Evaluations That Matter in Social Work
Örebro Studies in Social Work 19

ANNA PETERSÉN

Evaluations That Matter in Social Work
Abstract


A great deal of evaluations are commissioned and conducted every year in social work, but research reports a lack of use of the evaluation results. This may depend on how the evaluations are conducted, but it may also depend on how social workers use evaluation results. The aim of this thesis is to explore and analyse evaluation practice in social work from an empirical, normative, and constructive perspective. The objectives are partly to increase the understanding of how we can produce relevant and useful knowledge for social work using evaluation results and partly, to give concrete suggestions on improvements on how to conduct evaluations. The empirical data has been organised as four cases, which are evaluations of temporary programmes in social work. The source materials are documents and interviews. The results show that findings from evaluations of temporary programmes are sparingly used in social work. Evaluations seem to have unclear intentions with less relevance for learning and improvement. In contrast, the evaluators themselves are using the data for new purposes. These empirical findings are elaborated further by using the knowledge form phronesis, which can be translated into practical wisdom. The overall conclusion is that social work is in need of knowledge that social workers find relevant and useful in practice. In order to meet these needs, researchers and evaluators must broaden their knowledge view and begin to include practical knowledge instead of solely relying on scientific knowledge when conducting evaluations. Finally, a new evaluation model is suggested. It is called phronesis-based evaluation and is argued to have great potential to address and include professionals’ praxis-based knowledge. It advocates a view that takes social work’s dynamic context into serious consideration and acknowledges values and power as important components of the evaluation process.

Keywords: evaluation, evaluation use, knowledge use, phronesis, scientific knowledge, praxis-based knowledge

Anna Petersén, School of Law, Psychology and Social Work. Örebro University, SE-701 82 Örebro, Sweden, anna.petersen@oru.se
Acknowledgements

Frodo took on the mission to carry the ring to Mordor. I set out to write a thesis. Before Frodo took on the mission, the discussion sounded something like this:

“Okaaaay, here’s a beautiful ring, but that’s just on the surface. In the end it will kill you. Who can carry it to Mordor and destroy it?”

Silence. Then an intense discussion on who is the most appropriate ring-bearer. Gandalf interrupts the discussion and points at Frodo as their only option. It turns out that no one else can carry the ring.

Frodo reacts with surprise, horror, and a little bit of pride. He accepts the mission to carry the ring to Mordor, to become the ring-bearer. He receives quite an okay back-up team, consisting of among others, fun characters, evil characters, smart characters, and dumb characters. Then he sets out on his journey to Mordor. The journey offers, if nothing else, at least variation. In the beginning, it is challenging and fun, and in the end it is only a matter of survival. And it is only with their last bit of strength, that Frodo and his friend Sam climb the mountains of Mordor and throw the ring into the fire. Let me help you out here, the fire = the dissertation act. But before that, when the journey was at its worst and the burden of the ring was so heavy, do you remember what Sam says to Frodo?

“I can’t carry it for you, but I can carry you!”

No one said those words to me. But it doesn’t mean that I haven’t been carried. Because I have. First of all, I would like to express my deep and sincere gratitude to my supervisor Jan Olsson. You were one of the smart and nice characters in my team, and I have always felt I had your support. I can honestly say that without your support, it would not have been any thesis at all. Instead, I would have thrown myself into the fire of Mordor. Thank you for being my co-author and thank you for all the coffee and chocolate. Especially the chocolate. I would also like to thank my other supervisor, Anders Bruhn, who joined my team later than Janne. First, I considered you a little bit as Boromir, just another guy in the team that had to be there due to some strange obligation (like, let’s say professor in social work). But I was wrong, you have been a dedicated supervisor whose input and comments I have valued a lot. I look forward to our future collaboration. Kari Jess and Odd Lindberg – thank you both for reading the final draft(s). Kari, an extra thank you for helping me believe in myself, and Odd
an extra thank you for always being brutally honest. Thank you Verner Denvall for your valuable comments at my final seminar. A particular thank you to Jürgen Degner for reading and commenting, even though not getting a penny for it. You saw my desperation and understood that I needed an extra helping hand. I’m sad to say that your stapler has left you for good, but don’t blame yourself – it wasn’t you, it was her. I also want to thank Peråke, who has read and commented the method section, and who always has a reassuring smile on his face.

Thank you my fellow doctoral colleagues. You are the ones that make me enjoy everyday life at work: Anna, Karin, John, Louise, Daniel, Robert, Mathias, Sara T, Sara J and all the other doctoral students that have come and gone over the years. Maria Moberg Stephenson and Anna Meehan, thanks for helping me out when I did not understand the proof reader’s comments. Thank you Britt-Louise for letting me escape most of the meetings. I would like to send a hug to my colleagues and friends at semester one, who have been very patient while I have been writing and teaching at the same time. Thanks to the research administration staff, Mia and Kristina, and to the rest of my colleagues in social work and criminology. A special thanks and my sincere apologies to all of you who have read my sometimes pretty crappy drafts during the years! Of course, I will also take the chance to thank my respondents for letting me interview them.

Finally, my greatest thanks to the ones whose love have been carrying me through ups and downs in life. Erik, you have never complained, nor even sighed, when I have announced that I need to spend another weekend or night at work. Now, the time for working out harder than any time before has come. I know you have the elite seeded group in the Vasaloppet in the palm of your hand! Måns and Ellen, you haven’t exactly facilitated my struggle, but you give me such hope for the future! I love you kids, and I have tried my best for you. Thank you Kerstin and Roger, my parents in law, for your generous attitude and for always being there for our family when we have needed it. Finally, I would like to thank my Mom. From you Mom I have inherited stubbornness, perseverance, and a lot of ‘jävlar anamma’. Thanks for everything you have done and still are doing for us, we love you and we adore you!

I wanted to illustrate “to be an evaluator” or “to evaluate” by being half engrossed in the context being evaluated and half on the outside or above it. First, I thought of a swimmer diving into the water, but then the swimmer would be more or less completely under the surface and that was not exactly how I think about being an evaluator or to evaluate. Then, I thought about
when I have been running in a bog, where every step I have taken have made me sink deep into the ground. That is how I think about being an evaluator or to evaluate. Partly engrossed in the context and partly above it. Torgny Pettersson helped me take such a photo. You can find it on the book cover. I think Torgny did a great job.

Örebro, March 27, 2017
Anna
List of publications


Article I and II have been reprinted with permission from the copyright holders.
# Table of Contents

1. INTRODUCTION .................................................................................................................. 17
   Aim and research questions ......................................................................................... 22
   Disposition ..................................................................................................................... 22

2. EVALUATION CRASH COURSE ...................................................................................... 24
   Defining evaluation ........................................................................................................ 24
   Contextualising evaluation in New Public Management ............................................. 26
   Evaluation flashback ..................................................................................................... 29
   Evaluation and auditing in Sweden ............................................................................... 31
   Evaluation of temporary programmes and subsidies .................................................. 35
   Evaluation approaches and models .............................................................................. 37
      Variations in evaluation approaches ........................................................................ 37
   Evaluation use .............................................................................................................. 41
   Summing up .................................................................................................................. 43

3. KNOWLEDGE IN SOCIAL WORK ................................................................................... 45
   Aristotle’s knowledge triad ............................................................................................ 46
      Episteme .................................................................................................................. 47
      Techne ..................................................................................................................... 48
      Phronesis ................................................................................................................ 49
   Knowledge use in social work ....................................................................................... 51
      Social workers’ relation to knowledge from research .............................................. 56
   Summing up .................................................................................................................. 58

4. RESEARCH DESIGN AND METHODS ........................................................................... 60
   The empirical part of the thesis .................................................................................... 61
   Data ............................................................................................................................... 63
      The documents ........................................................................................................ 63
      The interviews ......................................................................................................... 66
      The interview guides ............................................................................................ 68
      Analysis of the empirical data ................................................................................. 69
   Reflections upon quality .............................................................................................. 70
   Ethical considerations .................................................................................................. 73

5. SUMMARY OF THE ARTICLES ..................................................................................... 76
   Article I ......................................................................................................................... 76
   Article II ....................................................................................................................... 78
   Article III ..................................................................................................................... 79
6. DISCUSSION ................................................................................................. 81
   The empirical perspective: Use and non-use ............................................. 81
   The normative perspective: A combination of knowledge forms .......... 85
   The constructive perspective: A phronetic knowledge strategy in social
   work evaluation practice ........................................................................ 87
   Incorporating a praxis-based vision of knowledge in evaluation .......... 87
   Doable evaluations and useable evaluations ........................................ 89
   Making evaluations matter in social work .......................................... 91
Final remarks ................................................................................................. 91
REFERENCES .................................................................................................. 93

APPENDIX I: Interview guide - the civil servants
APPENDIX II: Interview guide - the evaluators
List of figures and tables

Figure 1. House’s taxonomy of major evaluation models......................... 40
Figure 2. Social work governors.............................................................. 53
Table 1. The cases ............................................................................... 65
Table 2. The respondents ...................................................................... 68
1. Introduction

Almost a decade ago I finished my first evaluation together with a couple of colleagues. The evaluation was commissioned by a Swedish national authority and concerned the care of substance abusers. After delivering the report, I sat in my office in eager expectation: When would they start calling? Did they not want any ‘expert lectures’? Were they not anxious to know more about the conclusions in the report? Do I need to say that the telephone did not ring once? At least, not concerning that project. For an experienced evaluator, that would probably be ‘business as usual’. To me, as a novice, it was kind of a disappointment. Even though I kept my eyes open, I never read one sentence about the evaluation. As far as I know, neither the agency nor the government used the report in any manner. I do not even know if anyone read it. But I do know that the evaluation cost the Swedish taxpayers about 3 million SEK. My story is more than just a ‘funny’ anecdote; it is a very good illustration of evaluation practice in general. Afterwards, I found out that my experience is shared by several other evaluators around the world (see e.g. Alkin, 2013; Dahler-Larson, 2012; Henry & Mark, 2003; Lindgren, 2008; Patton, 2008). Evaluations just do not seem to be used as evaluators intend them to be.

One striking aspect of knowledge production in social work is the increase of knowledge that comes from evaluations. Therefore, it is of great importance that evaluations can provide knowledge that is both relevant and useful in practice. Much research on evaluation has shown that evaluation use generally is limited, and its impact is weak (see, e.g., Johnson, Greenseid, Toal, King, Lawrenz, & Volkov, 2009; Kirkhart, 2000; Patton, 2008). Questions on evaluation use in social work relate to the more comprehensive discourse on knowledge use in social work. Adopting knowledge from research and evaluation is a great challenge for social work (Gibbs & Gambrill, 2002; Nilsson & Sunesson, 1988; Rosen, 1994). There are a number of potential explanations for why social workers only partially consider such knowledge in their work. For instance, they may rather rely on personal experience, prefer to be guided by normative assumptions, or be unable to read and understand esoteric scientific texts (Herie & Martin, 2002). Literature also reveals that many professionals find research in social work

---

1 SEK represented approximately 0.113 USD in March 2017.

2 In this thesis, I use the terms ‘use’ and ‘utilisation’ interchangeably, just as other evaluation scholars in the literature do.
inadequate, unreliable, or even useless (cf. Petersén & Olsson, 2014; Thyer, 2001). Taken together, these explanations of poor research and evaluation use must be considered as shortcomings in the practice of social work, as well as in research and evaluation. From the point of view of evaluative inquiry, evaluators will need to meet these deficiencies by providing evaluation findings that can contribute with knowledge that social workers find both relevant and useful in their work. Moreover, it must be done in ways that guarantee that professionals can make use of it. That is easier said than done.

In the last couple of decades there has been an evaluation boom in society (Dahler-Larsen, 2012; Lindgren & Johansson, 2013). Evaluation has taken place as an organisational phenomenon that has become institutionalised in the public sector. It operates on all levels, from formal political practices to local ones (Taylor & Balloch, 2005). This institutionalisation is a part of what some have called the audit society, wherein monitoring activities such as evaluation, revision, and supervision are continuously increasing (Power, 1997). To adapt to this audit society, Sweden produces large quantities of evaluations. According to Forss (2007), the Swedish state administration provides approximately 2000 evaluations per year, and the local governments about 3000. This estimation does not include ongoing activities for delivering information for statistics, follow-ups, or quality indicators. There are several different authorities working exclusively with evaluation and supervision of the public sector, such as the Swedish Schools Inspectorate, the National Board of Health and Welfare, and the Swedish Work Environment Authority. Also the County Administrative Boards and local governments exercise auditing in different shapes and forms.

The quest for evaluation has become widespread across the public sector, and today evaluation is almost obligatory. One simply must evaluate, without further thoughts on why (Hertting & Vedung, 2009). There seems to be a mood indicating consensus that evaluations work as rational tools for a number of purposes. One may evaluate for measuring and accountability, determining efficiency, uncovering explanatory insights into social and public problems, understanding learning and transferability, and increasing agency responsiveness to the public (Chelimsky, 1997; Krogstrup, 2006), to mention a few purposes. In the light of comprehensive demands and high expectations of evaluative inquiry, researchers describe how evaluation as an activity has come to be taken for granted (see Dahler-Larsen, 2012; Lindgren, 2008; Rabo, 2006). Evaluation is thus seen as a ritual, which means
collecting data without any explicit purpose except to act as a modern organisation. Organisations are in general expected to act like this, keeping up with the rest of the audit society, because it is the appropriate behaviour today (March & Olsen, 1989, 1995). Dealing with evaluation ritually appears as a waste of time, money, and other resources. It also undermines the potentially good qualities of evaluations as contributors to learning and improvement. Nevertheless, the evaluation boom has more or less spread all across the public sector. In human service organisations, addressing citizens’ wellbeing, it ought to be of great importance to judge and value different interventions. Evaluation may contribute to improvement and effectiveness in practice; enhance professional moral purpose, progress in human service delivery, and responsiveness to service users; and fill other knowledge gaps (Blom, Morén, & Nygren, 2011; Lishman, 1999; Shaw, Mowbray, & Quereshi, 2006). Evaluations have, along with the increasing part of temporary programmes, received an additional task. Such programmes have become more common in social work since the 1980s, and they are characterised by being financed for a specific purpose during a limited time (Montin & Hedlund, 2009). They almost always end with an evaluation. As they usually are very costly, so are the evaluations. The programmes are intended to work as a catalyst for change and learning, which is supposed to live on after the programme has been terminated (Statskontoret, 2007). This is one strong motive for evaluating these programmes (Petersén & Olsson, 2014).

The contradiction between high demands for evaluation and doubtful ideas on how to make use of the results is remarkable. Evaluation use has been a subject of controversy for a long time now (Cook, 1997). Previously, evaluators anticipated that the results would be put into practice as an immediate consequence of the reporting of the evaluation. Such delusions have had to be abandoned, and most researchers consider the assumptions behind instrumental use unrealistic (Rossi, Freeman, & Lipsey, 1999; Weiss, Murphy-Graham, & Birkeland, 2005). Decisions do not seem to be made easily in politics. Besides that, Beck (1992) stresses that peoples’ trust in scientific knowledge has decreased along with an increasing mistrust of institutions and professionals. To compensate for such a development, institutions try to standardise their work and regulate the professionals. It is furthermore a way to seek control and minimise risks. Power (1997) calls this process ‘surrogate trust’. All in all, the conditions for evaluation use seem flawed.

Questions on use of evaluation relate to the more comprehensive discourse on use of knowledge in social work. Historically as well as currently,
adopting knowledge from research and evaluation is contested in social work. The field has been described as suffering from a long-standing gap between practice and research (Longhofer & Floersch, 2012; Nilsson & Sunesson, 1988; Taylor & White, 2005). Much literature has shown that social workers seem to have a half-hearted interest in considering scientific knowledge, that is, knowledge produced from research, in their work. At the end of the 20th century and the beginning of the 21st century, the Swedish National Board of Health and Welfare heavily criticized the working routines of the social services. This criticism consisted of arguments such as that social workers were relying on methods without scientific evidence, and that they lacked competence and knowledge to conduct their work professionally (Socialstyrelsen, 2000:12). Moreover, Swedish researchers found that social workers neither read the professional literature nor were they up to date with the literature available (Bergmark & Lundström, 1998, 2000, 2002; Hansson, 2003; Nilsson & Sunesson, 1988). Across the sea, researchers have noted a similar attitude towards scientific knowledge among American social workers (Rosen, 1994; Rosen, Proctor, Morrow-Howell, & Staudt, 1995). Instead of taking research findings into consideration, American social workers seemed to prefer to be guided by normative assumptions. In Britain, researchers have discovered social work to be influenced by intuition, values, and experience rather than research (Trinder, 2000), and similar tendencies has been found in Sweden as well. The Swedish social workers claimed that their own and their colleagues’ experiences were their most important source of professional competence (Bergmark & Lundström, 2000). Thereto, Ramsey, Carline, Inui, Larson, LoGerfo, Norcini, and Wenrich (1991) found that professionals’ knowledge use seems to decrease the more work experience they gain. However, more recent research has demonstrated a change of attitude towards research. An increased number of social workers are reporting that they are using professional literature (Bergmark & Lundström, 2007, 2008). This may be an effect of the massive efforts of implementing evidence-based practice (EBP) in social work.

There are several other possible explanations for why social workers only partially consider research in their work. Some of them may view their work as resembling an art (Schön, 1995), and therefore find research of limited importance. Others may be prevented from utilising scientific texts because of the texts’ linguistic inaccessibility (Herie & Martin, 2002). Another – contrasting – perspective on the limited use of scientific knowledge turns
things around: research and evaluation are focused on aspects that professionals find irrelevant, unreliable, or useless. Thyer (2001) criticizes research in social work in the USA as being too theoretical, rather than providing knowledge about the effects of different interventions. Above all, Thyer argues that researchers in social work should spend time on conducting systematic reviews of already existing research and evaluation projects. Representatives of the Swedish National Board of Health and Welfare have raised similar criticisms (see Tengvald, 1995, 2003). In common with those who condemn traditional social work research is their allegiance to the movement promoting EBP. EBP is often described as a way forward for professionals in social work to incorporate scientific knowledge in their work (cf. Svanevie, 2011, 2013). Despite this good intention, the evidence movement\(^3\) has not stood unchallenged. Antagonists claim the evidence movement’s epistemological assumptions to be unrealistic and not of relevance for practical social work (see, e.g., Måansson, 2000, 2001; Reid, 2001; Long, Grayson, & Boaz, 2006): EBP reduces social problems to measurable variables (Simons, 2004), it regards social problems from a sharper medical angle (Oscarsson, 2009), it devalues non-experimental designs (Denvall & Johansson, 2012), and it gives no or limited answers as to why effects arise (Blom, 2009). Additionally, Herz (2012, 2016) stresses that EBP passivates social workers by limiting their motivation to reflect upon their practice and decisions. In all, there are some really strong arguments that EBP, whose purpose is to inform practice with research, cannot solely bridge the gap between practice and research.

In summary, large numbers of evaluations are commissioned and conducted every year. This is partly a consequence of the high pressure of auditing in the public sector. Evaluations can, as best, provide knowledge for change and improvement. However, some research has shown that evaluation findings are not used at all. This can be explained by the fact that evaluations also can have other functions, such as being commissioned and conducted as a matter of routine, or as parts of the political game. This may be especially relevant concerning evaluations of temporary programmes, which

\(^3\) The movement promoting evidence-based practice (EBP) is, in somewhat simplified terms, the group of people representing the opinion that social work practice should be evidence-based and that research and evaluation in social work should focus on providing knowledge (evidence) about different interventions’ effects. Furthermore, the movement relies upon the idea of systematic reviews (Foss Hansen & Rieper, 2009, p. 141). Within this thesis, I call the movement ‘the evidence movement’, but in article II, I write ‘the evidence-based movement’ (see Petersén & Olsson, 2015).
almost always include an evaluation component. There simply are no further thoughts on how to use the results. In social work, the non-use can also be explained by referring to deficiencies in the social workers’ use of scientific knowledge. It has been argued that social workers do not search for, or read, scientific literature and that they instead rely upon other knowledge sources such as experience. Another explanation addresses weaknesses in the production of knowledge, which means that research and evaluation focus on things that the professionals find irrelevant, unreliable, or useless. 

Thus, there is a complex problem of the use of evaluation in the evaluation field, which is particularly challenging to social workers who generally are criticized of not using scientific knowledge. To improve evaluation practice in social work, all these problematic aspects need to be taken into consideration.

**Aim and research questions**

With departure in holding evaluation as an important, relevant, and potentially useful tool for enhancing knowledge production in social work, the aim of this thesis is to explore and analyse evaluation practice from an empirical, normative, and constructive perspective. The objectives are partly to increase the understanding of how we can produce relevant and useful knowledge for social work using evaluation results, and partly to give concrete suggestions on improvements on how to conduct evaluations. The research questions posed are as follows:

1. What kind of evaluation use can be identified in the field of Swedish social work, with particular focus on evaluations of temporary programmes, and what affects the different modes of use? (the empirical perspective)

2. In what ways can different forms of knowledge enhance knowledge use in social work? Which kind of knowledge should preferably guide social work? (the normative perspective)

3. How can we, through evaluations, best provide relevant and useful knowledge in social work? (the constructive perspective)

**Disposition**

The remainder of this thesis is structured as follows. Two theoretical chapters follow this introduction. In chapter 2, the field of evaluation is presented. The chapter aims to contribute with a so-called crash course in eval-
evaluation, starting with defining the concept and contextualising the phenomenon as a part of the administrative trend New Public Management. Thereafter follows a short summary of evaluation history, both internationally as well as in Sweden. Since temporary programmes play an important role in this thesis, I devote a section to dealing with evaluations of such programmes. At the end of the chapter, I present different ways to organise evaluations through approaches and models, concluding with a section about evaluation use.

Chapter 3 explores the topic of knowledge in social work, starting with a brief introduction of how knowledge from research has become a central element in the field. Early in the chapter, I present Aristotle’s knowledge triad, consisting of episteme, techne, and phronesis, as a point of departure for further reading. After that follows a number of sections concerning the relationship of social workers and other actors in social work to knowledge gained from research and practice. Evidence-based practice is presented as an important part of the knowledge debate.

In chapter 4, the research design and the methods of the thesis are outlined. Methodological considerations, as well as the techniques for data collection employed, are described. The empirical foundation consists of four cases of different large-scale evaluations of temporary programmes in Swedish social work. Within these cases, eleven people working as higher civil servants and evaluators were interviewed. The chapter concludes with a discussion about the quality of the study and ethical considerations.

Chapter 5 includes a summary of each of the three articles that are included in this compilation thesis. Briefly, article I is an empirical study of evaluation use, based upon the four cases mentioned above. Article II is a theoretical study and deals with the nature and quality of applied knowledge from research and evaluations. The third and final article presents a new evaluation model called phronesis-based evaluation. The article aims to provide new forms of evaluation in social work that will benefit learning and improvement in practice.

Chapter 6 concludes this thesis. In the beginning of the chapter, I first and foremost discuss the empirical aspects, with focus on article I. Further on, the normative and constructive perspectives on evaluation in social work are elucidated.
2. Evaluation Crash Course

In this chapter, I will provide an intense walkthrough of the concept of evaluation from an international and Swedish perspective. I briefly outline the history of evaluation, evaluation of subsidies and temporary programmes, and different ways to organise an evaluation. At the end of the chapter, I will discuss evaluation use, which is a central theme in this thesis. However, I start with defining evaluation, and then contextualise it as a phenomenon within New Public Management.

Defining evaluation

Evaluation – more than any science – is what people say it is; and people currently are saying it is many different things (Glass & Ellett, 1980, p. 211, in Shadish, 1998, p. 9).

There are authors within the literature who refer to God as the first evaluator ever, when, in Genesis 1, God “saw everything he has made, and behold, it was very good” (Genesis 1:31). God made a summative self-evaluation, so to speak. However, we never really responded to God until recently, and evaluation is considered to be a young discipline and an ambiguous concept.

Internationally, Michael Scriven’s definition of evaluation is one of the most cited (see e.g. Chelimsy & Shadish, 1997; House, 1993). It reads as follows: “Evaluation refers to the process of determining the merit, worth, or value of something, or the product of that process” (Scriven, 1991, p. 139). ‘Merit’ refers to the qualities of the evaluated object, and ‘worth’ refers to its meaning and utility within the system. Other definitions of evaluation sometimes include aspects of systematics. Carol Weiss is one of those researchers emphasising this: “Evaluation is the systematic assessment of the operation and/or the outcomes of a program or a policy, compared to a set of explicit or implicit standards, as means of contributing to improvement of the program or policy” (Weiss, 1998, p. 4). Just as Weiss does at the end of her definition, Michael Quinn Patton touches upon the purpose of evaluation when declaring his view: “Program evaluation is the systematic collection of information about activities, characteristics, and outcomes of programs to make judgements about the program, improve program effectiveness, and/or inform decisions about future programming” (Patton, 1997, p. 23). Eleanor Chelimsky (2006) offers a simple explanation for the different purposes of evaluation, which include ensuring accountability, or
generate knowledge, or improving organisations. Evaluation for accountability aims to hold key actors, such as policy-makers and programme managers, responsible for the outcome of the policy or programme. Evaluation for generating new knowledge targets critical reviews and more profound understandings of activities and problems, whilst evaluation for improvement aims to provide learning and improvement within organisations.

Swedish researchers also offer a plethora of definitions of evaluation (see e.g. Blom, Morén & Nygren, 2011, p. 18; Eriksson & Karlsson, 2008, p. 27ff; Jerkedal, 2010, p. 19ff; Karlsson Vestman, 2011, p. 24ff; Lindgren, 2012, p. 89). These definitions have in common that they all – more or less – reconnect with Evert Vedung’s definition. Vedung has been called the best-known evaluation scholar in Sweden (see e.g. Tranquist, 2015), and I have not found any Swedish book on evaluation from the latter part of the 1990s that does not mention or copy his definition. His definition is as follows: “Evaluation – df. careful retrospective assessment of public-sector interventions, their organisations, content, implementation and outputs or outcomes, which is intended to play a role in future practical situations” (Vedung, 2009, p. 3). He calls his definition “controversial” (p. 3) and explains its narrowness by claiming that evaluation has become such a popular activity that its meaning has broadened to a point where it has become relatively meaningless, as it can be applied to almost anything. Even the slightest effort can be defined as evaluation today (which makes me feel bad for referring to God as the first evaluator just for looking back at his own work). Kettil Nordesjö (2015) writes in his dissertation about evaluation as a contested concept (p. 31-33). A contested concept is characterised by lacking an agreed definition (Gallie, 1955), but in contrast, different groups and individuals may hold their own definitions of it. Such definitions usually start with how the concept is used by the specific actor(s) and whether they are aiming for a subjective or objective definition. Among other things, Nordesjö argues that evaluation should not be defined too narrowly as it would then risk losing value as the lowest common denominator. Evaluative inquiry could also include solely descriptive and non-explaining results. That kind of reasoning supports Vedung’s thesis about protecting the concept of evaluation and preventing it from becoming a semantic magnet (cf. Lundquist, 1976). Later in this chapter we will take a closer look at different variations of evaluation.

Evaluation is adjacent to research, but it is not the same. Several evaluation scholars emphasise the importance of being scientifically skilled in or-
der to be able to make accurate judgements (Chelimsky, 1985). That is similar to ordinary research. But, evaluators are tied to a commission, while, ideally, researchers are free to conduct any research. Research can be about describing, understanding, and explaining something without placing a value on the final results. Evaluation, on the other hand, is about providing judgements about an existing activity while it is ongoing or terminated (for a more detailed discussion on the differences between evaluation and research, see Coffman, 2003/2004; Karlsson Vestman, 2011; Rombach & Sahlin-Andersson, 2006).

In sum, the different definitions of evaluation diligently repeat three basic components: evaluation is the systematic and precise collection of knowledge; evaluation entails making a judgement; and evaluation aims to have some kind of impact. These three are, according to me, important components as they address the scientific element in evaluation of a systematic and precise course of action. What separates evaluation from research is that within evaluation one needs to make judgements, and that the results are supposed to have impact in practice. The results are not just aimed to become a contribution to the general knowledge base.

**Contextualising evaluation in New Public Management**

The demands of auditing and controlling public service have increased continuously over the last decades (Hood, 1991, 1995; Hood, James, Jones, Scott & Travers, 1999; Hood, James & Scott, 2000; Kitchener, Kirkpatrick & Whipp, 1999; Power, 1997). This means that monitoring activities such as evaluation, follow-up, supervision, and revision has become more and more important. This development is not particular for Sweden, even though Sweden has proven to be more receptive than many others; similar tendencies have also been acknowledged internationally. The ideas of this administrative trend in society, called New Public Management (NPM), originated from Thatcher’s and Reagan’s management in England and the USA (Ahlbäck Öberg & Widmalm, 2016). Since the late 1960s, it seems like the critique against the classical bureaucratic model for management has increased. Given the postwar expansion of the public sector, the model was considered insensitive and difficult to use for governing (Tarschys, 1983). Moreover, the public sector in itself was considered ineffective, being of low quality and poorly adapted to cater to service users’ needs. The critique against the governing of the public sector continued to grow during the 1970s and the 1980s, and both the right and the left wings joined the lament. Whilst the right wing wanted to cut government spending and make
the public sector more effective, the left wing wanted to democratise it as a way to improve service and to provide greater opportunities for public influence. Researchers in political science later stressed that changes in the public sector were necessary, and not just the outcome of public discontent. Due to the financial crisis, the detailed management, and the fact that most of the public agencies were financed by allocations without transparency of what they effected, the Social Democrats in Sweden started to approach the ideas of NPM (Ahlbäck Öberg & Widmalm, 2016).

NPM is usually described as consisting of a cluster of ideas retrieved from the private sector which challenge the traditional government models in favour of market and businesslike principles (Aldridge, 1996). The common catchphrase of NPM is that public service is expected to give more ‘value for money’ (Christensen & Lægreid, 2007; Hood, 1991). The reforms have been more or less implemented in every conceivable field of public life, from government machinery to personnel practices (Hughes, 2012). Briefly described, the ingredients of NPM are steering and control, disaggregation and competition, management and greater dominance, citizens’ and customers’ freedom of choice, and a new language (Abrahamsson & Agevall, 2009; Agevall, 2005). Steering and control are two ingredients that focus on improved effectiveness achieved through stricter control, goal steering, performance-based incentives, benchmarking and, not least, evaluation and monitoring (Agevall, 2005; Almqvist, 2006). Disaggregation, or fragmentation, refers to the separation between clients/customers and performers. The division between them is supposed to induce competition among probative actors, which in turn will lead to increased effectiveness. Furthermore, the ingredients management and greater dominance are more or less at opposite poles arguing for both centralisation and decentralisation (Christensen & Lægreid, 2007). First, it means that politicians will have more power over civil servants and the public administration, due to their role as rulers. Second, it means that managers in public administration should have more freedom to take decisions and concrete actions. At the same time, the citizens and customers of the public sector will have increased power through freedom of choice– they will be able to choose between different alternatives among welfare services. The final ingredient of NPM is a new language. Agevall (2005) writes that, along with the establishment of NPM, many terms within public administration have been replaced by synonyms from the private sector. Higher civil servants have become managers or directors and citizens have become customers.
As already mentioned, the public sector used to be governed through allocations, which meant that the agencies annually applied for public funds in order to carry out their functions. Then, the requirements of any reporting back were minimal, if not non-existent. However, along with the emergence of NPM, the new forms of governing came to focus much more on results, which in turn increased the demands for auditing and monitoring. Michael Power (1997) described this as the rise of the audit society, which has been established as an important concept for understanding the development. According to Power, the audit society risked working counterproductively in terms of cost efficiency, as auditing is an expensive activity. Ivarsson Westerberg and Jacobsson (2013) estimate its cost in Sweden to approximately 20 billion SEK every year, although that is a very rough figure (see also Forssell & Ivarsson Westerberg, 2016). Another investigation from 2007 estimates the total sum for public auditing to about 2-3 billion SEK (SOU 2007:75). One must also take into account that auditing costs for those who are becoming audited. These figures are not included in the numbers above. Power (1997) has criticised the methodology used when auditing the public sector, since he claims that we only measure what we are able to measure, and preferably measure easily. For example, the quality at a social welfare office can be measured by examining how many service users with substance abuse have become sober within a specified period. As long as there is agreement on what counts as ‘sober’, it is easy to measure just by counting. But, professionals and service users probably have completely different views of what quality at a social welfare office really is. They would probably also mention the meaning of relations, personal treatment, and participation, just to mention a few likely aspects. As for the discussion of measuring what is easy and possible to measure, we can acknowledge a second aspect, which is that agencies seem to try to adapt to the audit society by making themselves auditable in various ways (Ek, 2012; Lindgren, 2008). Some organisations put a lot of effort and financial resources into becoming auditable, which in turn can lead to a number of interesting behaviours. As a consequence, public agencies are now documenting and writing plans like never before.

All in all, auditing activities are an important element in public policy and management. Evaluation and similar forms of monitoring have become an institutionalised part of public service, but due to a narrow view of how measurement should be done, we can assume that NPM has paved the way for certain kinds of evaluations.
Evaluation flashback

Evaluation, as currently understood, did not develop until the 1960s when the American President Johnson tried to fight poverty (see e.g. Lundahl & Öquist, 2002; Wildavsky, 2007). Compared to auditing, which can be traced to the mid-1800s (Oldrup, 2013), evaluation is a young discipline (see Åberg, 1997, pp. 21-23 for a detailed historical overview) and there is no consensus about whether evaluation should actually be described as an autonomous discipline yet (see e.g. Andersson & Karlsson, 2003). The debate about how to define a discipline can be read in Toulmin’s writing from 1972, which House (1993) refers to when he carefully leans towards calling evaluation a discipline in its own right. Evaluative inquiry is characterised by engaging researchers from various other disciplines, with different scientific training. Scriven (2003) describes evaluation as transdisciplinary and predicts that in the future it will belong to an elite group of disciplines that have a ‘service function’ in relation to other disciplines. In other words, evaluation supplies other fields with tools whose subject matter is “merit/worth/significance” (p. 28). Even though the task of evaluation is to support others, it also generates research efforts of its own.

The history and development of evaluation have been illustrated in many ways, making reference to generations (Guba & Lincoln, 1989), waves (Vedung, 2010) and trees (Alkin, 2013; Carden & Alkin, 2012). Each author tries to capture the different trends in evaluation practice. Much evaluation history, here in this thesis as well, is described from the point of view of American writers. The USA has been, and in a way still is, the leader in evaluative inquiry. Development in the USA and in Sweden has been rather similar (Eriksson & Karlsson, 2008; Sandberg & Faugert, 2007). The catalyst for the comprehensive changes in evaluative inquiry was the War on Poverty and the Great Society (Wildavsky, 2007). President Johnson’s agenda embraced visions of a more equal society, thus major efforts were made in massive educational programmes in a number of sectors (Fitzpatrick, Sanders & Worthen, 2004). The phenomenon was exported from the USA to a number of countries, including Sweden (Ivarsson & Salas, 2013). To make a long and complicated story short, as a result of this expansion, interest in evaluation grew. Soon, evaluations began to be supplemented with analyses of causes and effects. In other words, along with the development of the welfare society, the appreciation of radical rationalism rose (Vedung, 2010). In those days, evaluative inquiry and rationalism seemed to be a match made in heaven. By that time, a positivistic understanding of science dominated (Albæck, 2001; Cook, 1997). Qualitative methods were hardly
even mentioned, and quantitative methods were assumed to be better suited for revealing causal consequences of a programme. Thus, experimental designs were considered superior to other designs, and qualitative methods were ranked lower than any statistical control (Cook, 1997). Vedung (2010) refers to this period as the science-driven wave.

But the times were changing and soon qualitative approaches came to challenge quantitative ones. In the middle of the 1970s, the strong belief in experimental evaluation began to fade (Vedung, 2010). At the same time, evaluators started to organise networks, develop programmes to train students, and also to publish scientific evaluation literature. The previous relation with federal governments got weaker and evaluative inquiry became self-sufficient. Still, the quest for evaluation increased throughout the 1970s. The dominance of the positivistic approach waned and scholars came up with new models. Many of these are still used today, such as Scriven’s goal-free evaluation (1973), Stufflebeam’s CIPP model (1971), Stake’s responsive evaluation (1975), and Guba and Lincoln’s naturalistic evaluation (1981). A later example of a new model is realistic evaluation, which is represented by Pawson and Tilley (1997; Pawson, 2013), and Kazi (2003) for instance. Many researchers argued for more pluralistic approaches in evaluation, involving a larger number of participants than before (Karlsson, 1995). Guba and Lincoln (1989) even took the development one step further by declaring a new paradigm called the constructive paradigm. The constructive paradigm, also called the hermeneutic or naturalistic paradigm, had a major impact in the evaluation circuits. Many evaluation scholars appreciated its ontological (reality is a social construction), epistemological (findings exist because the interaction between observer and observed creates what emerge from that inquiry), and methodological (hermeneutic-dialectic process) standpoints (Guba & Lincoln, 1989; Vedung, 2010).

In the late 1970s and the beginning of the 1980s, a transformation of the public sector started to take place. Society faced new types of business-inspired reforms, often under the label of NPM (Hall, 2012; Hood, 1991), which I have already described in some detail in this chapter. After about ten years of lagging behind other countries, Sweden adopted NPM as well. Within NPM, evaluation was established as a permanent function in the public sector. Perhaps it was also during this era that the concept of evaluation started to erode. Evaluative inquiry took the form of accountability assessments, performance measurement, and consumer satisfaction appraisal.
Evaluation practice in its current form is affected by the evidence movement, which took off in the 1990s. In Sweden, evidence-based practice (EBP) started to grow along with the development of NPM. EBP and NPM have some important traits in common, such as their top-down rationalistic view and acknowledgement of evidence and general techniques (Petersén & Olsson, 2015). The success of EBP meant a revival for experimental designs within evaluation. As a part of the evidence movement, two international cooperation bodies support research and evaluation through experimentation: the Campbell Collaboration and the Cochrane Collaboration. From a wider point of view than social work, the purpose of EBP is to make government more scientific and to base government activities on empirical evidence. Here, of course, state-initiated evaluations play an important role. In Sweden, alongside the evidence wave, is the fifth generation of evaluation, wherein evaluation is discussed in relation to learning (see Sjöberg, Brulin & Svensson, 2009). Learning evaluation has a formative approach and usually includes elements such as reflection and participation (Claesson, 2015).

Summing up, evaluation has undergone major changes since the 1960s. It has transformed from:

monolithic to pluralist conceptions, to multiple methods, multiple measures, multiple criteria, multiple perspectives, multiple audiences, and even multiple interests. Methodologically, evaluation moved from a primary emphasis on quantitative methods, in which the standardized achievement test employed in randomized experimental control group design was most highly regarded, to a more permissive atmosphere in which qualitative research methods were acceptable. (House, 1993, p. 3)

Most significant in this transformation is probably the shift in methodological focus from a fundamental belief in experimental methods to an emphasis on qualitative methods and case studies. Even if there is currently an ongoing evidence wave (Vedung, 2010) which makes us turn back to experimental designs, there are also strong opposing forces that have significant implications for the practice of evaluations.

**Evaluation and auditing in Sweden**

In the light of the international review of evaluation history, we now turn specifically to the evaluation tradition in Sweden. In many respects, the Swedish history is similar to the American one, and in the preceding section I noted some parallels between the USA and Sweden. Sweden was one of the first countries to embrace the ideas of evaluation (Hansen & Hansen, 2000),

ANNA PETERSEN  *Evaluations That Matter in Social Work*  31
along with the USA, Canada, the Netherlands, the UK, and Germany (Eliasson & Sunesson, 1990; Furubo & Sandahl, 2002; Vedung, 2004). The first efforts of evaluation in Sweden were made within pedagogy and aid for developing countries in the 1960s (Denvall, 2011; Lindgren, 2008). In Swedish social welfare, Robert Bell (1975) became a pioneer in evaluative inquiry with his thesis in pedagogy: To Evaluate Social Programs (my translation). A funny anecdote, which also reflects yesterday’s opinions about the term ‘evaluation’, involves the late Professor Harald Swedner’s critique of Bell’s phraseology (see e.g. Bell, 1975; Vedung, 2009). Swedner found the word evaluation “clumsy” and “misleading” (Bell, 1975, p. 225). This shows that there was no consensus on how to define evaluation at that time in Sweden. Evaluations of social reforms and programmes came into practice in the middle of the 1970s (Nilsson, 1993). Characteristic for Sweden is that evaluative inquiry has evolved in parallel in different fields. As has been noted, pedagogy has a rather long tradition of evaluation, but so does medical science, although it does not always explicitly mention evaluation. Instead, medical researchers talk about systematic studies of interventions. In the 1970s, treatment planning became an important part of social work. These plans for how to treat clients were supposed to be evaluated at a later stage, however, Eriksson and Karlsson (2008) make a sophisticated guess that just that particular part probably got neglected. In the 1980s, Sweden was affected by the introduction of NPM (see section Contextualising Evaluation within New Public Management in this chapter) and the eagerness to evaluate increased. During the 1990s the evaluative activities more or less exploded and new auditing methods for governing emerged and became institutionalised in the public sector. Examples of such methods are revision, accreditation, and certification (Power, 1997). Of course, these existed earlier too, but they did not receive the level of attention they got later. All in all, the methods are a result of the increasing demands of intensive auditing in more and wider contexts. The rise of the audit society has been acknowledged by researchers in quite critical terms. Ryan and Schwandt (2002) writes about auditing as a “mania” (p. 175); Lindgren (2008) compares evaluation to an insatiable monster (p. 19); and Albæk (2001) asks his readers why we are so obsessed with evaluation (p. 32).

Auditing methods for governing are often described in terms of making the public aware of what happens within organisations, especially in the public sector. Tools for governing can, hopefully, contribute to the maintenance of democratic ideals by making activities, processes, and outcomes visible. At the same time, such tools may improve efficiency and encourage
development. All in all, the expectations on auditing methods for governing are high. They are supposed to be methods for control and assessment, as well as for advancement (Lindgren, 2011; Power, 2003; Rombach & Sahlin-Andersson, 2006). To keep up with the rest of the audit society, more and more government agencies (including local ones) organise their own departments for evaluation or procure these kinds of services with companies that mediate the market of evaluation (Lindgren, 2008). There are also a number of authorities that have been created for the purpose of conducting evaluations. As a consequence of the major interest in evaluation, there is now a range of educations and courses available for those who want to specialise in this field. There are also different associations and networks available to people who are particularly interested in evaluation, such as the Swedish Evaluation Society [Svenska utvärderingsföreningen] and the Evaluation Circle [Utvärderingsringen], and some universities in Sweden have special centres or research groups for evaluation. In addition, Örebro University, in cooperation with Mälardalen University, was the first to offer a master’s programme in evaluation in social work in 2008. Within the scope of such activities, there is an ongoing professionalisation process of evaluative inquiry.

The government ambitions to control and regulate public sector activities are recognised by several Swedish researchers (see e.g. Alexanderson, 2006; Claesson, 2015; Ek, 2013; Hämberg, 2013; Sahlin-Andersson, 2006). Professionals within the public sector experience pressure to document and measure their activities as a means to ensure quality and accountability. The adaptation makes them “auditable” (Power, 1997, p. 91), so to speak. Concrete examples of greater government control are the development of quality indicators such as Open Comparisons [Öppna jämförelser] (Swedish Association of Local Authorities and Regions, 2013). Open Comparisons is a system for following up on and comparing a selection of activities within social work and health care. The general enthusiasm for evaluation in Sweden can partly be explained by NPM, and for social work, also the introduction of EBP (Denvall & Johansson, 2012; Liljegren & Parding, 2010). In the 1980s, the Swedish government published several reports that pinpointed the public sector’s lack of productivity and efficiency, and at the same time service users’ weak positions vis-à-vis the service producers were recognised. The solutions NPM brought seemed to be able to solve both

4 The master’s programme of evaluation in social work is now terminated.
problems. When EBP was introduced in the 1990s, the demands of evaluating and following up on interventions in social work increased further. Today, professionals have to report on their work on a regular basis. The development has been moving from unsystematic and non-existent documentation and evaluation to becoming a common part of daily work.

Sweden has, in comparison with many other countries, been a part of the gigantic evaluation wave that started in the 1960s and has grown stronger and stronger. The development in Sweden is quite similar to that in the USA, which must be considered the world leader in evaluative inquiry. From the beginning, Sweden has lagged about ten years behind the USA, and the concept of evaluation did not even exist in the Swedish vocabulary before the 1970s. Nilsson (1993) discusses why Sweden did not take up the ideas of evaluation from the USA earlier. He highlights a number of reasons: first, Swedish reforms have often been initiated in response to the working class struggle for better living conditions. In comparison, many of the reforms in the USA were by that time a consequence of professionals’ and experts’ influence on the public sector. Therefore, the USA found it necessary to firmly establish the reforms in research and evaluation. Second, in Sweden we have a tradition of investigating policies before taking action. These investigations probably worked as prohibitive on evaluation, as representatives from the different political parties had already had their chance to assess and affect the reforms under the investigation. Moreover, local politicians in Sweden are rather involved in the public administrations and have thus had extensive control. Nilsson writes, somewhat pejoratively, that the USA has had more of a tradition of ‘launching’ political reforms rather than investigating them (p. 51). Third, the Swedish welfare state grew enormously during the period 1932-1976⁵, and probably many stones would have been left unturned if attempts had been made to evaluate all reforms during this period of expansion. Fourth, the social sciences entered the Swedish scene quite late and at that time they were not really established in many of the fields that are strong in evaluation today. Consequently, Sweden lacked research environments at the universities that could push issues concerning evaluation.

In sum, after a slow start, Sweden has embraced evaluative activities in every field in social welfare. Evaluation in Sweden, as in other similar countries, has been established as important for monitoring due to the develop-

---

⁵ This period is usually referred to as Folkhemmet (in English, People’s home).
ment of NPM. Today, there is an ongoing professionalisation among evaluators. Many evaluators have organised themselves in societies and research environments, and Sweden has also had its first masters education programme in evaluation. Evaluation is also of particular importance in Sweden, as here have been established a new type of governing characterised by temporary programmes and subsidies.

**Evaluation of temporary programmes and subsidies**

Temporary subsidies are a growing feature in several industrialised countries and are considered to be an important government tool (Montin & Hedlund, 2009; Salamon, 1989; Vedung, 2007). They have become a very important source of funding for reforms, particularly in social welfare and education. Such subsidies are supposed to be used for special purposes during a limited time, and the effects are intended to live on after the programme has been terminated. Eventually, the programmes are expected to contribute to long-term learning and knowledge that can be integrated into ordinary work (Montin, 2007; Montin & Hedlund, 2009). Leeuw (2007, p. 78-79) defines subsidies as:

> the conditional transfer of funds by government to (or for the benefit of) another party for the purpose of influencing that party’s behaviour with a view to achieving some level of activity or provision.

The government consequently does not make any efforts by itself, but influences the local municipalities to work in certain ways. The rationales behind subsidies are that increased incomes motivate people to react by undertaking those activities the subsidy provider suggests (Leeuw, 2007).

Subsidies have been established as one of the new forms of governance, along with other policy instruments such as information and agreements (Lascoumes & Le Gales, 2007). Public policy instruments can be defined as a “set of techniques by which governmental authorities wield their power in attempting to ensure support and affect social change” (Vedung, 2007, p. 21). The use of such tools comes from the government’s lost control over the local authorities, which are supposed to be more or less autonomous (van Kersbergen & van Warden, 2004). The sudden need for more control can possibly be explained as an effect of the audit society. Depending on the kind of effects, and the strategies used to achieve those effects, the government can choose between different kinds of instruments, classified into three categories: carrots, sticks, and sermons (see Bemelmans-Videc, Rist & Vedung, 2007). Subsidies are categorised as carrots. To be able to affect local
activities to the extent desired, the government needs to replace ‘hard’ policy instruments, such as laws and regulations, with incentives (Pierre & Peters, 2000). In the literature, opinions are divided over whether subsidies are a hard or soft policy instrument. In the common combination with time-limited programmes and their mostly general character, subsidies are here considered to be soft (cf. Feltenius, 2011).

There are many negative aspects ascribed to subsidies. In short, critics say that they lack clarity and coherence and their goals are often vague. The conditions for receiving them are usually unclear and broadly formulated and, furthermore, as the efficiency and effectiveness are flawed, the possibilities to adequately monitor or evaluate them are limited (Leeuw, 2007). The Government’s Survey Support [Statskontoret] (2007) mentions that demands for evaluations often arise before any substantial effects can be expected. The subsidies are supposed to contribute in positive ways, but the recipients, that is, the local authorities, often experience them as short-term and unpredictable, and a cause of disjointed government. This means that it is difficult for the local authorities to predict what resources they will have in the near future (Statskontoret, 2003). Another negative aspect is that temporary subsidies take a lot of time and resources to administrate, which can be burdensome, especially for small municipalities.

It is more or less mandatory to evaluate time-limited programmes financed by subsidies. The Government’s Survey Support has repeatedly questioned the costly and resource-intensive evaluations, as it seems particularly difficult to measure any effects of subsidies. The difficulty is due to the human service organisations’ high degree of complexity, which make them complicated evaluation objects. Moreover, it is seldom possible to isolate the effects of a subsidy or programme that represents only a part of a larger organisation or activity. With this in mind, the Government’s Survey Support suggests follow-ups⁶ rather than ambitious evaluations. Finally, many researchers have indicated that subsidies and time-limited programmes are symbolic political expressions. Wiseman (1977, in Leeuw, 2007) and Jasinowski (1973, in Leeuw, 2007) claims that parties try to win votes by promising subsidies.

⁶ Follow-up seems to be a concept that has many different meanings. Vedung (2009) argues that within a follow-up, one continuously collects data but one does not relate the results to the intervention or any value criteria. A follow-up usually describes what is happening within a programme, but does not explain why it happens. Evaluation and follow-up are partly overlapping (Eriksson & Karlsson, 1998), but one can somewhat simplified say that a follow-up is a much less complicated evaluation.
Evaluation approaches and models

There are many evaluation approaches and models available, and many variations in how they are organised. The large number of approaches can be explained by the varied opinions on how to undertake evaluation and research – which is related to views on knowledge, its purpose and meaning and the possibilities of conducting an evaluation. There is considerable variation between different approaches. Some of them are very simple and can almost be compared to guidelines, whereas others are more like research paradigms. Stufflebeam (2001) makes a comparison between the concepts of approach and model, which I have not addressed earlier in this thesis. However, Stufflebeam writes that approach encompasses everything from illicit to laudatory practices, whilst model may not cover all kinds of ideas on how to conduct an evaluation (p. 9). Hereafter, I will continue to use the terms alternately. Obviously, there is no consensus within the evaluation community on why evaluation is needed, how to conduct it, or even how to define concepts related to it (Eriksson & Karlsson, 2008). My ambition is to present a brief overview some prominent evaluation approaches and models in order to contextualise my own research in the field.

Variations in evaluation approaches

Probably the best-known way to classify evaluation activities is the summative-formative dichotomy (see e.g. Scriven, 1996). Summative evaluation, or the mainstream view of evaluation, describes evaluation as an activity that has a starting point, a midpoint, and an end point culminating in an evaluation report. The report is supposed to serve as a basis for decision making. Formative evaluation, on the other hand, also called the alternative and extended view of evaluation research, is when the evaluator studies what is happening during and within the evaluation process. The final results are usually secondary (Ross & Cronbach, 1976, pp. 17-18). There are several evaluation approaches that are typically associated with the summative perspective, as goal-oriented ones focusing on decisions and accountability. As stated earlier, accountability is an important part of New Public Management. Therefore, summative evaluations have gained favour in public administration. Within formative evaluation, one can design and redesign the intervention while it is ongoing. Ordinarily, such a procedure positions the evaluator close to the stakeholders and participants. Different client-centred evaluation models are therefore good examples of a formative evaluation perspective.
Evaluation approaches reveal something about their ideological and epistemological backgrounds. House (1978) uses this information in his classical scientific article where he examines what he calls the “major models” (p. 4). These models can be found in Sweden also. The first approach House writes about is System Analysis, wherein one measures the output and then tries to relate differences in programmes to variations in test scores. System Analysis evaluation models often have an experimental design and have gained approval within medical research. Critics have pointed out that this model can be used for control rather than improvement, which tends to make the evaluands adapt to the auditing. Moreover, System Analysis has been accused of putting qualitative goals aside and focusing on quantitative, easier to measure, goals. The second approach in House’s sketch is Behavioural Objectives. The objectives of an intervention are defined by individual performances, which can be attributed to the specific individual’s behaviour. Behavioural Objectives are strongly related to measurements by quantitative techniques. Åberg (1997) relates both System Analysis and Behavioural Objectives to Vedung’s goal-attainment evaluation model (Vedung, 2009), as they both focus mainly on finding out whether or not the goals of the interventions have been fulfilled. The third evaluation approach is Decision Making, wherein the evaluation is organised by the decisions to be made (House, 1978). The model is therefore mainly aimed towards decision makers such as administrators and managers. It is usually not used after a terminated intervention, but is ongoing, supporting the organisation with information. The model is considered to be close to practice, which may increase its utility. Daniel Stufflebeam’s CIPP model is probably the best-known model within the category of evaluations for Decision Making (see Stufflebeam & Coryn, 2014). Goal Free evaluation is the fourth in this brief sketch of models. This is a model that has probably not been tried out in practice very often, but can still be important as it distances itself from how the established goals affect the evaluation results in twisted ways. Scriven (1973, 1974) introduced goal-free evaluation in sharp contrast to the widespread goal-based ones. Here, evaluators should disregard the goals of the programme and search for all kinds of effects, regardless of the commissioners’ or programme developers’ opinions. Goal-free evaluation is therefore an excellent tool for finding what would otherwise be considered as side effects. Stufflebeam and Coryn (2014) call the approach consumer-oriented because it is less prone to biases and more equitable in taking a wide range of values into consideration. Moreover, it is usually a low-cost evaluation approach. House’s (1978) fifth approach is Art Criticism, which
means that experts judge the important elements of a programme or an intervention. Art Criticism can be interpreted as similar to the connoisseurship and criticism approach (see Eisner 1985, 1998). Connoisseurship means ‘knowing’, and the approach involves in-depth analysis of a person considered to be an expert within a given substantive area who knows things by experience. The expert also needs to have a critical side and be able to view both sides of a phenomenon (Eisner, 1985). The sixth approach in House’s taxonomy is Accreditation, which is like a peer review function or judgement by colleagues (Åberg, 1997). This model is frequently used in a number of Swedish organisations, including universities, schools, and elderly care facilities. They gather in networks that aim to review each other according to external standards. The results are often supposed to be outlined as recommendations, and are intended to stimulate dialogue and development. House’s penultimate approach is Adversary. Briefly, adversary is a quasi-legal approach in which a jury, or something similar, presents the pros and cons of a programme. The model is supposed to give a picture of both sides. House’s final approach is Transaction, which usually uses a case study methodology (see Stake, 1995). Professionals and users are included in the evaluation and the methods used are described by House (1978, p. 5) as “informal”.

House presents his eight evaluation approaches in the order exhibited below, where he starts with System Analysis and ends with Transaction. Taken together, the approaches illustrate the large span in the evaluation field concerning everything from who comprises the main audience to consensus on goals, methodology, and outcome (see figure 1).
The major audience for the model at the top of the list is elitist-oriented and it becomes more oriented to clients and practitioners moving toward the bottom of the figure. There are greater demands for consensus on goals at the beginning of the list and more room for varying opinions at the bottom. The column dealing with methodology is interesting in more than one respect. The list begins at the top with objectivism and then moves downward towards subjectivism. This movement reminds of the evidence hierarchy, with experimental designs typical for an objectivistic ontological standpoint at the top and case studies and qualitative studies at the bottom (see e.g. McNeece & Thyer, 2004). Finally, when one talks about evaluation models in terms of outcome, House’s taxonomy goes from emphasising effects and productivity in the first four models (System Analysis, Behavioural Objectives, Decision Making, and Goal Free) to focusing on improvement, acceptance, resolution, and understanding in the rest.

House’s article ends with a division of utilitarian and pluralistic evaluation models. Utilitarian evaluation models (System Analysis, Behavioural Objectives, Decision Making, and Goal Free) focus on demonstrating effects and productivity, while pluralistic models (Art Criticism, Accreditation, Adversary Transaction) allow for more flexibility in methods and outcomes.

---

7 Figure 1 is inspired from House’s figure 1 (1978, p. 12) and Åberg’s interpretation of that figure (1997, p. 35).
Objectives, Decision Making, and Goal Free) aim to examine a programme’s or an intervention’s social utility, whilst pluralistic evaluation (Art Criticism, Accreditation, Adversary, and Transaction) has the interest of stakeholders at heart. Both of these groups include models that have been designed for the elite, that is, the commissioners and sponsors (System Analysis, Behavioural Objectives, Decision Making, Art Criticism, Accreditation) and models that have been designed for the masses (Goal Free, Adversary, and Transaction). The utilitarian approaches tend to use quantitative techniques to collect information and they find objectivity necessary for producing good evaluations. The pluralistic approaches, on the other hand, result in a subjectivist epistemology wherein qualitative techniques are diligently used. The advantages of viewing evaluation approaches from House’s perspective are that he clarifies how one can separate different approaches by examining how they deal with vital aspects of evaluation such as audience, goals, and outcomes.

**Evaluation use**

Evaluation researchers are prominent contributors when it comes to studying use of knowledge derived from evaluation and research (see e.g. Alkin, 2013; Balthasar, 2006; Contandriopoulos & Brouselle, 2012; Dahler-Larsen & Krogstrup, 2000; King, 2004; Rossi, Freeman & Lipsey, 1999; Saunders, 2012; Weiss, 1988, 1998; Weiss, Murphy-Graham & Birkeland, 2005). This may be so because of their negative experiences of repeatedly finding that their evaluations are not being used as intended, if they are used at all (Patton, 2008). This problem has been known since the 1960s and is well discussed in the literature (see e.g. Vedung, 2009; Weiss & Bucuvalas, 1980). Evaluation use is, in fact, the most researched area of evaluation (Christie, 2007). For a long time, scholars have debated the evaluation paradox: evaluations are continuously commissioned and seem to be mandatory in most fields within the public sector, but they do not seem to be of much use – at least there seems to be no visible, rational use. Evaluation use is defined as “the application of evaluation processes, products, or findings to produce an effect” (Johnson et al., 2009, p. 378).

In the literature, there are several potential explanations for limited evaluation use. Four common arguments are 1) flaws in the evaluation, 2) flaws in the communication between evaluators and sponsors, 3) recipients’ organisational and individual shortcomings, and 4) exaggerated or wrong belief in evaluation use in general. I would also like to add a fifth argument that changes in society, and maybe particularly in politics, happen so fast
that many evaluations become out of date even before they are finished (see Petersén & Olsson, 2014). Flaws in the evaluation can result in a badly conducted evaluation, with inadequate methods, avoidance of certain questions or issues, unlikely conclusions, and irrelevant theoretical frameworks, etcetera. Such flaws can first be ascribed to the evaluators, but they can also be a consequence of bad conditions for evaluating. Flaws in the communication between evaluators and sponsors could include issues such as a lack of clarity concerning delivery date of the evaluation results and how they are supposed to be designed and disseminated. Parenthetically speaking, recent research has acknowledged that evaluator competence plays an important role in evaluation use, for instance, with respect to the evaluator’s professional skills and ability to lead the work (Johnson, et al., 2009). Further, the recipients’ may be incapable of understanding or acting on the results. Such incapacity can depend on the recipients’ individual characteristics, such as lack of education or knowledge, but may also be due to organisational conditions that prevent the recipients from using the results. Finally, an argument most evaluation researchers seem to agree upon nowadays is that the belief in rational, instrumental use of evaluation results is exaggerated. Evaluations can have purposes other than direct use, such as enlightenment functions or setting issues on the agenda, even if there is a well-thought-out purpose from the beginning. However, different types of evaluation use are considered in the literature. Four of the most common ones discussed are instrumental use, conceptual use (also called enlightenment), symbolic use, and ritual use (Feinstein, 2002; Marra, 2000; Weiss, 1998; Widmer & Neuenschwander, 2004).

Instrumental use has been the dominant perspective within use of knowledge derived from research and evaluations. It addresses use as an immediate consequence of the results. For instance, instrumental use entails terminating a non-functional programme or supporting a programme that is working well. It is difficult to appreciate the extent to which new knowledge is used instrumentally, but almost everyone engaged in evaluation research agrees that it must be only to very small extent. Nevertheless, many people seem to perceive instrumental use as the ideal use and presuppose that it is how knowledge is used (Dahler-Larsen, 1998).

Even when evaluation results are not used as intended, the results can affect politics in other ways. Such use is defined as conceptual use, or enlightenment (Weiss, 1988, 1998, 2004). Conceptual use is based on the idea that evaluation and research influence ideas and conceptual understanding among policy and decision makers, public officials, and other stakeholders.
In the long run, the results can give insights into and better understanding of a programme, and one can talk about an inflow of ideas which may later lead to reorganisations in politics and administration. In such cases, evaluations would be important intellectual contributions and useful for further developing of policies. There are studies that confirm that conceptual use is more frequent than instrumental use (Amara, Ouimet & Landry, 2004).

These two knowledge use approaches describe the bona fide attempts to use knowledge. As a contrast, there are other types of use that underpin the assumptions that evaluations and research have limited influence in political contexts (see e.g. Sandberg & Faugert, 2007). Critical research has identified symbolic and ritual use as two unserious approaches. Symbolic use refers to an unserious attitude towards utilisation (Patton, 2008). The underlying assumption of such use is that the evaluation or research process and findings are not considered important or worth noticing in relation to the organisation in focus (Alkin & Taut, 2003). Anyhow, symbolic use is a part of the political game where the politicians at least have to pretend that they value research and base their decisions upon evidence. Skilful politicians can give the impression of rational and committed engagement and thus signalise credibility (Dahler-Larsen & Larsen, 2001).

Ritual use occurs when evaluations are commissioned as a matter of routine. It makes evaluation a taken-for-granted activity, and commissioners consider it an institutional obligation that gives the organisation a sense that it is acting legally and rationally. Evaluations work as instruments for acknowledging the organisation’s existence, as well as the political decisions that preceded a programme (Rabo, 2006). We can understand ritual use as following the logic of appropriateness. It is driven by social expectations about appropriate behaviour (values and norms) (March & Olsen, 1989). Thus, it is not a goal-driven activity; it is embedded in a social context of expectations and traditions. However, even though an evaluation is commissioned and performed as ritual, it can also be used for other purposes, like the ones mentioned above (Rabo, 2006; Vedung, 2009).

**Summing up**

Evaluation is considered to be a contested concept with many variations concerning how to be defined. Most definitions are based upon three basic components; evaluation is the systematic and precise collection of knowledge, to evaluate is to make a judgement, and evaluation aims to have some kind of impact. Evaluation is in general a young discipline and took
off in its present form during the 1960s. Since then, it has become institutionalised within the context of NPM, wherein it has a number of important functions. Sweden has been very receptive when it comes to adapt to NPM and implement evaluation in public welfare practices. There is a rich selection of different evaluation approaches and models to choose from. They can be divided into different categories, such as summative and formative, or pluralistic and utilitarian. But, no matter choice of model, evaluations seem to be sparingly used in practice. In the literature, one finds different categories of evaluation use, whereof four common types are instrumental, conceptual, symbolic and ritual use.
3. Knowledge in Social Work

Social workers, researchers, and evaluators in social work continuously face questions on knowledge. Nygren, Blom, and Morén (2006) line up questions that are both complex and important for the field of social work, such as What is knowledge in social work? Why are we searching for knowledge? Under which conditions and in what ways is knowledge created or produced in social work? How is knowledge used and under which conditions? (p. 17, my translation). These questions also apply to evaluation in social work, as evaluation has become an increasingly important part of the public sector in the last decades. Evaluations are today one of the most common tools for producing knowledge in social work. This chapter will cover a number of themes related to knowledge, and to knowledge in social work in particular.

During the social work profession’s formative years, its development was driven by moral and religious incentives (Bisman, 2004; Corby, 2006). Different volunteer organisations, which originated from earlier philanthropic movements, carried out social work (Meeuwisse & Swärd, 2016). In many respects, the period 1870–1920 is considered as the time when social work broadened from reliance on volunteer action to include political action (Bisman, 2004; Parton, 2000). With this development, a shift occurred from direct aid grounded in “altruism, love, idealism and Christian mercy” (Meeuwisse & Swärd, 2016, p. 31, my translation) to a stronger focus on knowledge. Social work’s aims and goals remained practical and moral, although the respect for science gradually became more and more evident. The professionalisation of social work continued during the 20th century, as the state slowly took over what had earlier been the task of philanthropic movements. Many western countries had by this time begun to introduce education in social work. In Sweden, the National Association of Social Welfare [Centralförbundet för socialt arbete] played a significant role as a lobbyist that strongly emphasised the importance of different forms of knowledge in social work, claiming theoretical knowledge as well as practical. Hitherto, the association had also underlined that elements of love and persistence are indispensable in social work (Meeuwisse & Swärd, 2016). I make a historical leap from when the first Swedish school of higher social work education was established in 1921 in Stockholm (Nygren, Blom, & Morén, 2006), to simply conclude that research and science have over the years received more and more attention in social work education. Dunk-West (2013) even claims that interest in research nowadays is fundamental to social work.
Before delving further into the theme of knowledge in social work, I will define the way I use the concept knowledge. There is a lack of agreement about the nature of reality and how to understand it. Even the ancient Greeks fought over alternative ways of knowing. There is no unified definition of knowledge that is viable in all situations, and in the literature one finds many implicit or explicit typologies (see de Brabander, 2000). Knowledge is thus an ambiguous concept. The Oxford Dictionary defines knowledge as follows: “facts, information, and skills acquired through experience or education; the theoretical or practical understanding of a subject”. This division of knowledge into theoretical and practical categories is similar to a long-standing philosophy rooted in the ideas of Aristotle. Aristotle considered knowledge as pluralistic, meaning that he was of the opinion that different forms of knowledge are needed; theoretical–scientific knowledge (episteme) for understanding how the world is structured, practical–productive knowledge (techne) for creating, and practical knowledge (phronesis) for developing good judgement (Gustavsson, 2000). This knowledge triad is the point of departure for how I use knowledge in the present thesis. The three knowledge forms are relevant, due to the imprints they leave on the modern debate about knowledge and the knowledge society (cf. Gustavsson, 2000, p. 15).

**Aristotle’s knowledge triad**

Originally, Aristotle’s triad of knowledge forms was not a triad at all, but a quintet. It consists of episteme, techne, phronesis, sophia, and nous (Aristotle, 2000). Here, focus will be on the first three, but to satisfy any eventual curiosity, I can add that sophia roughly approximates philosophical knowledge or wisdom, and nous intuition. It appears to me that sophia and nous have been forgotten by most scholars, except by philosophers perhaps, while the rest of the knowledge forms have been kept in mind or received new attention lately. Aristotle’s texts in *The Nicomachean Ethics* (2000) are the original source for reading about episteme, techne, and phronesis. Luckily, there are many scholars who have interpreted and explained Aristotle’s words in ways that are more up to date (see, e.g., Flyvbjerg, 2001; Gustavsson, 2000; MacIntyre, 2007; Ramírez, 1995). What is important to know when reading my texts about Aristotle’s knowledge forms is that I am only scratching the surface. Some interpreters of Aristotle’s work claim that one cannot use or write about episteme, techne, and phronesis without also involving the complete set of concepts from his comprehensive view of the-
ory and practice. However, most scholars are satisfied with limiting themselves to the triad, without giving any further explanations. For myself, I can, after some heavy reading, see the point of being familiar with The Nicomachean Ethics and its content. However, I find the knowledge triad, without sophia and nous, to be sufficient for answering the research questions in this thesis.

**Episteme**

The Greek word episteme refers to theoretical–scientific knowledge. One sometimes talks about ‘real’ science or ‘true’ science referring to episteme (Gadamer, 2004; Gustavsson, 2000), and by that one usually means knowledge derived from research, or knowledge that is not based upon opinions but evidence. In general, there are a number of different perspectives on how to define scientific knowledge, and these perspectives are often attached to specific subject areas. In short, the difference is especially noticeable between the natural sciences and the social sciences, where the latter have been more receptive to the view that research is also affected by values and norms due to social and cultural influences. Thus, the foundation of traditional ideas of science addresses episteme, that there is an independent reality that one can gain true knowledge from through observations and that is possible to verify or falsify. Episteme concerns knowledge that is assumed to be eternal and does not vary despite time and space. During antiquity, philosophers argued they could reach knowledge through theoria, the banal seeing. This seeing demanded concentration and detachment, a state of mind that Hannah Arendt (1988) called vita contemplativa. Today, most researchers do not usually devote themselves to such passive activities, and vita contemplativa has been replaced by the collecting of observations and other empirical strategies to obtain knowledge.

It is particularly in the methodological position of positivism that episteme appears especially clear. Positivism takes departure in ideas represented by Immanuel Kant (1724–1804), among others. Kant characterised scientific knowledge as nomothetic, meaning that the laws of nature are always applicable and constant. On the other hand, he did not separate practical knowledge from technical knowledge. In that sense, he attacked Aristotle’s original ideas of knowledge. Positivism was developed further during the 1800s and 1900s by advocates such as Auguste Comte, Otto Neurath, and Moritz Schlick. They belonged to a group of logical positivists whose most important criteria for science was verification. In order to verify some-
thing, it must first be observable. Value-related propositions were not considered possible to observe with the senses and therefore were meaningless to study scientifically. However, today positivism is more about falsification than verification. With the advent of Karl Popper’s central idea of keeping one’s theories open to criticism, positivists all over the world have made a U-turn. Instead of verifying by observations, one now creates hypotheses that need to be tested. If one cannot test the premise, it cannot be considered as scientific knowledge, and if it is not tested, it is not yet to be considered as scientific knowledge. It has to be possible to describe statements about patterns, which means that one cannot just ‘feel’ them. They must ‘be tested empirically according to a falsification principle and a correspondence theory of truth’ (Moses & Knutsen, 2007, p. 9). Knowledge derived from such research is sometimes called ‘explicit’ (Ellström, 1992) and is more or less easy to express verbally or in writing. If I were to dare to suggest an example of how a social worker might use episteme, without any elements of other knowledge forms, it might look like this: A social worker uses a an evidence-based standardised assessment instrument, developed by researchers, to measure a client’s situation and problems. The social worker asks the questions exactly as they are formulated in the guide, and the client responds, selecting from a list of predetermined answers. After finishing the assessment, the social worker enters the answers into a computer to find out how the client scored. The result may provide suggestions as to what kind of treatment is needed. The social worker is, without further thoughts on why, offering the client the treatment that the test suggests. The knowledge used here is based solely on scientific research.

**Techne**

Techne denotes practically applied knowledge. This knowledge is associated with the handicraft arts and represents practice, in contrast to theory. Ramírez (1995) calls techne the practical–productive form of knowledge and states that its value lies within its results; how to reach the goal is of less importance. One can rediscover the word techne in other concepts, such as technology and technique. However, techne should not be confused by an exclusive association with knowledge about technology. Instead, techne means more broadly to produce something or to create something. This type of action is called poiósis by Aristotle, and techne is the practical knowledge of poiósis (Gustavsson, 2000). In other words, techne is those skills that one needs to perform a handicraft, made of material or symbolic tools. The characteristics of techne are that it is pragmatic and context dependent.
From the perspective of techne, the process is – as already mentioned – secondary, whilst the result is everything. In social work, that would mean that the relationship between a social worker and a client is irrelevant, as long as the result is satisfactory. It does not matter if the social worker provides clients with unethical interventions, threatens them, or corrupts them – the final result is in focus.

**Phronesis**

Phronesis can be translated as practical wisdom and is about having knowledge of how to act in a specific situation, general truths notwithstanding. Other definitions relate phronesis to ethical sensitivity, suggesting that to enhance the genesis of ‘the good in life’, people need to resolve to make correct decisions. Such decision-making abilities are strongly related to phronetic virtues. In other words, practical wisdom takes departure in understanding what is best for the individual. It involves the ability to judge how to achieve a certain end and the ability how to reflect upon and determine good ends. For example, phronesis becomes concrete when an experienced social worker understands when to disregard general rules and instead use experience and value-based judgement for action. A very simple example would be that of a social worker offering a client a meeting immediately, and not after two weeks when she actually has a gap in the calendar, if thanks to her long experience she understands after just one phone call that the client is in some sort of crisis. Hence, phronesis is primarily relevant under concrete conditions and stresses sensitivity to the unique situation. Aristotle desired to balance instrumental rationality with value rationality, which here means to combine different forms of knowledge in order to explore how they can influence and gain from each other. Instrumental rationality is the leading vision in modernity, but there are many scholars who emphasise that value rationality needs to take a greater place in knowledge production to become more relevant for society (see e.g. Parton, 2000; Webb, 2001). Aristotle himself wrote about how the virtues of the knowledge triad relate to each other:

We may grasp the nature of prudence [phronesis] if we consider what sort of people we call prudent. Well, it is thought to be the mark of a prudent man to be able to deliberate rightly about what is good and advantageous. . . But nobody deliberates about things that are invariable. . . So. . . prudence cannot be science or art; not science [episteme] because what can be done is a variable (it may be done in different ways, or not done at all), and not an art
[techne] because action and production are generically different. For production aims at an end other than itself but this is impossible in the case of action, because the end is merely doing well. What remains, then, is that it is a true state, reasoned, and capable of action with regard to things that are good or bad for man. We consider that this quality belongs to those who understand the management of households or states. (N.E. 1140a24–1140b12)

So, practical wisdom is a question of how to act to make conditions as good as possible for humanity. It is not just any process of decision-making and actions, but the process that takes departure in true understanding of what is best for an individual and what is best for society. In contrast to techne, the result is not in focus at all, but rather the actions that cause the result. The practical wisdom is in these actions, a practical knowledge about how one should act in order to do good for humanity. Phronesis is counted as a virtue. Virtues can be considered not only as a means to reach a certain goal, but also as goals in themselves. It is usually desirable to be wise. To develop phronesis, one cannot just put theory into practice; one has to rehearse. Experience is the leading word for developing phronesis, and experience first comes from doing in practice. That means that practical wisdom only counts when it comes to concrete actions (Kalman, 2006). Things have been pretty quiet with respect to phronesis during the last, say, 2000 years. While the concepts of episteme and techne have lived on in modern language, phronesis has languished in the doldrums. However, in recent decades numerous scholars (see, e.g., Flyvbjerg, 2001; Nussbaum, 2001; Polkinghorne, 2004; Schön, 1987) have called for its return to use in discussions on professional knowledge and even for a reconceptualisation of social science (Kinsella & Pitman, 2012).

To me, phronesis has been really hard to understand, which has led me to incorporate the insights of other researchers to a great extent, of whom I will particularly mention Bent Flyvbjerg. Another important strategy for developing my understanding of phronesis has been to search for concrete examples. One of the first I came across I found in a book with the title Duty, Profit, and the Art of Being Human [Plikten, profiten och konsten att vara människa], written by Göran Rosenberg (2013). In the beginning of the book, the author describes the death of a close relative. When the author and his family come to visit the relative, he is lying on his deathbed.

---

8 According to Wikipedia, Aristotle lived between 384 BCE and 322 BCE, which implies that the concept phronesis is about 2360 years old.
to him sits an employee at the nursing home, holding the man’s hand. The employee is doing something more than just taking care of a patient, something that is not written in any job description and will not lead to any measurable result. Despite that, the employee does what he does. It is in his action that one finds value.

Phronesis has been giving rise to a number of discussions on how to interpret the concept. One common misinterpretation is to confuse it with tacit knowledge. Tacit knowledge is associated with the polymath Michael Polanyi (1967), whose idea of a tacit dimension in knowledge has gained both approval and criticism. Tacit knowledge encompasses all three of Aristotle’s knowledge forms, and not just phronesis. Ramírez (1995) explains what can be understood as the difference between tacit knowledge and phronesis, that phronesis cannot be described in words; rather, it is the describer who finds the knowledge in the words and in other expressions. Phronesis is therefore closely related to action, as it is knowledge that comes to be expressed in actions. It is unspoken, but not silent. Tacit knowledge, on the other hand, is silent and does not demand actions.

So, Aristotle came up with the idea of a number of different knowledge forms, whereof three have been discussed here. Episteme, techne, and phronesis – or, if one prefers, theoretical–scientific knowledge, practical–productive knowledge, and practical wisdom, – are all relevant in social welfare in different ways. They form jointly a framework of what needs to be taken into account when evaluating in social work.

**Knowledge use in social work**

Adopting knowledge from research and evaluation is, historically as well as currently, a great challenge for social work. The field has been described as suffering from a long-standing gap between practice and research (Longhofer & Floersch, 2012; Nilsson & Sunesson, 1988; Taylor & White, 2005). Furthermore, many scholars question whether social work may lack an exclusive knowledge base of its own (MacDonald, Postle & Dawson, 2008; Osmond, 2005; Pawson, Boaz, Grayson, Long, & Barnes, 2003; Trevithick, 2008). The latter can be explained by the fact that social work encompasses many broad and complex issues in society. Social workers are also to be found in quite different areas of work, such as in municipal activities, in health care, and also in the private sector or in volunteer organisations. The functions of their professional field can include relief of psychological distress and material need, social reform, and social control.
The field of governing is characterised by imprecision. There are a number of different formal and informal laws, rules, principles, and norms to take into account as a social worker. The knowledge base for social work must hold a plurality of theories and methods in order to meet the demands.

To understand the role of research and knowledge in social work, one needs to know a little about the nature of social work. I personally find social work to be a charged field, as it is very complex, sensitive, and dynamic. Many people have opinions about social work, and it also seems to be one of the media’s favourite topics. Below, I will make a short summary of the nature of social work, as defined by the International Federation of Social Workers (IFSW) (2014):

Social work is a practice-based profession and an academic discipline that promotes social change and development, social cohesion, and the empowerment and liberation of people. Principles of social justice, human rights, collective responsibility and respect for diversities are central to social work. Underpinned by theories of social work, social sciences, humanities and indigenous knowledge, social work engages people and structures to address life challenges and enhance wellbeing.

To answer to this definition, a social worker must be a person who can communicate, make judgements, help and strengthen people, and work under pressure, to mention a few traits (cf. Cree, 2003). In the light of this, social workers must be given a certain margin of autonomy and discretion to be able to conduct their duties (Freidson, 2001). This is the ‘room’ that the organisation the social worker is representing leaves for manoeuvre (Svensson, Johansson, & Laanemets, 2008). The room is designed from the organisation’s routines, rules, and principles, along with the social workers’ and the clients’ own individual characteristics. It is further affected by how the social workers use their discretion in this room, which varies depending on experience, for instance.

Social work is further very complex. Complexity can have a number of meanings. First of all, it perhaps would be more accurate to say that the field of social work operates in a complex world, as indicated above. The problems faced may require the engagement and cooperation of several professions besides social workers, such as teachers, psychologists, and police officers. The social worker may have to take different roles as well. A social worker can be a listener, an advocate, or an advisor, among other things.
A social worker can also be a person with the power to affect peoples’ lives deeply.

Within this context, social workers use different kinds of knowledge, of which knowledge from research is one type. Findings from research and evaluation can help in tackling different problems, but must also be combined and balanced with knowledge from other sources.

Social work is governed by many different actors, such as social workers, senior civil servants, and politicians. The distinction between these actors regarding their room for manoeuver is highly fluid. Their presence in numbers and hierarchy is illustrated by the pyramid below (Figure 2). Politicians, at the top of the pyramid, are democratically elected and the ones chosen to keep overall control. Senior civil servants have been called “the backbone” of public administration (Hysing & Olsson, 2012, p. 9), and they are supposed to be responsible for implementing the politicians’ decisions. Relevant examples of senior civil servants in this thesis are both people working at national authorities, such as the National Board of Health and Welfare, and those working with governance of public administration in the municipalities. The third group of knowledge users in social work are the social workers themselves, often called the front-line workers. They are usually also considered as civil servants, but lower in the hierarchy. They translate the political decisions into practice.

![Figure 2. Social work governors](image-url)
The collective term in research for people characterised as “the public bureaucracy’s outermost capillaries” (Esping, 1984, p. 72, my translation) is street-level bureaucrats (Evans & Harris, 2004; Hasenfeld, 1983; Lipsky, 1976, 1980). Street-level bureaucrats are defined as staff within human service organisations (Hasenfeld, 1983) who work directly with the clients of the organisations (Lipsky, 1980), and typical examples are social workers, teachers, and police officers. Johansson (1992) clarifies the concept further by describing a street-level bureaucrat as a civil servant within public administration, who works with the citizens on a daily basis and has considerable scope for action. Earlier, the legislation was designed in such ways that the authorities’ and their employees’ discretion was tightly constrained, while today, many of the newer laws are designed as framework laws. These laws set the frames for public service organisations, but are open to interpretation by the authorities and civil servants. That does not mean that the street-level bureaucrats can do whatever they want, but that their work extends between two poles, consisting of laws, rules, and principles, at one end, and the possibility to make their own decisions with respect to specific clients and situations, at the other (Johansson, 1992; Lipsky, 1980). In comparison to similar employees within other organisations, the street-level bureaucrats have rather broad latitude for action. Street-level bureaucrats are supposed to supply citizens with the public services and benefits of the society. At the same time, they are supposed to monitor the public order. Taken together, these two different tasks put street-level bureaucrats under political pressure (Lipsky, 1980), and they can be torn between eventually conflicting demands of being a perceptive fellow human and a representative for the organisation. In all, groups of knowledge users jointly control social work practice and use different forms of knowledge in different ways. Their preconditions for using knowledge differ according to their positions.

Use of knowledge from research among politicians and senior civil servants is complicated. Boswell (2009) writes: “Politicians and civil servants seem to be attaching more weight to using research in policymaking than ever before” (p. 3), and several scholars have testified to governments’ eagerness for evidence about ‘what works’ (Nutley, Walter, & Davies, 2007; Parsons, 2002; Sanderson, 2002; Sheppard & Ryan, 2003). At the same time, as discussed above, research has shown that decision-makers have a nonchalant attitude towards evaluations. As early as at the beginning of the 1980s, Cronbach and colleagues (1980) acknowledged that decision-makers did not base their decisions upon findings from evaluations. Despite their diligently ordering new evaluations, there is little indication that decision-
makers read the reports or take them into consideration when it is time to make decisions (Petersén & Olsson, 2014).

According to politicians and civil servants themselves, they are keener to use research knowledge if it is of – what they consider – high quality and comes from credible sources. What constitutes high quality research is a matter of debate, but the ‘evidence discourse’ is nowadays very present, and within that discourse a scientific ideal consisting of experimental designs and quantitative data is advocated. Along with New Public Management, this context systematically influences politicians and civil servants to be more receptive of research offering statistics and ‘objective’ data (see Landry, Amara, & Lamari, 2001). Most important, though, for this group of knowledge users seems to be that the research findings are clear and uncontested (Nutley et al., 2007). Research coming from authoritative sources such as universities and central governments seem to generally be perceived in such ways. People also tend to find research credible when it corresponds with their own personal or professional experiences. Weiss (1995) did discover that in those cases where the research findings challenges decision-makers’ own perceptions, the technical quality emerged as the more important factor. Further, the commission and aspects of time must be considered. Court and Young (2003, cited in Nutley et al., 2007) claim that research is more likely to be used if it is commissioned for a specific purpose, because then the commissioners find it hard to ignore the results. This seems to be especially true if the research also has a connection to local practice. A great barrier for the use of research is also the time aspect. It is a generally accepted truth that research usually takes a long time to produce, while the political practice can change quickly. Therefore, research findings may be found to be out of date or irrelevant when they are finally published. The presentation of research findings also has to have a nice package, which means it must be user-friendly and visually uncomplicated (Weiss, 1998).

According to several studies, people with higher education are more receptive than others to scientific knowledge use. This may depend on two things in particular: first, academically educated people tend to value research highly, and second, they are more likely to have the skills to find, understand, and interpret research. There may, however, be other impeding factors; for example, some professionals view research as inadequate in their work (Schön, 1995), and they may value other types of knowledge more highly (Booth, Colomb & Williams, 2003).
Social workers’ relation to knowledge from research

Earlier in this chapter, I claimed the relation between social work and research to be less than ideal, as social work has had a hard time using and implementing knowledge in practice. The relationship is, however, a difficult one for many reasons. Bengt Börjeson (2006), the first professor of social work in Sweden, had an idea that social workers suffer from poor self-esteem in relation to research. He argued that social workers do not consider themselves as equal partners to researchers and that they believe that researchers have to adapt their language in order to become understandable. As an adverse reaction, social workers criticize research in social work for being irrelevant for practice. In relation to this criticism, it is significant also to ask how important knowledge from research should be for social workers.

Bergmark and Lundström (2000) received a lot of attention in the beginning of the 21st century when they published their study on how social workers viewed knowledge from research in their work. The authors were quite critical of the social workers, as the study revealed that they seldom searched for new knowledge, but rather relied on experience. Another Swedish study, from the late 1980s, showed that social workers seemed to put more weight on internally produced research than on research from the universities (Nilsson & Sunesson, 1988). This result may support what Börjeson (2006) writes about social workers’ criticism of irrelevant research. Internationally, similar concerns about social workers’ relation to research knowledge have been raised. More and more methods have been added in social work, but studies have showed that no one seemed to know anything about the methods’ effects (Reid, 2002). Only a few social workers used knowledge from research in their work with clients (Gibbs & Gambrill, 2002), and they did not use it for decision-making (Trinder, 2000).

As an answer to the criticism of social workers that emerged during the 1990s, researchers and practitioners introduced evidence-based practice (EBP) as a way forward (Gray, Plath, & Webb, 2009; Newman, Moseley, Tierney & Ellis, 2005). EBP underpins the fundamental idea that social workers need to use scientific knowledge and practitioners’ expertise in systematic ways. It is acknowledged thanks to its ability to inform and improve practice (Nutley et al., 2007), and is believed to strengthen social work in terms of effectiveness (Cheetham, Fuller, McIvor & Petch, 1996), accountability, and efficiency (Gambrill, 2006). The movement has grown rapidly in social work during recent decades (Morago, 2006, 2010) and has been met by a mix of enthusiasm (see, e.g., MacDonald, 1999; McCracken &
Marsch, 2008; Roberts & Yeager, 2006; Sheldon, 2001), scepticism, and loud objections (see, e.g., Bergmark, Bergmark & Lundström, 2011; Rubin, 2007; Simons, 2004; Webb, 2001). The epistemological assumptions behind EBP are more or less positivistic (Avby, Nilsen & Abrandt Dahlgren, 2014; Webb, 2001), and the movement primarily supports scientific evidence that provides universal truths. This is borne out in the evidence hierarchy, wherein one finds that the ‘best’ – as in the most secure – evidence comes from randomised controlled trials (RCTs). RCTs are then followed by other kinds of experimental studies, and at the bottom of the hierarchy one finds case studies, cohort studies, and qualitative studies. These are thus deemed as less reliable, while RCTs are considered as the ‘gold standard’ in research and evaluation (Chen & Garbe, 2011; Evans, 2003).

As argued in a previous article, two camps within the evidence movement have been discerned (see Petersén & Olsson, 2015). The first camp takes a position that we call ‘hard’ or ‘pure’. By that, we mean that there is a group stressing that practice should first and foremost be based upon the ‘best research evidence’. To determine what the best research evidence is, one relies upon the evidence hierarchy and keeps it as a yardstick for selecting the best practices (cf. Aveyard & Sharp, 2013; Thyer, 2008). This position is also described as the positivist position and has been criticized for ignoring flaws in research and placing evidence in better light than it deserves (Rubin, 2007). Cohen, Stavri and Hersh (2004) argue that this camp within the evidence movement relies too heavily on empiricism and has too narrow a definition of evidence. The other camp constructs a ‘softer’ version of EBP, wherein one prefers to talk about ‘the best scientific evidence available right now’ (McNeece & Thyer, 2004; Thyer, 2003). Here, qualitative studies also can be included (Nevo & Slonim-Nevó, 2011), and clients and especially professionals can be ascribed more active roles. However, it is also suspected that this softer version is just a strategy for avoiding criticism, implying that the methodological rigour is weakened (Gray, Plath, & Webb, 2009).

Alongside the criticism of the evidence hierarchy, EBP has had to endure much other debate. A lot of it has concerned problems and challenges in implementing research in practice (see e.g. Briggs & McMillin, 2012). One of the most common critical arguments is that the position of the evidence movement is too narrow when it states that practice should be based upon evidence from research. Can professionals in social work be such rational agents in a strict scientific sense that they can solely base their decisions
upon such evidence first? Many scholars answer no to that question. Furthermore, to reconnect to the discussion on social workers as street-level bureaucrats, professionals within the public sector usually have to take a number of different perspectives into consideration in their work. Such perspectives can be organisational demands and conditions, and clients’ wishes. Nevo and Nevo-Slonim (2011) suggest a broadening of the evidence concepts and prefer to use evidence-informed practice instead of evidence-based practice. This suggestion for development has support from a number of other researchers as well (see, e.g., Chalmers, 2005; Culyer & Lomas, 2006; Epstein, 2009; McSherry, Simmons & Abbott, 2002).

EBP has developed within the well-oiled machinery of auditing, and must be understood in relation to NPM (for a summary of the ideas behind NPM, see chapter 2). EBP and NPM have a number of important traits in common. Both have a knowledge approach inspired by episteme, stressing evidence and general techniques to deal with issues. They also have a top-down rationalistic view, incorporating the belief that policy goals, rules, and evidence have general relevance, independent of context. Public officials and professionals are supposed to be able to fulfil their implementation functions successfully in practical situations through standardised treatments and general techniques. In the audit society, mistrust and control are central components (Power, 1997). Therefore, it is considered essential to assess and evaluate the performance of local units, and it is understood that the control function of evaluation is superior to other functions. The ideals of NPM optimised the possibilities for establishing EBP.

So, the answer to the criticism of the social services and the social workers has not been so smooth at all. EBP has resulted in an awareness of the relation between different interventions and their effects, and the need for documentation, follow-up, and evaluation in social work practice. Another contrasting perspective on EBP is that its advocates have too much confidence in evidence from research, believing that research can have a more direct role in practice than it most often can, and that the evidence movement neglects knowledge from practice, value judgements, power relations, and the adjustment to unique situations (cf. Hammersly, 2013).

**Summing up**
Knowledge in social work has become more and more important during the last century. Earlier, social work practice was based upon philanthropic motives, but in recent decades, scientific knowledge has been increasingly acknowledged. Knowledge is an ambiguous concept, but most people will
agree that knowledge can be both theoretical and practical. Aristotle identified several forms of knowledge, whereof the three most significant for this thesis are episteme, techne, and phronesis. Episteme is theoretical—scientific knowledge, techne is practical—productive knowledge, and phronesis is practical, value-based knowledge. Today, episteme is usually the kind of knowledge that is valued the highest; however, Aristotle did not make such a ranking between the knowledge forms.

Social work has been described as suffering from a long-standing gap between practice and research. Social workers are in need of different forms of knowledge to meet the demands of a diverse field. Besides social workers, other relevant knowledge users are politicians and civil servants. Professionals’ use of scientific knowledge is affected by many different factors. A selection of affecting factors are first, that the possibility to actually use knowledge from research depends on the professionals’ room for manoeuvre, which can both be very limited and hold considerable scope for action; second, that how they perceive the quality of knowledge and how well it fits their own ideals matters; and third, that scientific knowledge can be presented in ways that are rather hard to read and understand. In the beginning of the 1990s, many governments in the western world started to implement EBP. The motive behind this change was the will to increase social workers’ use of scientific knowledge. EBP has attracted much critique, both positive and negative. Many social workers and researchers have appreciated EBP and consider its potential to increase the quality in social work, while others think it puts too much focus on scientific knowledge at the expense of other knowledge forms.
4. Research Design and Methods

This chapter outlines the overall design of the thesis. The study locates itself within a tradition of qualitative social work research, based upon interviews and document analysis. By way of introduction, I present the research design and process. Then follows an account of the data, and in addition, the sampling techniques are described and discussed. The section concludes with a discussion about the quality of the study and reflections upon ethical considerations. The purpose of the chapter is to record the research process transparently.

My standpoint is that the social sciences differ substantially from the natural sciences. I share this opinion with a large number of scholars (see, e.g., Burr, 2003; Flyvbjerg, 2001; Weber, 1949). Traditionally, people have perceived research to generate objective, true knowledge by using scientific methods. This ‘true’ knowledge is empirically based and is supposed to be generalisable to other, similar contexts. Such ontological and epistemological ways of viewing research are, in simplified terms, characteristic of the natural sciences. The social sciences have, been characterised by many observers in a comparable fashion. Today, there is greater disagreement than previously regarding how to interpret, understand, and undertake research. I consider humans as self-reflecting, acting subjects in society, whose actions are made possible or limited by different structures. These procedures give rise to new structures and characteristics of the actors. Society is complex and dynamic, even though changes may not happen fast enough to be able to say that it is constantly changing. From this point of view, I present three different perspectives: the empirical, the normative, and the constructive perspective. One encounters these perspectives every now and then in political science (see, e.g., Lundquist, 1993, 1998; Rothstein, 1994; Sjövik, 2004). The empirical perspective addresses empirical aspects of evaluation, meaning questions of how something is. The questions can be answered empirically, but here I claim that they can credibly be synthetised from normative and constructive aspects.

The normative perspective addresses values. Values refer to something being desirable or undesirable, good or bad, right or wrong. Whether one is aware of it or not, most ideas or entities embody such values. A normative analysis aims to examine these kinds of value statements and to justify or condemn them (Badersten, 2006). Despite the value-laden character of social work, normative questions and theories do not seem to be explicitly
outstanding in social work research. Social workers and social work researchers often give advice and recommendations in their professional roles, but without clarifying the epistemological and methodological choices on which they base their proclaimed assumptions. Here, we can find the importance of incorporating a normative perspective. Rothstein (1994) sounds a note of warning concerning normative analysis: Normative analysis is not one’s personal opinion or determination of which values that are the best ones. Normative analysis answer the question of how something should be, whilst empirical analysis answers questions of how something is. What remains then, is the question of how something can be, which leads us further to the constructive perspective.

Taking departure in the constructive perspective, one can answer questions of how one can deal with something. Constructive theory builds upon normative and empirical statements taken together, whereby one considers both what is empirically possible and normatively desirable. Very simply said, a constructive perspective consists of a fusion of ‘is’ and ‘should’, albeit reality is, of course, much more complicated than that. Most fair would probably be to view the relationships between the three perspectives as dynamic and successively changing and affecting each other (cf. Sjövik, 2004, p. 54). In short, that means that the constructive perspective is not just a product of the other two, but also – after being constructed – affects them in return. Hence, a complex process is started wherein the three perspective have impact on each other.

Summing up, taking – for social work – an unusual approach to methodology by mixing empirical, normative, and constructive perspectives, this thesis seeks to explore and analyse evaluation practice.

The empirical part of the thesis

The empirical data have been organised as four cases. Within a case study, one analyses the specific and unique to develop knowledge about the general. The construction of a case requires overlooking some of its characteristics and situations; it is a subgroup of a larger phenomenon or condition. Consequently, there must be other cases of the studied phenomenon or else the definition of case study methodology is pointless. In short, a case study limits itself to studying one or several subgroups to draw conclusions about the more general category. Stake (1995) holds that a case is a specific, unique, or bounded system. It is defined by being a unit or object whose smallest parts are kept together by criteria decided in advance. The advantage of the case study approach is its potential to generate knowledge.
applicable to other situations and to produce new causal mechanisms for cumulative building of social science theories.

I have studied four empirical cases of large-scale evaluations of state-instigated temporary programmes in social work. Each evaluation considers a number of documents such as policies and memos. As a starting point in these cases, I interviewed the authors of each of the evaluations and a selection of civil servants of interest because of these cases (see Table 1 and 2). Evaluations of temporary programmes are the results of efforts made within special circumstances. State-instigated temporary programmes are in these cases financed by subsidies and funds meant for a specific purpose during a limited time. The effects of such subsidies are usually expected to live on after the formal programme has been terminated and also to contribute to new knowledge and long-term institutional learning (Montin & Hedlund, 2009). These programmes are spread all across the country and can be studied both locally and nationally. Evaluations of temporary programmes must be understood in their real-world contexts (Flyvbjerg, 2001; Stake, 1995; Yin, 1994). The case study researcher needs to be aware of the complexity that surrounds a case in order to include a variety of aspects concerning contextual circumstances (George & Bennett, 2005). The evaluations of these temporary programmes in social work are conducted in a multifaceted and varying environment. It embraces ideological conditions as well as political and historical ones. Moreover, we can further locate these programmes within the wider context of Swedish governance, and within the narrower context of social work. The strategy of using case studies enables in-depth qualitative research, which yields rich, nuanced, and extensive descriptions of evaluations and their processes in social work (cf. Merriam, 1994).

The four cases are treated separately, that is, the main purpose is not to compare them in detail, but to use them as examples of evaluations of temporary programmes. Together, they form a platform for discussion of issues concerning evaluation in social work. The cases differ both in the aim behind the evaluation and in design. In that sense, flexibility on how to use theoretical concepts was needed and provided opportunities to elaborate different theoretical ideas. Concerning generalisations, some researchers argue that it is not possible to draw any wider conclusions from one single case (for a more detailed discussion, see Flyvbjerg, 2006, pp. 224–228). I argue that one can generalise from only one case; however, such generalisations may not be in the tradition of positivistic science, which favours statistical or empirical generalisation (Danermark, Ekström, Jacobsen &
Instead, I claim that one can do analytical generalisations from cases. Here, such generalisations are made by carefully premeditated assessments regarding the extent to which the results from the study can guide the content in other similar situations. These cases represent phenomena greater than themselves. In other words, the in-depth knowledge one gains from the cases says something about evaluation practice in general. The potential for generalisation does not lie within the number of different cases, but in the theoretical framework I use to analyse the results (Flyvbjerg, 2001). I emphasise “the force of example” (Flyvbjerg, 2006, p. 228), and it is my quest to shed light on both local results and results that are generally applicable. Additionally, the cases originate from large governance efforts, in the sense that they include rather large funds, encompass a great number of service users, and have more or less impact in social work practice. Moreover, most social work interventions are aimed at groups of people who generally are ‘politically weak’, in the sense that one cannot expect them by their own powers to bring about changes within social work practice. For such reasons, it is even more important to examine cases from social work. In sum, the purpose of conducting case studies is not for comparison as such but to achieve as much knowledge as possible about these four cases (cf. George & Bennett, 2005) and to find patterns and developing theories that may be applicable to other cases.

Data
The empirical material consists of two sources of qualitative data: documents and interviews. The documents comprise evaluation reports of temporary programmes in social work, and each evaluation is the foundation of one of the four cases. The evaluation reports are thus documents that represent the final outcome of the evaluation process. They give less information about the process at hand. Therefore, they needed to be supplemented with other types of data. The interviews supplement the cases with inside information that one cannot find in the evaluation reports or other public documents.

The documents
All documents that serve as data within this thesis are official ones and at the point of the data collection could be found on the websites of the National Board of Health and Welfare, the Ministry of Health and Social Affairs, and the National Board of Institutional Care. The documents consist
of a mixture of evaluation reports, policy statements, memos, and decisions, whereof the evaluation reports are in particular focus here. They provide the foundation of this thesis and often supply the rationale behind the answers in the interviews. The documents have an informative and descriptive function, to explain the formal motives and intentions that preceded the evaluations. In sum, they serve as major parts of the cases and help to empirically contextualise them (Guba & Lincoln, 1981; Merriam, 1994). For instance, a document consisting of the commissioning of an evaluation provides information about the programme, the particular wishes for the evaluation design, and any time limits. A document consisting of an evaluation report brings a description of the field (e.g. substance abuse in Sweden) and descriptions of how the evaluation was conducted. This type of information was important for designing the interview guides and for recruiting respondents.

The number of cases, and documents, relevant for this study was limited, due to my ambition to study quite new and large state-financed evaluations in social work. Thousands of evaluations in social work and related fields in social welfare are conducted every year, and four were chosen on the basis of the following specific criteria established beforehand: First, the evaluated programme (and consequently also the evaluation report) had to be fairly recent. None of the programmes chosen had started earlier than 2003. This criterion was two-fold, for both practical reasons such as having access to the evaluation reports on the Internet, and for ‘timely relevance’. The latter aspect addressed my intention to study a topic of actuality and current interest. The second criterion was that the programmes had to be financed wholly or partially by subsidies, to fulfil an underlying purpose of addressing the government’s administration of evaluations. The third criterion was that the programmes had to target fields that are typically related to social work. The four evaluations within this thesis deal with issues such as substance abuse, elderly care, institutional care of adolescents, and honour-related violence. Finally, the fourth criterion was to include only ‘larger’ programmes, to exclude those below $25 million. By excluding ‘small’ programmes, I could leave out locally evaluated ones. The limit of $25 million was used because it represented a large financial gap to the next programme that could possibly be selected. All four programmes were nation-wide and not directed only towards selected cities or parts of Sweden. There is a possibility that I have missed a programme that ought to have been included, because if it has not been published on the Internet, it was not included in
the sample. Table 1 illustrates an overview of the four cases (see also Petersén & Olsson, 2014).

<table>
<thead>
<tr>
<th>Title of the programme</th>
<th>Time of the programme</th>
<th>Programme funding</th>
<th>Target</th>
<th>Evaluation design &amp; method</th>
<th>Commis-</th>
<th>Evaluator</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Care Agreement</td>
<td>2007–2009</td>
<td>$53 million</td>
<td>Substance abuse care</td>
<td>Pre-/post-test design, using both questionnaires and interviews</td>
<td>The govern-</td>
<td>The government and the National Board of Health and Welfare Orebro</td>
</tr>
<tr>
<td>Reinforced Care of Adolescents</td>
<td>2006–2008</td>
<td>$36 million</td>
<td>Institutional care of adolescents</td>
<td>Process evaluation, interviews, and observations</td>
<td>The govern-</td>
<td>The government and the Swedish National Board of Institutional Care Lund University</td>
</tr>
<tr>
<td>Action Against Honour-related Violence (HRV)</td>
<td>2003–2007</td>
<td>$25 million</td>
<td>Victims of HRV and a general awareness of the problem</td>
<td>Programme evaluation with post-test design and stakeholder perspective; interviews, questionnaires, document studies</td>
<td>The govern-</td>
<td>The government and the National Board of Health and Welfare Umeå University</td>
</tr>
</tbody>
</table>

Table 1. The cases

There are many strengths of using documents. They can be examined as often as needed and their contents do not change, in contrast to respondents, who may adapt themselves to an interview situation or when answering a questionnaire. The documents consulted covered a lot of information and were largely very precise. However, they only reproduced formal and official information, whereas the interviews captured informal dimensions.
Documents are of special importance when it comes to validating other information. In this thesis, the documents and the interviews validated each other in turn.

**The interviews**
The qualitative research interview has a central role within social science, as it can help obtain descriptions of peoples’ motives and intentions towards, and experiences and understandings of, a particular phenomenon. The purpose of doing interviews within the frame of this study was to achieve in-depth knowledge about evaluation practice and evaluation use. The informants contributed information that I could not obtain through reading, observations, or questionnaires (cf. Patton, 2002). Yin (2006) claims interviews to be one of the most important sources for information in case studies, for reasons I find equally compelling for other qualitative studies as well. Interviews are flexible when it comes to collecting data, as they can be either very concrete and rigid or unstructured and floating.

The sampling strategy for the interviews is a combination of purposive sampling and snowball sampling, which are two non-probability sampling techniques that rely on the judgement of the researcher in selecting interviewees. It is primarily used in qualitative studies and is characterised by the selection of individuals, groups of individuals, or organisations aimed at answering the studies’ research questions.

The power of a purposive sampling is that it offers information-rich cases, which were required here. Another reason for using a purposive sampling technique was to achieve representativeness in order to be able compare the four cases to each other (cf. Teddlie & Yu, 2007). To select informants, I identified a few criteria as starting points (cf. Coyne, 1997). The main criteria was that the informants had to work at the Ministry of Health and Social Affairs, or the National Board of Health and Welfare, and administer evaluations as part of their working duties; otherwise, the informants had to be contracted as evaluators of any of the four relevant programmes.

I started the sampling process in 2010 by contacting those civil servants at the Ministry and the Board who were administering the programmes and/or the evaluations. One person, associated with the programme Reinforced Care of Adolescents, declined participation. This person did not work at the Ministry or the National Board of Health and Welfare, but at the National Board of Institutional Care. To increase the number of informants, a snowball sampling technique was used. That method identified one more informant from the Ministry, a civil servant who was nominated by a
colleague. The method was also used successfully at the Board and gave another two interviewees. In all, with help from the snowball sampling technique, the material became richer, and several new angles were included. It is unlikely that a snowball sample would be representative for a population (Bryman, 2011; David & Sutton, 2011), but that is – as already argued – not the point in the present study. Although the population is very small, there is a good chance that other people with tasks comparable to those of the informants would have similar experiences and stories to tell, the point thus being that the selected informants also have varying experiences and insights, and in all, their stories provide a nuanced picture of evaluative inquiry and questions of knowledge in social work.

As indicated above, evaluators responsible for evaluations of each of the programmes were contacted for an interview. In one of the cases, I interviewed two of the evaluators, both the one designated as the ‘head evaluator’ in the evaluation report and the one who had spent the most hours working on the evaluation. In another, divergent case, I interviewed a person (official F/evaluator) who had kind of a dual position, alternatively conducting evaluations and also being the person who regularly commissioned evaluations. All interviewed evaluators were experienced researchers and PhDs, except for one who at that time was a PhD student. Their names were repeatedly mentioned in contexts of social work research and evaluation in Sweden, and to some extent also internationally.

All interviews, except one, were conducted by me. The one exceptional interview was conducted by my co-author (see articles I and II), but in my presence. The reason I did not do the interview by myself was that I had been involved in conducting the evaluation, as a part of the evaluation team. Each interview took about one hour. All interviews but two were recorded and transcribed verbatim. For the two not recorded, notes were taken during the conversation. In one case, the interviewee did not want to be recorded, and in the other case I contacted the person to schedule an interview at a later date. However, this person was slated to go on a longer vacation, and if I wanted any interview at all, I had to adapt to the situation and conduct it immediately. I have not used any quotations from the two interviews that were not recorded, but not because they were not recorded. These two interviewees simply did not say anything that lent itself to quotation. Otherwise, I have transformed spoken language into written language (cf. Kvale & Brinkmann, 2009). About half of the interviews took place at the informants’ workplaces, in their offices or in meeting rooms where we could talk undisturbed, and the rest were conducted over the telephone.
The interview guides
The interview guides (see Appendices 1 and 2) were semi-structured and to a large extent based on my reading and analysis of the documents. They started with questions on the position of the interviewee, mainly about the job title and descriptions of the officials’ professional tasks. These descriptions often included vital information on the processes, from commissioning an evaluation of a programme to administering and then closing it. The interview guides consisted of a number of themes but were not very detailed and allowed flexibility to switch topics during the interview. It also gave room for unforeseen follow-up questions. The three main themes were ‘the programmes’, ‘the evaluations’, and ‘use of the evaluation findings’. After asking the informants about their professional tasks, a typical initial question might be, ‘Tell me about the programme The Care Agreement.’ I would then continue the interview by asking more detailed questions, such as
‘What was the idea behind the programme?’ Additionally, I used probing and follow-up questions to gain more in-depth information. The semi-structured interview guide was used in ways such that when an interviewee gave an unexpected or exciting answer, there was room for more questioning (cf. Kvale & Brinkmann, 2009). One interview guide was used for the officials at the Ministry of Health and Social Affairs (see Appendix 1), while I used another guide for the evaluators (see Appendix 2). The difference between the guides relates to the informants’ different positions in evaluation practice. The officials were asked questions about how they planned the evaluation commission, whether they had any particular evaluation design in mind, and how they used the final results of the evaluation. The evaluators were asked questions such as how they reacted to the evaluation task, whether they had any opportunities to affect the evaluation design, and if so, how. They were also asked what happened after finishing the evaluation.

**Analysis of the empirical data**

In a qualitative study, the analysis begins when the data collection starts and continues during the complete research process (Kvale & Brinkmann, 2009; Layder, 1993; Merriam, 1994). Concerning the interviews, I followed Kvale and Brinkmann’s (2009) recommendations to start analysing as early as during the interview situation. Because of this, the same questions were sometimes posed more than once, however, usually reformulated. Re-asking questions is also a matter of validity, as it helps the interviewer to understand the respondent and the other way around.

Further, I used content analysis to establish themes and patterns in the material, both the documents and the interviews. Content analysis has become a widely used qualitative research technique due to its user-friendliness of classifying large amounts of data into manageable categories (Hsieh & Shannon, 2005; Kvale & Brinkmann, 2009). The interviews were organised using NVivo 8, which is a qualitative data analysis software package. I prearranged a number of categories into ‘tree nodes’, and other categories that came up during the analysis were sorted under ‘free nodes’. The tree nodes were initially created from the themes in the interview guides, which in turn were inspired by the documents. Reading and rereading the nodes separately and relating to the interviews and documents as a whole caused patterns and nuances to emerge. These patterns were interpreted, generating new themes and creating ways to make further comparisons and search for alternative explanations. Through this process of analysis the categories alternately grew and shrank in number and content until the moment I felt
that there was nothing new to be discovered in the material by this circular examination.

Concerning the documents, I analysed them by first reading them many times. This was necessary to understand the evaluation process. I then created a couple of categories, which can more or less be found in the interview guides. Examples of categories are ‘the programmes’, ‘evaluation design’, and ‘findings’. Later in the process, I refined the categories and developed new ones. An example is the category ‘evaluation design’, from which I created new categories such as ‘demands’ and ‘conditions’.

**Reflections upon quality**

The traditional concepts for valuing research quality are validity and reliability. They are here replaced by concepts more adapted to the constructivist view of knowledge that pervades this thesis. Validity and reliability are intimately related to a positivist ideal of knowledge, which suggests that conclusions about the studied object are only valid if they follow logical rules that are universal (Flyvbjerg, 2001), and only reliable if the study can be replicated under comparable conditions. In the 1980s Guba and Lincoln substituted validity and reliability with the concept ‘trustworthiness’ (Guba & Lincoln, 1981, 1982; Lincoln & Guba, 1985). Trustworthiness contains four aspects labelled credibility, transferability, dependability, and confirmability, which are the concepts that will be used here to discuss the study’s rigour. Roughly, credibility corresponds to internal validity, and dependability to reliability. Transferability relates to external validity, while confirmability addresses the presentation. No matter what terminology one prefers to use for qualitative inquiry, all research needs to be addressed in terms of rigour in order not to become fiction but to be useful and valuable (Morse, Barrett, Maynan, Olson & Spiers, 2002).

Working for the trustworthiness of a study is a continuous process, involving a constant questioning of the procedures, questions, results, and interpretations within the study (Kvale & Brinkmann, 2009). It may already be possible for the reader to value some of the quality of the present study by reflecting upon the description of the research process. In any event, each of the four aspects of trustworthiness will be discussed to facilitate the understanding of the researcher’s knowledge claims and the readers’ further examination of the thesis.

In qualitative research, credibility means confidence in the truth of data and analysis (Polit & Beck, 2012). A number of aspects affect the credibility of the present study. First, my earlier experiences and mistakes in research
projects and evaluations helped me avoid some of the worst pitfalls during the data collection (as lack of battery in the recorder, ‘wrong’ interviewee, and similar). Second, the respondents seemed genuinely interested in the topic of the study, which made them keen to provide rich and informative data. They also described their weaknesses and failures, which strengthened the picture of them as reliable sources. The informants’ impressions of me seemed to affect them in two contrasting ways: most of the researchers appeared to view me as a novice, and they were glad to tell me ‘how it really is’. I perceived them as helpful. The other group, consisting of people who had more administrative duties, treated me as though I were ‘more educated’ than they were, and they presented their stories humbly and sometimes a little bit defensively. The open character of the interview guide assisted the informants to reflect widely and openly on the different issues, and that helped provide rich data. Follow-up questions were posed when more detailed or clear answers were needed, or when touching upon especially interesting subjects. Credibility was further enhanced by the analysis of data, which was done both individually and in cooperation with experienced researchers. The results have been presented in public, such as in higher-level seminars and conferences, several times during the last couple of years.

Dependability is more or less the same as reliability and usually of less relevance in qualitative research. Some items are thus important to discuss. For instance, by way of introduction, the study was presented in the same way to all participants. All civil servants were interviewed from the same interview guide, except for small variations depending on their employment, and all evaluators from another interview guide. I conducted all interviews, except one that was conducted in my presence, and I transcribed them all by myself. The same occurred with the documents, which were all read and analysed in detail by me. These procedures brought me closer to the research objects and their content. The recordings of the interviews are of good quality; however, there are a handful of words where one cannot hear properly the voice of the informant. In those cases, the speech is noted as ‘inaudible’ in the transcriptions.

Could anyone replicate this study? Possibly, but not with the same results as reported here. The analysis of the documents would probably give similar results, while the interviews are more likely to differ owing to reasons of time aspects and influence of the researcher. Earlier notions of my effect on the informants suggest they would respond differently to another researcher. That is just supposition; the basic information would probably be
similar, while the informants might behave more or less timidly or confidently. That, in turn, would affect the interpretations of their narratives. The transcriptions make it possible to repeat the follow-up questions that were not detailed in the interview guide, but times change, and the participants would probably have new insights and experiences today.

Confirmability deals with the presentation and objectivity of the study. Is it, for instance, possible for another independent researcher to confirm the findings, and in such a case, to what extent? There exist assumptions that qualitative researchers’ subjectivity increases the likelihood that they will gain in-depth insights into the complexity of the studies’ phenomena (Campbell & Wasco, 2000; Hewitt, 2007). Hence, each researcher would bring a unique perspective to the qualitative study, and it would be strengthened. In any case, most researchers agree that one needs to control for bias in interpretation. To make data reflect the content of the documents and the voices of the respondents, I have, according to the recommendations of content analysis, read the material several times. A number of reviewers of the texts have played the role of ‘devil’s advocate’ and constantly questioned the interpretations made. To enhance the transparency of the interviews further, quotations are used as a strategy to validate the interpretations (Henriksson & Månsson, 1996).

The possibilities of transferring the results of this study into similar contexts are mixed. Critical readers of the thesis can draw their own conclusions about the study’s transferability from the descriptions of the context, the sample of the documents and participants, the process of analysis, and the results (Merriam, 1994): ‘the burden of proof for claimed transferability is on the receiver.’ (Guba & Lincoln, 1989, p. 241).

Additionally, I recognise the limitations of the study and suggest ways to improve the applicability of these results to other contexts. The transferability is reduced by the small sample and limited number of informants. If there exist any deviant cases, which are not included in the sample, they will change the study’s result to a large degree. No matter the number of cases at hand, a researcher can study one case that can give knowledge applicable to a wider context or the other way around – a researcher can study a large number of cases, but discover only a few similar patterns (George & Bennett, 2005). On the other hand, Morse (1999) argues that a purposive sampling, as here, extends the knowledge in the field further. The context of such evaluative inquiry as studied within this thesis, that is to say, evaluations of activities in a society that is in constant change, is varying. Absolute generalisation, as the positivist paradigm requires, is therefore impossible to
claim. For this study, I have set out extensive and careful descriptions of the context to facilitate judgments on transferability for those who have an interest in such questions. Furthermore, the ambition of this method section has been to establish clear and explicit arguments for the interpretations made within the rest of the study.

Finally, I have some additional reflections concerning the interviews and how I perceived the interview situations. Critics claim interviews to be more or less biased, depending on how the interviewer affects the informants. Semi-structured interviews are thus considered as more neutral than unstructured ones, because of the specific questions (Marlow, 1998). My influence as an interviewer can relate to several features, such as my age, gender, or role as a researcher. I did perceive that I had impact on the respondents in a few cases. The most obvious to me was when the respondent refused to meet me for an interview and also did not want to be recorded. My interpretation of this reaction is that the civil servant did not want to risk the feeling of being criticised in one’s professional role. Additionally, three of the respondents from the Board and the Ministry said things like, ‘We are not as educated as you are,’ trying to gloss over when they did not have the answers they perhaps thought they were supposed to have. This behaviour can be related to what Alvesson (2013) calls impression management, which means that the interviewees want to give a good impression of themselves, or – what is possibly more relevant here – make their actions appear rationally correct. In one situation, I perceived that a respondent was trying to advise me in doing research, and in another situation I got the impression that the respondent was becoming defensive and trying to engage me in her/his perception of research and evaluation. During those conditions, I tried to stay neutral or act curious to avoid a reactive affect and hence limit the dialogue (cf. Marlow, 1998).

**Ethical considerations**

In the introduction of this thesis I related a story of when I conducted my first evaluation. What I did not say is that I have also been involved in one of the evaluations represented here – as one of the cases (the Care Agreement). The relationship between me, as researcher and evaluator, and the research object, risks affecting the subjectivity. In the literature, the closeness between a researcher and the research object is often acknowledged as something positive. My earlier experiences, especially of the Care Agreement, provided me a number of benefits. First, it gave access to interviewees at the National Board of Health and Welfare and the Ministry of Health
and Social Affairs. Moreover, I was rather familiar with the processes of commissioning and conducting evaluations, which aided in all interview situations (cf. Clandinin & Connelly, 1994). Importantly, I brought my own set of values, beliefs, and emotions into the interviews and into my attempts to understand, interpret, and explain the content of my research (see Campbell & Wasco, 2000). I consider the notion of objectivity impossible for examining the social construction of the reality; nevertheless, the researcher–researched relationship, which has long been in the shadow of more technical aspects of qualitative research (Richard & Emslie, 2000), needs to be acknowledged to provide the readers the opportunity to do a critical reading. There are several guidelines that aim to guide researchers in their work. In Sweden, we heavily rely on the national Swedish ethical guidelines for humanities and social science (see Vetenskapsrådet, 2010). They consist of four principles, which are voluntarism, participation, confidentiality, and utilisation.

By way of introduction, I informed all respondents about the rules of voluntarism and participation. Hence, they knew it was voluntary to participate, and that they could decline to continue, even during an interview in progress. All respondents who were asked to participate accepted, but one of them refused to see me in person and did not want to be recorded. It was clear to me that all respondents were already familiar with the Swedish ethical guidelines because of their professional roles as researchers and civil servants commissioning and conducting evaluations and research.

The rule of confidentiality requires the researcher to protect the respondents’ identities and sensitive information. I cannot claim to have fulfilled this criterion successfully, as I name the evaluations by their titles. One can therefore find the evaluators who conducted them, even though one cannot know which one of the evaluators was interviewed (there was more than one evaluator in each case). The interviewees thus are guaranteed neither anonymity nor confidentiality. They have been consulted in their roles as public officials and scholars. The questions of the study are impartial, impersonal, and do not treat sensitive topics. The groups of respondents cannot be considered as vulnerable groups (cf. Raudonis, 1992). I am still divided in my opinion as to whether I ought to have tried harder to protect their identities. As professionals, they need to accept that it is in the nature of science not only to critically review the work of others but also to make one’s own research available for review. As a researcher, one is accountable for one’s own work, especially when the activities are publicly funded. That people with insight can identify some of the respondents can be mitigated
further by arguments such as that the results are first and foremost interesting on a structural and organisational level and not on an individual one. Ethics within document analysis are similar to any other research ethics:

The general principles involved in handling documents are no different from those involved in any other area of social research, but that specific features of documentary sources do require the consideration of their distinguishing features and the particular techniques needed to handle them. (Scott, 1990, p. 2)

Turning to the principle of participation, the documents had already been disseminated in the public domain, and any reader could undertake analysis without obtaining any consents. The issue for the researchers is then the lost control over their published data. Others’ new analyses and selective use of the material can be in sharp contrast of the original purpose of the research (Darlington & Scott, 2002). Because of this information, I have tried to be “ethically aware” (Grinyer, 2009, p. 4) in my reading and interpretations of the documents. Finally, the restriction on utilisation means that collected data are not to be used for anything other than research purposes. Data within this study are used only for that purpose and remain unavailable to outsiders.
5. Summary of the Articles

The three articles of the thesis are briefly summarised below. They are all published in English, but one can find an abstract in Swedish in article I. In short, the first article takes departure in an evaluation paradox in social work. Evaluations are continually prioritised, but seem to be less used for learning. In the second article, we call evidence-based practice (EBP) into question as a strategy for relevant applied knowledge in social work. Instead, we suggest praxis-based knowledge informed by phronetic knowledge as a better option in social work research and evaluation. In the third and final article, phronesis-based evaluation is presented as a new evaluation model.

Article I


This article contributes to the debate on evaluation use by analysing temporary national programmes in Swedish social work. It considers the possibility that there is an evaluation paradox in social work, which consists of two items: first, evaluations are a common form of knowledge support and are continually prioritised. A growing number of evaluations are commissioned every year. Second, in practice, evaluators often find their evaluations are not used as intended. Evaluations can have symbolic or ritual functions, or be misused or not used at all. In sum, evaluations are repeatedly ordered, but do not seem to be used for learning and development in practice. The study addresses questions such as these: How are evaluations of temporary programmes in social work designed and used? Is there an evaluation paradox within social work, and if so, how can we understand it and possibly solve it? How can evaluations of temporary programmes be designed to make them more suitable for the development and improvement of social work?

In the theoretical section two different perspectives are discussed, one that deals with the use of evaluation as an empirical–theoretical phenomenon and one that contributes in a normative–constructive way. The first perspective elaborates four frequently used theoretical concepts: instrumental, conceptual, symbolic, and ritual use. The second perspective, on the
other hand, takes departure in three evaluation approaches to enhancing evaluation use: utilisation-focused evaluation, fourth-level evaluation (responsive evaluation), and EBP.

The empirical material consists of documents and interviews. The documents are four selected evaluations of larger-scale national social work programmes, and the respondents are officials from the Swedish Ministry of Health and Social Affairs and the Swedish National Board of Health and Welfare and evaluators from different universities in Sweden.

Results show that the evaluation design process is complicated and varied. Usually, a number of participants were involved, both officials from the Ministry and the Board and evaluators. They all experienced that they had more or less influence in designing the evaluation. However, the evaluators reported being forced to adapt to the interests of the commissioners, in order to receive the evaluation task. At the same time, the commissioners gave two contrasting perspectives. They partly represented themselves as not competent enough to make demands regarding the design, but also had to emphasise their superiors’ eventual suggestions concerning design. Moreover, one commissioner found that evaluators adjusted their evaluation projects to fit their former interests.

The use of the evaluations examined was limited. The officials at the Ministry took a somewhat reactive approach when the evaluation reports arrived. There were several reasons for this behaviour, such as that they did not have time to read the reports and that they found them irrelevant, out of date, and seldom offering answers to the questions initially posed. The evaluators, on the other hand, were a little bit more positive in their responses about the evaluation use. They could certainly not remember or give any concrete examples of how the evaluations had been used by the commissioners, but they could describe their own use. The evaluators tended to use the empirical data from the evaluations for their own research interests, such as in dissertations, articles, book chapters, and educational material. Ritual use is a main conclusion in this study, but the evaluators use the results conceptually.

The results overall support an evaluation paradox in social work. Evaluations are prioritised, but the results go largely unused. Limited evaluation use can be seen as a problematic conclusion. It may undermine potentially important functions of evaluations in society, as evaluations ideally have democratic functions to inform citizens and provide transparency respecting
political actions. Nevertheless, the evaluators are diligent users of the evaluation results, and in a broader perspective that contributes to an extended bank of knowledge in social work.

Article II

This article is of theoretical character and deals with the nature and quality of applied knowledge from research and evaluations. The aim is twofold. First, EBP in social work is described, along with how it became the dominating approach. Then, praxis-based knowledge (PBK), founded in phronesis, is presented. There is a particular focus on elucidating how evaluative inquiry can be designed in a phronetic sense, since evaluation practice suffers from flaws in terms of use.

EBP, supported by the evidence-based movement, is increasingly influential and important in social work today. It is strongly supported by governments in many countries. EBP has partly gained approval because of its claimed contribution of scientific assurance of evident knowledge for professional social work. Relevance is thus tightly connected to evidence, but there are few indications of how to apply this knowledge. In the article, EBP is found inadequate as the sole provider of knowledge in social work. Instead, PBK is presented as a major important strategy for this purpose. Within PBK, the knowledge form phronesis is given much more attention than it has hitherto in social work. Phronesis is a part of Aristotle’s knowledge triad, the other elements being episteme and techne. Episteme refers to scientific theoretical knowledge, while techne, on the other hand, addresses a heteronomous and practical form of knowledge. Phronesis is roughly translated into practical wisdom or prudence. It is primarily relevant under concrete conditions, and is context dependent and stresses sensitivity in relation to the unique situation. The Danish researcher Bent Flyvbjerg has formulated four value-rational questions as a point of departure for phronetic research: (1) Where are we going? (2) Is it desirable? (3) What, if anything, should be done? (4) Who gains and who loses and by which mechanisms of power?

PBK is open to multiple sources of knowledge, and in particular, a phronetic knowledge view. Besides traditional research, PBK acknowledges knowledge that is value-based, context-dependent, and sensitive to power
relations, and grounded in practical experience. The PBK strategy needs influential professionals to generate praxis knowledge. Social workers can effectively help researchers to study social problems, because of their close relation to these problems. They can also show researchers the complexity and relevance of these problems and help the researchers to gain access to study and evaluate both problems and interventions. Accordingly, social workers should be allowed and engaged to play a central role in research and evaluation. However, they should also be encouraged to use different forms of knowledge in practical situations, instead of having manual-based ‘solutions’ imposed on them. The PBK strategy can also play a vital role in evaluation practice in social work. We discern five conditions which may advantageously serve as starting points for phronetic evaluations: (i) using Flyvbjerg’s value-rational questions as a point of departure for doing relevant evaluations in social work; (ii) starting in practical situations; (iii) being open to different types of methods and materials; (iv) being familiar with the field in question, which means avoiding consultants and researchers without knowledge of the subject; and (v) organising the evaluation through dialogue as a method, open to different claims of truth. There are a number of evaluation models that seem especially compatible with the PBK strategy, such as different stakeholder models, empowerment evaluation, and user evaluation.

After more than a decade with EBP as the dominant discourse in social work, we can easily identify EBP’s weaknesses in practice. Instead of getting seduced by the evidence-based movement, social service providers need to consider more praxis-based approaches when producing and using knowledge. Bringing phronesis back onto the agenda is of vital importance for improving knowledge production and use in social work, and the PBK strategy is our concrete suggestion for how to do it.

**Article III**


In this article, a new evaluation model for social work is suggested, called phronesis-based evaluation (PBE). In contrast to what many understand as the dominating evaluation ideal, EBP and its experimental designs, this model does not emphasise any particular design or method. Instead, the model takes departure in Aristotle’s concept phronesis, which can be translated as practical wisdom. PBE aims to provide knowledge relevant and useful in social work practice, by informing peoples’ praxis-based knowledge.
Besides phronesis, PBE also relies upon the ideas of pragmatism and Ernest House’s thoughts on democratic deliberative evaluation. Within pragmatism, one lets the research questions guide the choice of methods. That means that the researcher rejects any demands of choosing between positivism, critical realism, or constructivism, but instead welcomes a pluralistic perspective on knowledge. Pragmatic researchers argue for approaching a problem with an open mind, often by combining several different ways to reach knowledge. Therefore, pragmatism is often claimed to be the rationale or the leading contender for a mixed-method approach. In chiselling out PBE, much inspiration is taken from democratic deliberative evaluation, a model that emphasises stakeholder involvement and value-aware judgements. These two components are particularly important in PBE as well.

The central elements of PBE are starting in practical social work, letting the value-rational questions guide the evaluation process, acknowledging value and power as important components of an evaluation, being open-minded concerning methods, and making use of the results. These five items constitute the evaluation process of PBE, even though they may not follow each other in direct order, but may overlap or even change places with each other. The article embodies PBE by means of a fictive case about care for unaccompanied minors in Sweden.

Social work is about supporting and empowering people to live their lives well and in ways they have decided for themselves. Therefore, interventions in social work must be valued for their processes and not just for the final results. It is within the process that one can find the knowledge that the professionals need to develop their practical wisdom. The article points to the importance of evaluating social work in ways that help the field to learn and improve. PBE is acknowledged as a model to meet social work’s need for praxis-relevant knowledge. Social work aims particularly to work against social injustices and instead to promote and support peoples’ rights of protection and empowerment. In a world in constant change, the challenges for social work are varying in their nature and take new, unfamiliar forms. Consequently, researchers and evaluators must be able to supply social work practice with relevant knowledge. Relevant knowledge comprises findings from research and evaluation that can provide learning and improvement (change) based upon stakeholders’ voices and through analysis of values and power. Then, evaluation in social work can be used as a tool for overcoming social injustices.
6. Discussion

In this chapter the most essential results for the three articles will be discussed based on the aim of exploring and analysing evaluation practice from empirical, normative, and constructive perspectives. The ambition is to increase the understanding of how to produce relevant knowledge and to give concrete suggestions for improving evaluation in the field of social work. The research questions posed are as follows:

1. What kind of evaluation use can be identified in the field of Swedish social work, with particular focus on evaluations of temporary national programmes, and what affects the different modes of use? (the empirical perspective)

2. In what ways can different forms of knowledge enhance evaluation use in social work? What kind of knowledge should preferably guide social work? (the normative perspective)

3. How can we, through evaluations, best provide relevant and useful knowledge in social work? (the constructive perspective)

The empirical perspective: Use and non-use

Discussing the obstacles and opportunities of evaluations as knowledge contributors in social work needs to be done in the light of current evaluation trends along with the conditions of social work practice. These trends and conditions have to be confronted with the limitations and possibilities offered by present-day evaluation practice. In this section the empirical aspects of evaluation will be elaborated further.

Evaluations can hold a number of benefits. Evaluation provides knowledge about the evaluation objects, and sometimes even more information than that (so-called ‘side effects’; see Vedung, 2009). The new knowledge can be used to improve, develop, or terminate a programme or an intervention. Sometimes, the findings can be generalised to other, similar contexts, or at least others can learn from the results. Hence, the new knowledge can become valuable in many different ways. Despite all these good things that can be ascribed evaluations, research shows a gloomy picture of evaluation use. In article I, an evaluation paradox is identified, that is, evaluations are universally demanded, but rarely utilised (Petersén & Olsson, 2014). The findings are in line with previous research, but extend it further by providing new insights pointing to one group of actual users being the evaluators themselves.
Civil servants from the National Board of Health and Welfare, and from the Ministry of Health and Social Affairs, only passively engage themselves in evaluation use, and I label this kind of engagement as ‘ritual use’. Ritual use occurs when evaluations are commissioned as a matter of routine, and it is embedded in a social context of traditions and expectations of auditing (Rabo, 2006). One can probably not expect ritual use to lead to any particular changes at the commissioners’ initiative, as it permeates the whole evaluation process, from the beginning to the end, when the evaluation report is received and is supposed to lead to action. Thus, if there were no ideas at the outset about why the evaluation was being commissioned, there is likely to be a lack of intention to use it later. Larger programmes, such as the four cases in this thesis, seem to be intimately related to a subsequent evaluation. The interviewees talked about evaluation in ways that revealed patterns of evaluation as a well-established activity, but with unclear intentions. Commissioning evaluations helped the civil servants and their organisations to fulfil the institutional obligations of being auditable. The administrative trend of NPM and the audit society stresses the importance of evaluative activities. Consequently, the number of evaluations has been increasing at staggering rates, and this may be a reason why civil servants, as sponsors and commissioners, do not pay the evaluation any particular attention after it is concluded. Another reason for ritual use, or non-use, could be that the civil servants are too distant from the practices being evaluated. They do not seem to have any real insight into practice and what is happening therein as a result of the intervention. Just as research has shown that evaluation results are more likely to be used in practice if stakeholders are seriously engaged in the process, perhaps commissioners need to be engaged in similar ways. As the situation is right now, no one seems to be engaged in large-scale evaluations of temporary programmes.

The strong trend of evidence-based practice (EBP) in social work seems to affect the choice of evaluation design in favour of experimental designs having so-called ‘epistemic qualities’ (see Flyvbjerg, 2001, p. 3). To be given an evaluation task, evaluators seem to experience that they must suggest a design including a before-and-after measurement, even if they are hesitant about its value in the specific case. The commissioners are apparently ignorant of what is possible to do (evaluate) under the given circumstances, despite their eagerness to fulfil their commitment to developing an EBP. My critique here should not be ascribed to EBP as such, but to how its importance has spilled over also to situations where it cannot be applied. If
one wants to commission an evaluation of effect with an experimental design, it needs to be incorporated into the initial planning for the programme or intervention. That would be optimal no matter what the design, but it is particularly important when it comes to designs that require a before-and-after measurement. Otherwise, commissioners and evaluators risk becoming initiators and performers of, in short, badly conducted evaluations. There are some empirical findings indicating that this is the actually case. The commissioners emphasised that they prefer evaluations of effects, while in two of the four cases the evaluators tried to adapt to such demands by using a before-and-after design. In both cases, the evaluators were rather dissatisfied with how the evaluations turned out. It is problematic if evaluators, scientifically educated and highly experienced, agree to use a particular design even though that may not be possible because of contextual conditions. Researchers and evaluators thus seem to be willing to compromise on professionalism in order to receive funds.

Does the evaluation design matter to the commissioners in the end? Is it because of the lack of proper evaluations of effects that the civil servants do not pay the reports any attention? The civil servants at the Ministry of Health and Social Affairs report other reasons as well for not taking part in the evaluations, such as they do not consider it to be part of their duties. Instead, they suggest that it is the National Board of Health and Welfare that really should read and analyse the evaluation reports (see Petersén & Olsson, 2014). Another result from the interviews reveals that the evaluation reports may be of no significance at all to the Ministry. One of the respondents representing the Ministry stated that when they finally received the evaluation report, there were new issues that were high on the agenda, and the old ones had become irrelevant. The evaluations are, as the respondent put it, considered as passé. The time aspect is obviously an important reason that evaluations are not used, which has also been pointed out in earlier research (Petersén & Olsson, 2014; Weiss, 1998). The gap between the time when the evaluation is commissioned and the point when the final results are delivered seem too big to handle. The critical moment has passed; governments may even have been replaced by others with new political ambitions, and there are fresh, new interventions to implement. This is also distinguishing for the politics of temporary programmes in Sweden, which have been criticized for being unpredictable and short sighted (Fäldt, Storbjörk, Palm, Oscarsson, & Stenius, 2007; Statskontoret, 2007; Socialstyrelsen, 2009). People with insight into project politics, such as the civil servants and evaluators within this thesis, are probably well aware that a programme
financed by temporary subsidies will not be continued. This may decrease their motivation to use the evaluation findings. Knowledge derived from such programmes is hence hindered from being acknowledged and utilised because of its temporary character.

There are some more aspects, also concerning time and the temporariness of the programmes, worth recognising. The respondents, both evaluators and some of the civil servants, told about the difficulties of evaluating temporary programmes, as these sometimes are terminated, and thereby ‘forgotten’ in practice, long before the evaluation has even started. The interviews also reveal the converse, that evaluations may be started and perhaps even finished before there any possible effects to discern. A final explanation of poor evaluation use of large-scale evaluations may depend on the receivers’ abilities to understand and assess the evaluation reports. The interviews expose uncertainty among the civil servants at the Ministry and the Board as to how to interpret the results (cf. Jacobsson & Sundström, 2001; Mcdavid & Hawthorn, 2006). One of respondents, for instance, called the evaluations ‘too theoretical’. Such uncertainty may give a clue to the reports’ limited user-friendliness and open up room for other innovative ways to present the results.

Let us take a closer look at the actual use of the evaluations. In article I, we confirm, just like several studies before (e.g. Patton, 2008; Weiss & Bucuvalas, 1980), that evaluation use is very limited indeed. Earlier research has showed that decision-makers and other potential users prioritise their own ideology and interests, and the programme’s viability, before evaluation results (Shadish, Cook, & Leviton, 1991). In our interviews, none of the informants could give any concrete examples of instrumental use, which is to say, use as an immediate consequence of the evaluation. However, there is a glimmer of light concerning utilisation, and that is the evaluators’ own use of their findings. In all interviews, the evaluators stated that they had used the evaluation in one way or another. In three of the four cases, scientific articles had been written and published, and lessons learned from the evaluations were incorporated into teaching and passed on further to students. It should be added that it is the evaluators from the universities who have used the evaluations in this way, probably as a result of their inherent opportunities to write and teach. One of the interviews reveals that the evaluator thought of the task as a way to ‘do research’, and it implies that he was conscious of the possibilities it would bring even before he received the commission to do the evaluation. In the interviews, ‘doing research’ (evaluating) is regarded in the same way as being financed to spend
working time on the evaluation task, writing articles and books, and maybe being able to finance a doctoral student. These items can be the motivation to secure an evaluation task, perhaps more than the actual purpose of the evaluation at hand. In all, evaluators’ use of evaluation helps to broaden our understanding of the potential scope of utilisation. It is, however, not an entirely positive picture, and we need to reflect upon how different knowledge forms should guide social work.

The normative perspective: A combination of knowledge forms

From the previous section, we can draw the conclusion that use of evaluations of temporary programmes in Sweden leaves a great deal to be desired. If government persists in treating evaluation as an important element in public administration, evaluation’s content and use must be problematised from different aspects. Here, focus will be on what kind of knowledge should preferably guide social work and its potential consequences for evaluation use.

In article II the debate over applied knowledge in social work is highlighted. Social work is often described as a field in need of both scientific knowledge and practical knowledge (cf. Nordlander, 2006), and in the article it is stated that applied knowledge must reflect evidence as well as relevance (Petersén & Olsson, 2015). The point of departure for the critique is that much focus in the last decades has been on evidence – based on the knowledge form episteme – on behalf of relevance. At the same time, social workers have been criticized for overlooking scientific knowledge in their work. As a consequence, the National Board of Health and Welfare in Sweden started to implement EBP in social work. Using an evidence-based approach was supposed to increase social workers’ use of scientific knowledge, and thereby, practice would come to be based on science to a larger extent than before. In the article we object to a one-sided trust in scientific expert knowledge as the solution to human and social problems, and we remind the readers of another, contrasting form of knowledge, phronesis, which refers to practical knowledge. In social work, which is a multifaceted field of human activities, such knowledge that aims to understand and react to the specific and unique in a complex situation is needed. Phronesis requires the ability of good judgement of how to know what is the right thing to do. Questions of human practice cannot be studied under the same conditions as scientific–theoretical questions. Human actions cannot be explained or understood strictly scientifically, and instead of searching for ‘a truth’, re-
searchers should search for ‘the most likely’. Depending on the viewer, epistememe and phronesis can be considered to either compete with each other, or strengthen each other. However, scientific knowledge has by tradition been regarded as superior to practical knowledge (Brante, 2014), and one reason for that is because it is believed to be able to provide evidence.

In the second article it is claimed that social work should preferably adopt a praxis-based knowledge (PBK) strategy, which means a combination of different forms of knowledge and an upgrading of practical knowledge in particular (Petersén & Olsson, 2015). Social work operates on a field that is dynamic and sensitive to changes. It is impregnated with values and power relations (Healy, 2005) and includes individuals, groups, and organisations. To fulfil social work’s task of overcoming social injustices and supporting empowerment, social change, and development (cf. the International Federation of Social Workers’ global definition of social work), social workers will have to use their entire available set of knowledge forms and skills, which can be summarised as their ‘professional knowledge’ (cf. Hudson Drury, 1997). Practical knowledge and personal characteristics that can be ascribed to a social worker are not enough for performing professional social work; however, they are necessary for using scientific knowledge in practice. There is a danger in using scientific knowledge without further reflection. Studies have shown that social workers refer to evidence-based methods and manuals, or even such as they perceive to be evidence based, as crucial for informed choices and decision-making (see, e.g., Herz, 2012, 2016). Thereby, they also have their backs protected from any potential criticism, which seems to have become more and more important within NPM. Accountability and someone to hold responsible are central elements of NPM, and this, in combination with citizens’ increasing tendencies to give feedback on public services, may be a strong reason why social workers like to defend their work with references to scientific evidence. The upgrading of phronesis has much support in the literature. Many scholars plead for practical wisdom to take greater place, as they see a risk for social workers to distance themselves from the client-oriented work if they solely rely on and consult what they regard as an assured epistemic standard of knowledge (cf. Delmar, 2012). There are also many researchers who are positive towards EBP and still raise warning fingers for allowing EBP to be interpreted as always putting scientific evidence first. Brown (2014, p. 119) writes:

True expertise in evidence use ... provides a vision of policy makers as social actors who intuitively develop responses to situations. They will have
an intuitive holistic reading, based on the amalgamation of the formal knowledge they have adopted to date, an understanding of the specific case they are dealing with, and their understanding of other environmental factors that might be influencing the policy decision.

This ‘understanding’ Brown stresses comes from practical wisdom and is necessary for using scientific knowledge. Otherwise, social workers could be replaced by robots as well. If phronetic expertise were to continuously be associated with use of evidence, the distance evidence-based methods risk creating between social workers and their undertaking of social work could be bridged. The field of social work operates in a critical space, wherein social workers can have substantial influence, or even be life-changing, for people and groups that are discriminated against, socially excluded, and marginalised in various ways. Using and producing knowledge for learning and improvement in social work cannot be disclaimed from taking values and power seriously, not even if one is using scientific evidence.

The constructive perspective: A phronetic knowledge strategy in social work evaluation practice

On the basis of what has emerged due to the insights and results from the empirical and normative perspectives, I have sought an answer to the question of how evaluations in social work can best provide relevant and useful knowledge. I have come to some conclusions:

- Programmes and interventions in social work should be evaluated by incorporating a praxis-based vision of knowledge.
- Doable evaluations are useable.
- There are implications for practice and research on how to make evaluation matter in social work.

Incorporating a praxis-based vision of knowledge in evaluation

The first point, that programmes and interventions in social work should be evaluated by incorporating a praxis-based vision of knowledge, is concretised in article III, where a new evaluation model called phronesis-based evaluation (PBE) is presented (Petersén, forthcoming). The background of developing a model is that I hold evaluation to be a powerful force of influence in social work, as long as it can contribute relevant and useful knowledge. PBE takes departure in phronesis, and advocates a view that takes social work’s dynamic context into serious consideration as well as
acknowledging values and power as important components of the evaluation process. The central elements of PBE have five aspects, as follows: starting in practical social work, using value-rational questions, acknowledging values and power, being open-minded concerning choice of methods, and taking evaluation use seriously. A central point of departure for PBE is that the field of social work is continuously in need of new knowledge in order to fulfil its task of working for social change, development, protection, empowerment, and social justice. However, most current evaluation findings do not seem to be relevant and useful (see Petersén & Olsson, 2014). If we then think it is a good idea to make some changes in knowledge production, PBE may be a successful way to recover a doubtful evaluation practice. By adopting a phronetic approach, abstract methodological thinking is set aside, and focus is instead on how things can be improved in practice.

PBE also means handing over the interpretative prerogative of useful, relevant knowledge to social workers and people making use of social work. Evaluation practice, as it appears today, is very much about conducting evaluations commissioned and influenced by those in positions of power. In other words, those in power commission new knowledge, receive it, and put it on the bookshelf. Those with less or no power do not get the same access to engage in the knowledge production, and neither are they given the same possibilities to use it (cf. Nutley et al., 2007). Evaluation in social work many times concerns politically weak groups in society, which is to say, people who cannot be expected to change the state of things through the exercise of their own power (cf. Socialstyrelsen, 2009). With this awareness at hand, poor evaluations and minimal utilisation seem even more problematic.

Acknowledging values is fundamental in PBE, and in many other models with more or less constructivist approaches (see House, 1978). In social work, there are several reasons to discuss value-related questions, and that is specifically what PBE takes into account. The reasons can be expressed as new fields to study. Social work is constantly challenged by ‘new’ social problems, ‘new’ groups of people in need, and ‘new’ kinds of norm-breaking behaviours. This can be exemplified further, concretely, by migrants from Eastern Europe begging in Swedish cities; people belonging to the LGBT population, whose rights have received attention in the last couple of years; and the increasing numbers of unaccompanied refugee minors. These are three examples of groups that we do not know much about, but their situations are surrounded by a discourse marked by stereotypes and normative assumptions. Evaluation can be applied as a tool for elucidating the
service users’ situations, including from their own perspectives. At best, in a true democratic spirit, clients should be engaged early in the evaluation process, in order to inspect the evaluation questions and design in relation to their interests.

I have put forward the notion of PBE as a model to meet social work’s need for praxis-relevant knowledge. Relevant knowledge is here interpreted as knowledge that can provide learning and improvement for change in practice. PBE is thus a constructive contribution to the debate in social work on applied knowledge, and how to produce it.

**Doable evaluations and useable evaluations**

The lack of evaluation use, stated in article I (Petersén & Olsson, 2014), is both remarkable and not. In comparison to earlier research, the result was expected. On the other hand, here are four different cases of evaluations that cost millions of Swedish crowns. It would be disrespectful to say that lack of use is not outstanding in any way. I have touched upon a number of reasons of why evaluations are not used to any particular extent, and in this section I will try to make the arguments clearer. Much of the text in this thesis circles around episteme and phronesis, which can easily be (mis)understood as episteme versus phronesis. The point is, however, not to make any judgements about what kind of knowledge form is ‘the best’, but to elucidate how social work’s quest for knowledge has become equivalent to a hunt for ‘best evidence’ or ‘what works’. This has affected evaluation in social work in several ways: evaluations shall preferably be evaluations of effect, with randomised controlled trial (RCT) designs, or at least some kind of before-and-after measurement. Nevertheless, social work practice seldom provides the necessary conditions for setting up an RCT. One lesson learned here, which reminds us of what I wrote in chapter 2 concerning temporary programmes and subsidies, is that many programmes are approved with only a very short time between their formulation and the allocation of funds. The opportunities to plan for an evaluation are limited, especially if one wants an RCT. Even in this phase of the evaluation process the circumstances have basically made it impossible to do effect evaluations. A second lesson is that the social work context is dynamic and difficult to control. Seldom can one say that nothing other than the intervention has affected its target. Despite this, commissioners continue to pressure evaluators to do effect evaluations. Evaluators, in turn, seem to accept working under these doubtful conditions in order to secure the evaluation task. This vicious circle ends up wasting the potential of evaluations. The result is neither a good
effect evaluation nor any other good version of evaluation. This, the commissioners discover when the evaluators write in the evaluation reports that the results are not valid. As far as I can see, there are two possible solutions to this problem. Both solutions hold the idea of improving the conditions for doable evaluations, either by providing opportunities to conduct proper effect evaluations in those cases where other circumstances permit that, and the evaluation’s purpose is to find out about effects; or by upgrading other evaluation designs that can be put into practice without high demands of before-and-after measurements, control groups, and controlled environments. Of course, this would be made easier if evaluators would stop playing the game simply to receive funds.

The way of dealing with evaluation described above has been heavily criticized by scholars all over the world. Much of the criticism consist of arguments suggesting that studies focusing solely on effects do not give any information about what happens during the process. What about the quality and content of the intervention in itself, and not just the final results? Drawing upon Aristotle’s notion of phronesis, the quality of an intervention depends on both the final results and how the intervention is implemented. In fact, Aristotle values knowledge about the process more highly, because the results are nothing less than the consequences of a cluster of well-executed actions. So, evaluations of effect generate knowledge of a specific form, namely, episteme. But, evaluations in social work are not limited to determining the effectiveness of something; they can also be a means of overcoming social injustices and underpinning social change (cf. Lishman, 1999).

A final reflection of doable and usable evaluations concerns the time aspect. Drawing upon article I, it is worth noting that the time aspect is crucial for whether the evaluations will be used at all (Petersén & Olsson, 2014). We cannot affect time, and we cannot stop the political activities awaiting an evaluation report. With this in mind, evaluations as phenomena should be designed and conducted to provide knowledge that can benefit the field of social work in ways other than just revealing effects or highlighting any eventual reached goals. This seems an especially good idea when it comes to evaluations of temporary programmes, as the programmes do not always last, anyway, and therefore, the knowledge may not be applicable in the specific context in which it occurred.
Making evaluations matter in social work

Much research reports a lack of use of evaluation results. In social work, we can talk about a general lack of knowledge use as well. The Swedish government has tried to improve the use of scientific knowledge in social work by implementing EBP. Nevertheless, it seems as findings from evaluations still are not used as evaluators intend them to. Currently, it does not seem as evaluations matter at all. Instead of unreflectively seek to apply the models of the natural sciences to social work, I suggest conducting evaluations in ways that matter to practice; we need to start in social work practice when evaluating issues concerning social work practice. The content and context of practice is too crucial to be overlooked. Social workers, who are one of the important groups of potential users, are not tabula rasa (see Leitton, 2003), but live and work within that context, and besides scientific knowledge they rely upon several other knowledge forms in their work. In the field of evaluation in social work, we also have to acknowledge issues of values and power to a greater extent than before. Social works’ uttermost aim is to work against social injustices and therefore evaluations of social work interventions need to rely on similar values and with similar focus. Finally, evaluators of social work have to engage in communicating evaluation results by means that facilitates use. Within PBE, I have tried to formulate strategies that can meet the deficiencies of current evaluation practice, and at the same time meet the needs of social work practice. It is one serious attempt with potential to make evaluations matter in social work.

Final remarks

In this study, I have developed a new evaluation model called phronesis-based evaluation, which addresses serious limitations in the evaluation practice of social work. This model can hopefully bring some novelty to the evaluation practice as well as to the academic debate on evaluation models and evaluation use. It fulfils a number of significant functions, such as addressing professionals’ and clients’ opinions seriously, taking values and power into account, and focusing on learning and improvement in practice. However, the model has still not been tested in real life, and therefore we cannot really identify its potential weaknesses or how its strengths will materialise in practice. I can, however, predict some challenges concerning PBE, whereof the first is that the model requires very engaged evaluators. The evaluators must have a curious and open-minded attitude, and be prepared to have a lot of face-to-face contact with stakeholders and to actually stay on the sites for the intervention (cf. Barley, 2011). Furthermore, the
model places high demands on ingenuity concerning communicating the results, and also requires a touch of educational talent when taking responsibility for trying to maximise evaluation use. It is my belief and hope that the model can help the field of evaluation to become more creative concerning communication and utilisation.

Is there a place for PBE in a world favouring evidence of ‘what works’? Many scholars and politicians have been clear that NPMs’ influence in society must decrease. With such a development there is a good chance that we will once more start to appreciate knowledge that comes from ‘below’, context sensitive and grounded in peoples’ real-life experiences. Despite the fact that PBE has not been tested yet, the model has good prospects of becoming a fundamental part of evaluation practice in social work. It builds on ideals of classical knowledge and develops them to fit the dynamic context of social work. It opens up room for using intersectional analysis on complex social problems and is a tool for challenging power in a value-impregnated field. It clearly is an alternative for commissioners and evaluators holding the values of democracy and participation in high regard.
References


94 ANNA PETERSÉN *Evaluations That Matter in Social Work*


Bergmark, Å. & Lundström, T. (2007). Att studera rörliga mål. Om villkoren för evidens och kunskapsproduktion i socialt arbete [To study mobile targets: on the conditions for evidence and knowledge production in social work]. Socionomens forskningssupplement, 19, 4-16.


Eliasson, R. & Sunesson, S. (1990). Om utvärderingens ideal och realiteter [About the ideals and realities in evaluation]. In R. Eliasson, C. Levin,


Hood, C., James, O. & Scott, C. (2000). Regulation of government: has it increased, is it increasing, should it be diminished? *Public Administration, 78*, 283-304.


Månsson, S-A. (2000). Kunskapsutvecklingen inom socialtjänsten och den akademiska forskningen: drar vi åt samma håll? [Knowledge development within the department of social services and academic research – are we drawing the same direction?]. *Socionomen, 8*, 4-9.


Nordlander, L. (2006). *Mellan kunskap och handling. Om socialsekretares kunskapsanvändning i utredningsarbetet* [Between knowledge


Richard, H. & Emslie, C. (2000). The ‘doctor’ or the ‘girl from the University’? Considering the influence of professional roles on qualitative interviewing. *Family Practice*, 17, 71-75.


c


Appendix I

Interview guide – civil servants

Background information
1. Name
2. Title
3. Working duties
4. Earlier working experiences
5. Tell me about your work here at the Ministry of Health and Social Affairs/the National Board of Health and Welfare.

The programmes
1. Tell me everything you know about how a temporary programme is initiated and planned.
   - what characterizes a temporary programme?
   - who are involved (different professions)?
   - what are the ideas behind a programme? Ideology? Research? Theoretical ideas?

The evaluations
1. Tell me about the evaluations that belong to the programmes.
2. Why do you want the programmes to be evaluated? What is the purpose?
3. Who decides what the evaluation will look like?
   - design
   - content
   - focus
4. How does it work when you are contracting an evaluator?
5. Are there any other actors involved in the evaluation processes (besides the interviewee and the evaluators)?
6. Tell me about your and your colleagues’ competence on evaluations.

Evaluation use
1. What happens after a finished evaluation?
2. How do you use the evaluation results?
3. Were there any plans of how to use the evaluation findings already when planning the evaluation?
4. Can you comment on the allegation that evaluations are not used, or not used as intended?

Clearance issues
1. Is there anything else you would like to mention due to everything we have talked about?
2. Can I contact you if I have any more questions later?
3. Do you have any colleague I should talk to in order to find out more about temporary programmes and evaluations?
Appendix II

Interview guide - evaluators

Background information
1. Name
2. Title
3. Working duties concerning evaluation
4. Earlier working experiences concerning evaluation

The evaluation
1. Tell me about the evaluation.
2. Which was the evaluation commission?
   - How was the evaluation task negotiated between you and the commissioner?
   - Was there a “clear” task or was it designed by you and the commissioner (others) in cooperation?
3. Did you complete the evaluation task as agreed?
   - Complications?
   - Changes?
4. Can you tell me about your relations to the commissioners during the ongoing evaluation?

After the evaluation
1. Were there any plans of how to use the evaluation findings already when planning the evaluation?
   - from you (the evaluator)
   - from the commissioners?
2. How was the evaluation received?
   - Any feedback from commissioners or stakeholders?
   - Any meetings or conferences related to the evaluation?
3. What do you know about the evaluation’s impact?
   - Has it been acknowledged?
   - Any known effects?
4. What expectations did you have on the evaluation’s impact?
5. Do you know anything about any eventual long-term effects? What?
6. Have you seen any citations or anyone referring the evaluation?
7. Do you have any earlier experiences of evaluation use? What?

Clearance issues
1. Is there anything else you would like to mention due to everything we have talked about?
2. Can I contact you if I have any more questions later?
Publications in the series
Örebro Studies in Social Work


14. Åkerström, Jeanette (2014). “Participation is everything”: Young people’s voices on participation in school life.*


*Doctoral thesis    **Licentiate thesis