaluation practice. The thesis was that if there is a constitutional court then the common conception on what should be considered relevant evidence in evaluations of public policy emphasizes evidence in a legal sense. (1a 15 min)

References.

The Danish Evaluation Institute, 2003, Educational Evaluation around the World – An International Anthology. Copenhagen: The Danish evaluation Institute.

Furubo, Jan-Eric, Ray C. Rist and Rolf Sandahl, 2002, *International Atlas of Evaluation*. New Brunswick, N.J.: Transaction publishers.

Petersson, Gustav Jakob and Ove Karlsson, 2006, "An Evaluation Tree for Europe". Paper presented at the Symposium "Evaluation roots in the USA and in Europe: tracing traditions", Stockholm, June 17-19, 2006.

Stame, Nicoletta, 2006, European Evaluation (Theory) Tree. Draft of the presentation at the Symposium "Evaluation roots in the USA and in Europe: tracing traditions", Stockholm, June 18, 2006.
References for the citations from this paper:
- Davidson E.J., 2006, "review" of Alkin M, Evaluation Roots, in American Journal of Evaluation, vol. 27, n. 2.
- Derlien H., 1990, "Genesis and Structure of Evaluation Efforts in Comparative Perspective", in R.
Rist (ed.), Program Evaluation and the Management of Government: Patterns and Prospects across eight nations, New Brunswick, NJ: Transactions Publishers.
- Manski C. M. and Garfinkel I., 1992, "Introduction", Evaluating Welfare and Training Programs, Cambridge, Mass: Harvard University Press.
- Wildavsky A., 1979, Speaking Truth to Power, New Brunswick, NJ: Transaction Publishers.

Recordings from the Symposium "Evaluation roots in the USA and in Europe: tracing traditions", Stockholm, June 18, 2006:
- 2a
- 2b
The recordings are preserved by Gustav Jakob Petersson.

Comment on the Paper: An Evaluation Tree for Europe

Wolfgang Meyer and Reinhard Stockmann, CEval Saarland University

1   An alternative View on the Classification Scheme

1.1 Discussing the Tree Model

We welcome the idea to systematize European evaluation and to compare it with American evaluation. But we are not sure if the Alkin Tree Model is the best basis for doing so. The tree model was developed for structuring the great variety of evaluation *theory*. It is not directly evident that the Alkin Model can be used for a description of the European Evaluation *structure*:

1. We can't see a single trunk. At least there are two different scientific roots (constructivism and positivism) which influenced the foundation of evaluation and the discourse until now.

2. Disciplinary thinking formed evaluation and differentiated it. Different trunks are developed from economy, sociology, education, psychology, public administration, policy analysis. So there are different roots leading to different trunks and not a single trunk with the same roots. The European Evaluation structure is more like a briar with woven and sometimes tangled branches. Therefore, a bush (maybe even a jungle) is more an appropriate picture for the Evaluation situation in Europe than a tree.

3. Why there should be only three branches to structure the Evaluation situation (discourse, institutions, practices)? There is no clear reason for that. We think there should be more categories used (see 1.2).

We agree in the position, that national history, political frameworks and other contextual factors are central in forming a national evaluation tradition. We believe that the individual circumstances which formed a national evaluation culture are so different from each other, that there is no European tree with a single trunk in the moment. Hence, under the pressure of the EU some kind of harmonisation in evaluation in Europe takes place. This tendency may lead to the formation of a single trunk of future evaluation research in Europe.

In the moment we would prefer another more clear classification scheme which includes the three mentioned dimensions (branches) but which operates with more than these three categories.

1.2 Other Classification systems

We think the classification system of Furubo and Sandahl (2002) is an interesting alternative approach. To assess the maturity of the national evaluation culture they use nine different indicators (number of policy domains, supply of domestic evaluators, *national discourse*, professional organizations, degree of institutionalization government/parliament, pluralism of institutions/evaluators performing evaluations, evaluation within the Supreme Audit Institution, and proportion of outcome in relation to output/process evaluations). These nine indicators are used to make a ranking of 21 European and Non-European countries (p.10).

What are the arguments for and against the Evaluation Tree compared with the Classification Scheme of Furubo and Sandahl?

From our point of view, the Furubo and Sandahl classification scheme has more advances than disadvantages compared with the tree model. Nevertheless, we are not absolutely satisfied with the scheme, because of its complexity and the visual classification difficulties. For example, we both tried to rate Germany separately following this scheme and our results differed significantly from each other as well as from the published rating.

The main problem is the overall judgement on a *national* evaluation culture. By using different weightings of evaluation practises in various policy fields and sectors, we came to different results on the national level. With respect to the huge differences between sectors, the national 'mean' is not meaningful and it seems to be insufficient in referring only on nation as a classification principle. Our different judgements on the evaluation practice in Germany were caused by the variation of evaluation practices in different sectors and our diverse opinion on the importance and contribution of each single sector to a national evaluation culture. We assume this not only being a German problem and we strongly recommend keeping an eye on intra-national variations.

The utilisation of Furubo and Sandahl's classification scheme is limited to comparisons of *nations*, because of the criteria used for classification. Therefore we want to suggest another opportunity, giving us the possibility to classify both nations and sectors (and maybe other aspects like, for instance, theoretical schools). However, this is also a starting point for discussion and we should try to develop a proper and satisfying solution.

Another Opportunity

As a starting point for another classification scheme the following idea could be used: By conducting an evaluation, one has always to answer the following question: What will be evaluated by whom on which criteria with which objectives?

These questions can be used to structure the national evaluation field(s). Additionally, the key understanding of evaluation can be added to outline differences between sectors, nations, schools etc.

The advantage of this classification system is its universality: the guiding question must be answered for all kind of evaluations in all circumstances and independent from time, sector, nation, theoretical school or whatever. Compared with Alkin's tree, it does not use a fixed structure for organisation and gives way for other figures and models to describe the evaluation structure. Compared with Furubo and Sandahl's classification, it does not refer to national policy items and criteria, reducing the classification mode on nations and political systems. Therefore, we favour this classification scheme.

Figure 1: Classification Scheme

Indicators	Operationalization	Country: Germany	Sector: Development Aid (Germany)
Understanding of Evaluation	constructivism vs. positivism, implementation vs. impact assessment, formative vs. summative evaluation, self (internal) vs. external evaluation	More positivism than constructivism. More formative than summative, More external than internal.	More contructivism than positivism. More formative than summative, More internal than external.
'What'? (Subject of Evaluation)	Laws, policies, program, project, sector, organisation, individuals	Mainly programs and projects	Mainly programs and projects
'Who'? (Institutions of Evaluation)	Independent public evaluation institution (e.g. state agencies), ministry evaluation departments, university and research institutes, consultants, evaluation departments within implementing organisations	ministry evaluation departments (1-2), seldom university and research institutes, mostly consultants, very seldom evaluation departments within implementing organisations	ministry evaluation department, seldom university and research institutes, mostly consultants, evaluation departments within implementing organisations (less in NGO)
'Which Criteria'? (Criteria of Evaluation)	Effectiveness, Efficiency, Outcome, Impact, Sustainability, Relevance, Significance (DAC-Principles)	Mostly Effectiveness, in some parts also Efficiency	Mostly Effectiveness, Outcome, Relevance, Significance; Efficiency only for huge investments (KfW)
'How'? (Approaches and Methods of Evaluation)	Objectives-oriented, Management-oriented, Consumer-oriented, Expertise-oriented, Participant-oriented (Fitzpatrick et al 2004)	Objectives- and Management-oriented	Objectives-/ Management-often Participant-oriented
'Which Objectives'? (Target of Evaluation)	Knowledge, Accountability, Development (Learning), Legitimacy (Stockmann 2006)	In general: Accountability (Erfolgskontrolle), Legitimacy.	Accountability (Erfolgskontrolle), Legitimacy, Development (Learning)

1.3 Sector- vs. Nation perspectives

As we mentioned above, we share the opinon that national history, political frameworks and other contextual factors over time play an important part in forming a national evaluation culture. So we agree that the national perspective is an important one. But we have to keep in mind, that organisation of Evaluation in Europe happens in disciplines, not in nations. Different scientific traditions are leading to different concepts and understanding of evaluation. But: the scientific exchange is discipline-oriented, not nation-oriented. For sure, this is a research and scientific oriented view. In contrary, the use and utilization of evaluation is more nations specific and depends on the political system and policy styles. Therefore, the nation perspective is important, but a sectoral perspective should be at least added. *We assume the sectoral differences being bigger and probably more important than the differences between nations in Europe.* Moreover, we assume that the EU will homogenises the nation differences, but not the sectoral differences.

2 Comment on Thesis'

2.1 Discourse

The evaluation discourse is an important issue for analysis. The differences between Europe and USA are worth to be explored.

Europe is an importer of evaluation ideas from USA and there is no comparable export viewable at the moment (S. 4). For example, the standards discussions in Switzerland, Austria and Germany (and some other countries) have overtaken the US-discussion from the 1980s. As far as we know, the standards discussion in these countries are absolutely unknown in the US-context and they do not have any influence on recent debates within the US-community.

There are main differences between the US and the European discourses on evaluation in the evaluation communities. In several European countries there are national traditions which are significantly different to the US-traditions. In Germany, for instance, there is almost no evaluation theory discussion which can be recognised by three aspects: first, there are no evaluation theory books in German language written by German authors available. Second, in the German Journal of Evaluation, no theoretical article has been published jet. Third, there is no special interest group on evaluation theories in the DeGEval, the German Evaluation Society. This is a significant difference to USA and AEA where theoretical issues are quite important from the very early beginning.

Nearly the same can be said concerning the methodological issues of doing evaluation: while there is a strong discussion on evaluation methods in USA, there are no specific debates and interest groups regarding this topic in Germany. Another difference is the so-called 'self-evaluation' debate in German Evaluation Society. The German Evaluation Society developed specific standards for such kind of internal evaluation run by the actors itself. Even the term 'self-evaluation' can not be found in the USA evaluation discourses. The German discussion on self-evaluation is by no means overtaken from the US-discourses and it seems to be a national specific discussion.

There are also strong differences in the evaluation fields of interest between USA and Germany (and probably other European Countries). For example, in Germany there is a branch on implementation research which can not be found in USA. The German tradition in social inquiry is accompanying research ('Begleitforschung') which is strictly oriented towards implementations, not towards impacts. The policy experiment of implementing comprehensive schools in the early 1970s is a very good example for this kind of research strategies. In Germany, policy research is not an impact research, it is more historically and

phenomenologically oriented than the more empirically and positivistic oriented discourses in USA (and UK and European countries).

Besides this directly social science oriented evaluation discussions, there are also different views in other disciplines on the issue of evaluation. For example, in Germany there is a strong tradition on evaluation of technical equipment and instruments. There are specific institutions on this issue, for example, the TÜV (Technischer Überwachungsverein – Technical monitoring association). The German tradition of 'Technikfolgenabschätzung' (engineering results assessment) is strongly influenced by this technical understanding of 'evaluation' which is not recognised as a comparable tradition of evaluation in the social science discourse. A comparable example is the environmental debate on evaluating of several technical issues which ignores the social science branch of evaluation.

2.2 Institutions

We see some evidence for the thesis, that European states are characterised by public institutions for evaluation while public bodies in the USA tend to out-source evaluations to private enterprises (p. 5).

In Germany, there is a domination of public institutions in evaluation, but only in a few sectors. We see the reason especially in the high importance of control, inspection and supervision in German ministries. Some of them build up subordinate institutions, like the Institut für Arbeitsmarkt- und Berufsforschung, IAB (labour market), Bundesinstitut für Berufsbildung (vocational training), Umweltbundesamt (environment) etc..

But there are other sectors which are more privately organised (e.g. development cooperation). Moreover, there is a strong development from state to market at least in the ideologies of political leaders, resulting in a trend of 'outsourcing' evaluation. E.g. University Education.

In Germany, the institutionalisation of evaluation is extremely differentiated in policy fields and sectors. For example, evaluation is well established in sectors like labour market policy, development cooperation or university education. However, there are strong differences between these three fields concerning the way of implementation.

In labour market policy, evaluation refers to the economic tradition and is mainly limited to quantitative econometric analyses. It is well established in a state-run research organisation IAB. But there is no use of any qualitative methods in evaluation and no program evaluation at all.

In the field of development cooperation much more qualitative methods than quantitative methods are used, the evaluations are mainly restricted to project and program evaluation. Country- or sector evaluations are very rare. Evaluations are conducted by internal evaluation departments within implementing organisations (less in NGO) and a variety of private consultants. Scientific Institutes and universities are only weakly integrated.

In university education evaluation is a completely private business, run by accreditation and evaluation societies and companies.

These examples from Germany show the variations between sectors especially concerning the institutionalisation of evaluation practices. By using a cross-national comparison perspective, such sector differences will be ignored and artificially harmonised for producing a 'national type' of evaluation culture. We are not sure, if the differences between policy fields and sectors are 'typical German' and we assume other countries having similar or other fractions.

2.3 Practices

In Germany evaluation is concentrated on a few policy sectors only. This situation has not much changed in the last 30 years. We have not the impression that German researchers get very often involved in high-level public evaluations for public sector activities or that evaluation projects tend to be of a greater scale than the ones of the USA (p.6). At least the roots of evaluation research in USA are laid by some huge policy programs and evaluation studies in the early 1960s. We do not know any comparable national programs and evaluation studies in Europe yet.

There are some evaluation studies of greater scale in Germany, but they are quite seldom. Most of them are international evaluation studies from the OECD (e.g. PISA) and especially from the EU. One of the very rare national examples is the evaluation of HARTZ IV (labour market policy), which is merely an evaluation of economic impacts and not a program evaluation. The huge majority of evaluations in Germany are of smaller scale, sometimes even of extremely small scale. Even the international high-level public evaluations from the EU are broken down to small scale evaluation projects at the local level (e.g. INTERREG, EQUAL, ESF).

3   Explanations und Alternative Explanations of National Difference

In general, we find the theoretical argumentation on the roots of the European evaluation trees very fruitful and we suggest following this direction. There are a lot of interesting differences between the national states, political systems and political traditions. Compared to the development in USA, the political history of Europe is more broken and divided, bearing a lot of interesting variances that may be used to explain differences in evaluation culture. Just to mention some:

-        The east-west-confrontation between socialism and capitalism divided Europe for half a century. The polarised debate between state and private-ownership (and leadership) both in politics and in economy never happened in the same extents in USA.

-        The pluralisation of languages hinders the exchange of ideas. By calling USA a multi-cultural society, one has to keep in mind, that the USA are ONE state and there is at least a leading culture of being 'American'. The division in Europe is obviously much deeper and we are just at the beginning to develop an European identity.

-        Languages are more or less a symbol for different cultural traditions. In Europe, one can find the Anglo-Saxon tradition in the North, the Francophile tradition in the West (and some parts in the East), the German tradition in the Middle, the Balkan Tradition in the East and so on. The differences in policy and living stiles are significant and they are not bound to national borders or forms of state organisation.

This list is of course limited and only used as a short hint. It is worth to explore this issue more systematically and to test its influence on the development of evaluation culture in different nations and sectors.

3.1 Explanations

Some Explanations are presented in the paper. For the 'long roots' of evaluation Stein Rokkan's approach in describing the political formation and nation-building processes in Europe is suggested as a tool. We strongly agree with this argument and we believe that Stein Rokkan's approach can be a good starting point for exploring differences in political systems

and in evaluation culture. Moreover, Stein Rokkan's approach may help us to understand the 'new nation-building processes' in Eastern Europe (especially in the new states Latvia, Lithuania, Estonia, Slovakia, Slovenia, Bosnia, Croatia, Serbia, and – the latest one – Montenegro) and its influence on new evaluation cultures.

Hence, we strongly agree that Stein Rokkan's approach is not the only model to explain long-lasting developments of political systems. Harald Swedners attempt to structure the European history of ideas may be used as well as Immanuel Wallersteins World System Theory (Wallerstein 2004). These are only examples for explaining differences in long-lasting political development processes that may influence the process of establishing an evaluation culture. Having these approaches in mind, additional dimensions (besides nations and sectors) may come into the focus of analysis.

The paper is giving an example by using Gösta Esping-Anderson's theory on the changing nature of welfare states in Western societies and his model of 'Three Worlds of Welfare Capitalism' – liberal, conservative, and social democratic welfare states. Without any doubt, the welfare state model is another very important variable that may influence the development of evaluation cultures. But one should not forget that evaluation research is a very young discipline in Europe. Most evaluation societies are built in the 1990s and even if we take the longest roots, they will not reach much farer than into the 1970s. In our opinion, the differences between the 'three worlds of welfare capitalism' are weakened in the last decade according to welfare state 'reforms' which are more or less financial 'cut-backs'.

However, we suggest to follow this long-term developments and explanations, but we also suggest to strengthen the argument, how the aspects mentioned by authors like Esping-Anderson, Rokkan, Swedner or Wallerstein should influence the development of evaluation.

3.2 Alternatives
We like to mention some alternative theories and approaches for exploring the variation of evaluation culture in Europe. This is also just an 'add-in' to start with and we do not claim to cover the spectrum in total or to give a systematic overview.

3.2.1 Role of the EU

As mentioned above, the role of the EU in homogenising institutions not only in politics but also in science and in economy should not be underestimated. We assume that the EU will be the most important actor for evaluation research in Europe during the next twenty years and it is not easy to forecast the development of evaluation in the EU. We assume a decline of national differences in evaluation practice, but a stable variation (maybe even increasing) between the sectors.

3.2.2 Phases of reforms
Another important aspect may be the varying dynamic of reforms, resulting in phases of stagnation and phases of innovations. One important approach may be the assessment on life-cycles of rules (Schulz, 1998; March, 2000; Beck u. Kieser, 2003). In the long run, rule innovations in rule-systems may lead to over-regulation and blockades ('bureaucratic-scleroses') and to a collapse of the whole rule system. A new rule-system with a new life-cycle replaces the old one in a political reform process. Regulative deficits are causing new rules and an increasing amount of regulations. Finally, the new rule-system will also come to a blockade and will be followed by the next reform.

This model is not only useful for organisational reforms but also for policy research and reforms. By using this model, each sector is on a different state of 'rule reform dynamics' and evaluation maybe an important tool to influence this development (and it maybe used different on different stages). National or sectoral differences in evaluation practise maybe a result of different dynamics or phases of reforms.

For instance, in Germany the "Reformstau" (reform jam) in labour market policy has been publicly criticised in mid-1990s. As a result, the government started the HARTZ-programs to renew the labour market with great optimism (2 Mio. new jobs were promised). Other reforms, for example in the health system, had been started too. In general, the Schroeder-area is a time of huge reform programs, while the Kohl-area was a phase of stagnation. But: in some sectors, nearly no reform happened during the Schroeder-area and reforms occurred during the Kohl-area (e.g. in development cooperation).

There are no studies on the effect of this 'reform climate' on the development of evaluation practice, but we think it is worth to work on this issue.

3.2.3 Governance
Up to now, evaluation has always been seen as a state task. However, the paper assumes a difference between USA and Europe regarding the extent of public and private initiatives. In doing so, the question of 'governance' is getting important (Kooiman, 2003; Sharma u. Bareth, 2004; Held, 2005). Is the involvement of civil-society, interest groups, private companies supporting or hindering the development of evaluation practices? Our experiences offer examples for both – in most cases, huge state organisations are running and developing evaluation systems. Moreover, they force private organisations to follow their way and they claim for evaluation practices to legitimate public investments. Private organisations, especially non-profit organisations, sometimes blame evaluations for being a needless new condition for getting public money. But on the other hand, there are also some non-profit organisations making use of evaluation (or 'self-evaluation') without any pressure from the state. The role of the private sector for developing evaluation practice is at least not quite clear in the moment.

4 Answers to the Further Questions

4.1 Is there a typically European evaluation tradition? Are there differences within Europe so pronounced that a generalizing European evaluation tree would also mirror the Evaluation tradition of the United States?
No, not yet. The EU might form one, but at the moment it is fragmented according to historical and cultural spefica (not only national, but also language regions or sectors).
No generalisation possible, no mirroring of the USA possible.

4.2 If Alkin's tree structure is used, which European names would you like to place an the three branches?
From the German perspective, we do not see comparable influential names to build an evaluation tree on theories or schools. The reason is the lack of theoretical debates on evaluation issues in Germany (see chapter 1.3 and 3.2.1).

4.3 If you compare the formal institutions for evaluation and the evaluation practices in the EU with such institutions and practices in your respective home countries, which differences do you see?

The EU is sectoral differentiated, but centralised within the sectors. There is no 'bottom-up' accumulating process for finding a new evaluation culture by making use of the different experiences and varieties of approaches in European countries. Instead, the evaluation system is a bureaucratic 'top-down' model, following the principles and ideas of EU-administration (see chapter 1.3). Therefore, the EU forces the evaluators in all European countries to use their own evaluation system and, up to now, there seems to be few opportunities for improvement and for initiating a learning process. At least in some sectors, the German government and other state organisations deviate significantly from such a bureaucratic 'top-down' steering model and evaluation methods are developed more participatory and in dialogs between state, consultancies, scientific research institutes, and stakeholders.

4.4 Do you agree with us on that evaluation tradition is bound to its historical and societal context?
Yes, definitely (see chapter 3). However, we believe the historical and societal context and its development should not be reduced to a national perspective but should also include at least variations between sectors and policy fields.

Literature:

DAC-Principles

Beck, N. & Kieser., A. (2003), The Complexity of Rule Systems, Experience, and Organizational Learning. In: Organization Studies, 24, 793-814.

Fitzpatrick, J.L., J.R. Sanders & B.R. Worthen (2004), Program Evaluation. Alternative Approaches and Practical Guidelines, Boston u.a.: Pearson (3.).

Furubo, J.E. & R. Sandahl (2002), Diffusion Perspectiv e on Global Developments in Evalution in: Furubo, J.E., R.C. Rist and R. Sandahl (eds) International Atlas of Evaluation, New Brunswick: Transaction Publishers, 1-26.

Held, D. (2005) (eds.), Global Governance and Public Accountability, Malden u.a.: Blackwell.

Kooiman, J. (2003): Governing as Governance, London u.a.: Sage.

March, J. u.a. (2000), The Dynamics of Rules. Change in Written Organizational Codes, Stanford: Stanford University Press.

Schulz, M. (1998), Limits to the Bureaucratic Growth: the Density Dependence of organizational Rule Birth. In: Administrative Science Quarterly, 43, 845-876.

Sharma, B.M. & Bareth R.S. (2004), Good governance, globalization, and civil society, Jaipun: Rawat.

Stockmann,R. (2006), Evaluation und Qualitätsentwicklung, Münster: Waxmann.

Wallerstein, I. (2004), World System Analysis. An Introducti

Attachment 2.9: Nicoletta Stames handout of June 18 2006.

	US	EUROPE	EUROPEAN STATES
roots	Pragmatic administration	Different administative cultures, from more to less centralized EU admin: modelled on French	France: Napoleonic centralized UK: scandinavian:
	Empirical research methods	Similar as US	
	Scarce state intervention	State intervention widespread	Different welfare states: Social-democratic Liberal Conservative
	Consequence: need for evidence of success	Consequence: Different attitudes to evidence	UK: evidence based policy
	Federalism	EU between intergovernmental system and multi-level governance	Federalist: Germany Centralized: France, UK, Italy until recently
	Programs (big and small)	Laws not programs. EU introduced programs (big ones) and projects	
branches			
Evaluation discourse	methods	Similar as US ? Realist evaluation	
	values	Similar as US ?	
	utilization	Similar as US ?	
Institutions	Institutionalized: PEMD, GAO	Institutionalized . EU: evaluation units	France: CNE Italy: evaluation units
practices	outsourced	Outsourced: independent evaluation? In house	
	Big evaluations with War on Poverty,	Big evaluations of EU programs	
	More small evaluations	Big and small national evaluations	
		EU evaluation guidelines	
	Systematic reviews	Systematic reviews Realist syntheses	

This table is an unfinished attempt at showing how the various elements (branches and roots) would fare in comparative analysis:
- Branches are those listed by Karlsson and Petersson
- Roots are those listed in the present paper.

Attachment 2.10: Notes from panel discussion October 5 2006.

Evaluation roots in the USA and in Europe: tracing traditions?
Session 3, EES/UKES Conference, London, U.K.
Thursday, October 5, 2006, 11–12:30

Notes prepared by Rahel Kahlert

Important: The notes were taken electronically during the session and are not a transcript based on a recording.

Marvin Alkin
In our book, we talked about prescriptive theory of individual evaluation experts how they believe how evaluation should be conducted. For example, Patton describes how he thinks evaluation needs to be done. We looked mostly at methodologists, and not at people who do not have a fully developed theory, also not at practitioners.

"Roots" is a theory about theories. Often people state that "You are really a theory historian." I disagree: "I am a theorist of evaluation theories." The organizational structure of the book is that theories are concerned with three issues: *Methods*, *Valuing*, and *Use*.

The theory tree is based on these three issues. The book is North American centric. For the purpose of this meeting, we eliminated some issues and reexamined the tree. This is the 3rd edition. We made minor positioning changes on the branches.

These three issues are now used to characterize the historical path. For example, Stufflebeam represents *Use*, Campbell represents *Method*s, Scriven and Guba&Lincoln represent *Valuing*. Then follow Owen, Wholey and Tyler in the *Use* branch. Stake and House are *Valuing*, and Cook is *Methods*.

The *Roots* book looks at statements that people made in their chapters. We then categorized the statements by the branches of the tree. Great preponderance of the same branch has enormous influence on other people on the same branch.

Franz Leeuw was key in doing this type of analysis in the Dutch context.

Christina Christie
[Christie shows a paper copy of the new tree.] This is the third published version of the evaluation tree. We did a content analysis of the theory of evaluation to develop the tree. Feedback came also from students. This is a great teaching tool for evaluation courses to focus on the North American perspective.

The *Valuing* branch was always a sticking point. The *Use* and *Methods* branch make sense. When explaining the tree, brakes are typically put on the *Valuing* branch. For example, it is not intuitive why Scriven and Guba&Lincoln end up on the same tree branch. The issue of the *Valuing* branch can be traced back to the philosophy of science, especially the constructivist and relativist paradigm.

What about splitting the branch up? An administrative assistant—a history student—liked to sketch out what is going on, by taking into account the theory of cognitive psychology. He sketched the *Valuing* branch with a little hole. Stake and House are on the bridge between Scriven and Guba&Lincoln. They endorse the idea of deliberate democratic evaluation, and also endorse procedures with the *Methods* branch. Stake introduced case study evaluation. He was trained as a psychometrician. Some inconsistencies may exist in the writing of an evaluation theorist. Scriven is seen as a post-positivist who is somewhat lonely. Then House and Howe are more perceived as constructivists.

Is there particular tree for the European context? It poses a challenge to this tree idea.

Nicoletta Stame
I like the idea of an American tree of theorists. But there is something common of people historically since the 1960s in a system different from Europe. The tradition is more federalist and more pragmatist. The most important characteristic of this system is that in the U.S. there has been a great debate among theories. This U.S. context is favorable to discussion. There is not a culture that tries to assess that there is the right way. They do not have the characteristic of the state that we have.

What is a difference between U.S. and Europe? We add the importance of the state. The U.S. does not trust the state. We trust the state, so we do not need to show that it is good. This is my main assumption, which has a negative side. We have been a welfare state for centuries. We have a different culture that is not favorable to evaluation. We introduced it because of the fiscal crisis of the state. Also because the European Union introduced programs such as poverty programs like in the U.S. There has been some change in the working of the state. Nevertheless, evaluation was introduced in a different environment than in the U.S.

There is the idea that there is one way in which evaluation should be done. Europe has the distinction of academic environment where we debate about theorists. Actually we have very good thinkers. Unfortunately, this is an academic debate in Europe, in which evaluation has been institutionalized in the state.

I like the metaphor of the tree. We should inquire about the roots. The trunk has social accountability and fiscal control. This would be similar in Europe. But we also would need to add some type of bureaucracy of which we are so proud. Bureaucracy is disappearing. I think that evaluation in Europe needs to open its mind to this kind of debate, and to be more open to the possibilities that evaluation can be conducted in different ways. I hope that the theory tree will prosper and blossom, but there is some need to change the roots.

[Polemical comment:] Alkin is revising his tree every three months…

Frans Leeuw
It is wonderful to have this tree, but "Wo sind all die Maenner her?" (where do all the people come from?). Where do all the organizations and institutions go? Right now, there are only vague ideas, no strong statements.

The whole session is an invitation to join us to develop the "roots" of a European theory. We are the court of justice of theory and theoretical roots. Skip lunch and start to work together with us.

Why pay attention to institutions and organizations? Why do we not find this in the U.S.? In the Netherlands, since the brick and mortal institutions in the 1930s, everything is embedded in state institutions. In France, there is Bourdieu, but not in the field of evaluation. A Scriven type is hard to find in the UK and France.

The U.S. *Roots* by Alkin is a study about the brains and power of pushers, or wheelers and dealers. See James Coleman's *Foundations of Social Theory* about individual behavior and corporate behavior. Birthrate of corporate institutions is often higher than of natural individuals.

Europe is a country in which individualism is not as high amongst cultural values. (Leeuw comments: Gary gives me a positive incentive by nodding). The U.S. is 91%, but Europe is less, such as Austria is 55%.
There is a brain-drain from Europe to the U.S., such as Lazarsfeld. Believing that institutions matter and organizations codetermine is higher in Europe.

Do you agree on this preliminary statement that persons are less important in European history of the evaluation roots? Young Go West?

Owe Karlsson Vestman
We have a paper written about the evaluation tree for the European situation with Gustav Petersson. We tried to follow the same metaphor. But we have different names on the branches.

The three branches are: Evaluation discourse, institutions for evaluation, and evaluation practices:
- Evaluation discourse: What is evaluation? Why do we do evaluation? What is the evaluand? We recognize Scriven and others behind that. Discourse is international. We do not think it is a good idea to make a big difference between the U.S. and Europe. This branch is related to *Valuing*, you find other ways of thinking. We think it is because of the ways we look at the state and society. It is about the importance for actually doing evaluations.

- Institutions for evaluation: Instead of names of heroes, there are institutions, such as Sweden's National Board for Higher Education, which implemented a system for universities for a six year review. Behind such a model is not a name, but an institution. This means that institutions are the important actors.

- Evaluation practices: This differs from discourse as the talk of evaluation. Evaluation practices take different shapes in different policy sectors in society. These sectors include social work sector versus education, health versus foreign aid etc.

Gustav Petersson
We pursued the ambition to elaborate models to explain differences between countries and regions. We believe that formal institutions are very important in forming evaluation traditions. European countries are mainly Western welfare societies. How these countries distribute welfare and allocate welfare in society should influence what evaluation traditions look like. We try to follow this logic. For example, how welfare is distributed by which formal organizations should have an impact which questions should be justified in evaluation. It could also influence how you structure the evaluation in the state.

Other questions are: Institutions, where are they, how are they structured in a state, what questions do they raise? If you try to explain differences in this way, we see a short root of difference in evaluation.

Why do countries choose to structure services in different ways depending on different welfare states? We will see whether this will be useful in the future.

Reinhard Stockmann
The evaluation tree is a wonderful model for the U.S. For structuring the theoretical discussion it is fine, but it is not the best model for European evaluation. It is not suitable to compare. The problem is that they needed three trunks rather than one. All panel participants agree in the special role of institutions. For example, Leeuw will talk more about this.

Almost all institutions are formed and developed within specific sectors. Examples are multinational and big companies that work in selected branches such as mechanical engineering, but not in the entire economic spectrum. Boundaries within subjects are more important than national boundaries. Then we have administrators who take action in special policy fields. The organizations of the not-for-profit sector are concerned about different issues such as fight for human rights, environments. This differentiation is very similar to evaluation.

In society as a whole, governments, with their institutions of the nation state, are influential, but nation states decline in importance. Institutions establish evaluation systems and shape evaluation practices. We need to pay special attention to sectors and policy fields and disciplines. There is a different evaluation discourse; and methods and evaluation systems are different. There are intensified differences that evaluation is not a well defined system established in different countries. They are highly correlated with policy fields. A national evaluation culture exists. But even larger differences between sectors exist. There is no national average. Nations cannot be a classification principle. Doubtless, national histories have an important role in forming national evaluation cultures over time. But development typically happens in sectors not nations.
The nation perspective is important, but a sector perspective should be added. Sector differences are greater and more important than differences between nations. There is not a systematic investigation among these.

The three dimension of Petersson/Leeuw or Sandahl could be used. I prefer a simpler classification scheme:
What will be evaluated by whom, using what criteria, and how will the evaluation be conducted for what objectives? The advantage of this classification system is that the guiding question should be answered in all evaluations. Compared with Alkin's tree model, it does not use a fixed structure. Compared with Furubo's characteristic, it does not emphasize nations.

The five questions are:
1. What is evaluated: subject of evaluation, e.g., laws, policies, programs, sectors, organizations, individuals etc.
2. Who does the evaluation: e.g., independent public evaluation, ministry, department etc.
3. Which criteria are used: e.g., DAC principles, effectiveness, efficiency, outcome impact, relevance, significance etc.

4. What approaches are used: Fitzpatrick differentiates between consumer-oriented, management-oriented, objectives-oriented, expertise-oriented, etc.
5. Towards which objective: knowledge-based objective etc.

These five questions are now applied to three sectors: development aid, education and labor market:

What is evaluated?
- Development aid: programs, projects etc.
- Education: organizations, courses, individuals.
- Labor market: laws, policies, privatized sector where governments determine only general conditions.

Who is doing it?
- Development aid: ministry, evaluation department, but seldom universities.
- Education: Merger of universities, private institutes doing evaluation work.
- Labor market: state authority, IAB.

What approaches are used?
- Development aid: Participant-oriented.
- Education: Expertise-oriented.
- Labor market: Objectives-oriented.

What is the target?
- Development aid: accountability, development, learning.
- Education: accountability, development learning.
- Labor market: accountability, legitimacy.

There are great differences between policy fields. Looking at sectors instead of nations gives us a more complex picture in Europe.
The different roots of national evaluation culture form different trees, different heights, different branches etc.

As you can see what we are into here is more of a jungle in this garden of evaluation.

Leeuw
Evaluation system was the topic of the last conference of EES. We wrote a short article about this topic. The reason for highlighting this topic is going back about 20 years, when the situation was quite different from today.

Evaluation systems for evaluation have been created in the last years. This kind of system is basically a somehow opposite ad-hoc perspective on evaluation. The system is characterized by factors such as pre-arranged organizational arrangement, created in advance to conduct and commission evaluations in certain policy domains. Systems have permanence and history. They have been created to exist for some time. They have also an idea of intended use. They have created institutional arrangements, which link the system to decision-making agents. These systems create a lot of questions.

What are examples of systems of evaluation? We brought together a number of empirical characteristics to make a book that sells even better than Alkin's.

You will be able to read all this on the website. I do not know which website, but there will be one.

In these systems, money is involved, it can destroy careers, put ministers down. Now we are in the flame.

What is the relevance for this *Roots* work? It is not only the search for persons or academic schools or paradigms, but also for systems. Systems will function as restrictions of opportunities for new evaluations. They help socializing new evaluators. They focus on collective action.

The discussion could be: Who disagrees with all this? Who thinks Furubo and I are victims of the reification disease?

Discussion

Sandra Speer
Is this just another approach? In Europe, we discuss more the political question of evaluation. But this approach is typical European, such as what is being discussed at EES conferences. For example, who is evaluating why and whom? This is another approach which is genuinely European.

Rahel Kahlert
You presented the U.S. tree and a possible European tree. What about analyzing the connection between the two? Since the U.S. has had a longer evaluation tradition.

Could the U.S. tree also be supplemented with institutions rather than individuals? For example, RAND, Abt Associates, Mathematica. I hear that some representatives of these institutions even came to Europe to promote certain approaches.

Alkin
We tried to come up with a way to incorporate institutions.

Why do we not have streams, what Patton suggested? We needed something that is understandable. The notion of sectors could be equally applicable to the States. For example, sector differentiation is not unique to Europe, but applies to across-the-pond as well. We do not have the same kind of disciplinary perspective. We are able to mix disciplines in the pursuit of understanding a program.

The European perception about U.S. evaluators is about the big studies you have read. Small evaluations are not read in Europe. These are not the conception that you have, which is working with program people to help them. There is this giant picture of what you think of U.S. evaluation in the European Union. Similar, what the U.S. thinks about European evaluation are these large projects in the EU. Both sides generalize about concepts.

Gary Henry
The classification system has simplicity in reducing complexity. Doing a classification system is to separate what needs to be separated. "Branches of a tree" is a good metaphor. The nature

is different in Europe. A hypothesis how the tree could be different is: A prescriptive base in the United States, but a more descriptive approach in Europe.

I am not sure what would constitute the branches. Because in the U.S. the folks on the tree are mainly academics, leading to a prescriptive set of branches that could spread across sectors and nations. What Europeans are dealing with establishing experimental traditions?

Zenda Ofir
What do we do in Africa? Do we have a tree? Evaluation has been externally driven. But there is a huge movement in development evaluation being indigenously and locally led. We try to foster that discussion in three ways:

- Explore in our cultures what prevents us and facilitates thinking about evaluation
- What is our understanding of evaluation?
- How have evaluation paradigms affected us?

The January 2007 conference in Africa invites to think through these three points. It is critical to start our own evaluation tradition. The last evaluation conference in Cape Town had more than 500 participants from over 50 states.

Alkin?
Gary's point is appreciated about descriptive and prescriptive. The third conjecture is that it is not only the difference between the prescriptive and the descriptive, but the number of collaborative arrangements could differ. For example, professors in the U.S. have incentives to work. Numbers of teams between professors, governments, and organizations work together in the U.S. The collaborative-arrangement approach could be used as criteria for difference. How does this criterion link to prescriptive and descriptive?

Stame
I do not believe that there is a difference between descriptive and prescriptive between Europe and the U.S. There are powerful systems in the United States, such as the GAO as the biggest institution of evaluation in the U.S. We have put on the table many things, but it would be interesting to pursue what we do with the descriptive approach. For example, issues about welfare states. We need to move toward a European tree.

3. Attachments of case study 3.

Attachment 3.1: Programme of the paper sessions of the SVUF conference 2005.

This programme of the paper sessions of the SVUF conference 2005 is to be found at
http://www.svuf.nu/papers_konferens2005.shtml (2007-03-30)

1:1 Metod I
Moderator: Eva Åström, eva.astrom@hsv.se

Kollegiegranskning med inslag av självvärdering.
Hanna Adolfsson, hanna_ulrika@hotmail.com

Självutvärderingssystem vid IVF vid MdH.
Ingalill Ekelund-Nordenmark, ingalill.ekelund.nordenmark@mdh.se

Utveckling av fokusgruppsmetoden för nätmiljö.
Pernilla Westerberg, pernilla.westerberg@ucer.umu.se
och Lisa Almqvist, lisa.almqvist@ucer.umu.se

P 1:2 Villkor och värderingar I
Moderator: Kim Forss, kim.forss@mailbox.swipnet.se

Utredning som politiskt verktyg.
Jenny Öhman Persson, jenny.ohman-persson@riksrevisionen.se

Utvärdering – vår tids bekännelseskådespel.
Charlotta Karlsson, charlotta.karlsson@hsv.se

Evaluering och verksamhetsuppföljning i sociala frivillighetsorganisationer.
Stig Linde, stig.linde@svenskakyrkan.se

P 1:3 Styrning, mål, intressenter
Moderator: Sven Faugert, sven.faugert@faugert.se

Utvärderingssystem i Sverige – till vad och för vem?
Eva Björklind, eva.bjorklind@mdh.se

Uppdrag granskning – redovisning av erfarenheter från metautvärdering av Storstadssatsningen.
Helene Lahti Edmark, helene.lahti_edmark@soch.lu.se och
Joakim Tranquist, joakim.tranquist@arbetslivsinstitutet.se

P 2:1 Metod II
Moderator: Kari Jess, kari.jess@mdh.se

Platsbesök som metod vid utvärdering av utbildning
– vilken information kan ett platsbesök egentligen ge?
Britta Lövgren, britta.lovgren@hsv.se

Evaluation workshops – practice-based development in health care and social services.
Per-Åke Karlsson, per-ake.karlsson@hb.se

Saknas Utvärdering av flexibel utbildning.
Eva Åström, eva.astrom@hsv.se,
Magnus Johansson, magnus.johansson@hsv.se m.fl.

P 2:2 Villkor och värderingar II
Moderator: Marja Lemne, marja.lemne@statskontoret.se

Dags att överge organisationsmetaforen?
Oscar Öquist, oscar.oquist@skolverket.se

Begreppen utvärdering, politik och forskning med utgångspunkt i värdebegreppet.
Elisabeth Andersson, elisabeth.c.andersson@mdh.se

P 2:3 Lärande eller ansvarsutkrävande utvärdering
Moderator: Lennart Widell, lennart.widell@esv.se

Utvärdering som verktyg för lärande.
Berit Goldstein, berit.goldstein@zeta.telenordia.se

Programlogikmetoden som ansvarsutkrävande redskap.
Gustav Petersson, gustav.petersson.9403@student.uu.se

P 3:1 Nationell utvärdering, system och ansvar I
Moderator: Joachim Schäfer, joachim.schafer@av.se

The design of evaluation systems.
Kim Forss, kim.forss@mailbox.swipnet.se

Produktion, beställning och användning av utvärderingar i den politiska beslutsprocessen.
Åsa Sohlman, sasohlman@privat.utfors.se

Storstadssatsningens utvärderingssystem.
Evert Vedung, evert.vedung@ibf.uu.se och
Nils Hertting, nils.hertting@ibf.uu.se

P 3:2 Utvärdering ur EU-perspektiv
Moderator: Per-Åke Karlsson, per-ake.karlsson@hb.se

Projektstrukturering och utvärdering av EU-projekt.
Veikko Pelto-Piri, veikko.peltopiri@karlskoga.se

Utvärderingspraxis i EUs socialfonder.
Kicki Stridh, stridh@interkomp.se

P 3:3 Roller, interaktion och inflytande
Moderator: Ingalill Ekelund-Nordenmark, ingalill.ekelund.nordenmark@mdh.se

Metaforiska dilemman i den interaktiva utvärderarrollen.
Joakim Tranquist, joakim.tranquist@arbetslivsinstitutet.se

Utvärderaren, de utvärderade och den värderande texten.
Verner Denvall, verner.denvall@soch.lu.se

P 4:1 Nationell utvärdering, system och ansvar II
Moderator: Tony Angleryd, tony.angleryd@riksrevisionen.se

Saknas Om att använda utvärderingar
eller inte i försvarssektorn.
Marja Lemne, marja.lemne@statskontoret.se

Utvärderingssystem för FoU-finansiärer?
Sven Faugert, sven.faugert@faugert.se
Inga Hallgren, inga.hallgren@faugert.se och

Johanna Nählinder, johanna.nahlinder@faugert.se

Evaluation of interventions in the government
sector in Sweden: a need for change.
Ove Eriksson, ove.eriksson@av.se

P 4:2 Aktörsmöten, samarbete och prioriteringar
Moderator: Annika Nordlander-Finn,
annika.nordlander-finn@riksrevisionen.se

Utvärdering av samarbete mellan
kommun och landsting.
Susanne Carlson, susanne.carlsson@phs.ki.se
och Julia Jäder, julia.jader@mdh.se

Samhällsekonomisk utvärdering av socialt arbete – är det möjligt och önskvärt?
Kari Jess, kari.jess@mdh.se