

Fjerde Generations Evaluering

Cand. scient. pol.
Konsulent, www.policy.dk
Peter Gorm Larsen

Juni 2008.

Eksamens nr.:167884
Vejleder: professor Peter Dahler-Larsen
Masterafhandling i evaluering, Syddansk Universitet 2008.
60 sider eksklusiv forside, indholdsfortegnelse og litteraturliste.

Indholdsfortegnelse

	Abstract	3
1.0.	Indledning	3
1.1.	Undring over Fjerde Generations Evaluering	4
1.2.	Problemformulering, afgrænsning og metode	5
1.3.	Disposition	7
2.0.	Evalueringsteori	8
2.1.	Dansk evalueringsteori	8
2.2.	Udenlandsk evalueringsteori	9
2.2.1.	En metateoretisk begrebsramme	14
2.3.	Er evalueringsteori sand?	19
2.3.1.	Pragmatisme	21
2.4.	Er FGE evaluering?	24
2.4.1.	Evalueringsdefinitioner	26
2.5.	Opsamling	30
3.0.	Er FGE en dårlig evalueringsteori?	32
3.1.	Viden og FGE	32
3.2.	Værdier og FGE	40
3.3.	Sociale programmer og FGE	44
3.3.1.	Repræsentativt demokrati og FGE	44
3.3.2.	Magt og FGE	47
3.4.	Anvendelse og FGE	50
3.5.	Evalueringspraksis og FGE	53
4.0.	Konklusion	60
	Litteratur	63

Abstract

This dissertation is about, why a so-called fourth generation evaluation theory (FGE) by Egon G. Guba and Yvonne S. Lincoln from 1989 seldom is applied in practice. Could the reason be, that FGE is a bad evaluation theory? This hypothesis is the point of departure for this dissertation. It is very difficult to say, what is a bad theory of evaluation. Both because theory of evaluation is a new discipline with a lot of disagreements and because it is difficult for theory of science to say generally what is bad or good theory. After an overview of evaluation theory, a short passage through theory of science and definitions of evaluation, takes the analysis its point of departure from the conceptual frame developed in the book *Foundation of Program Evaluation* (1991) by Shadish et al. FGE has troubles regarding construction of knowledge. The main reason is an unclear concept and definition of constructivism. The proposed methodology is not in accordance with constructivism. It has further repercussion. It opens up for relativism and solipsism, but nevertheless emphasises FGE empowerment and support for weak groups. FGE has no theory of value. FGE makes no considerations regarding power, theory of social programmes and how they function in representative democracies. Social programmes are only mental constructions. Its adherence to consensus between stakeholders makes FGE too unrealistically. FGE has a theory of use, but it is confined to a narrow context. There is only little evidence of use of FGE in practice. A review of literature shows common problems with FGE. It is difficult to use FGE in practice especially the so-called hermeneutic dialectic process, where stakeholders negotiate claims, concerns and issues regarding the evaluand. The final conclusion is that FGE is a bad evaluation theory. This can properly explain why it is not used very much in practice. This does not mean that FGE can't be developed and adapted for use in the future.

1.0 Indledning

I 1989 skriver Egon G. Guba og Yvonna S. Lincoln bogen "*Fourth Generation Evaluation*" (FGE). Første Generations Evaluering handler om måling (ibid.22-26). Den er udviklet i begyndelsen af det 20. århundrede. Fokus er på test af f.eks. intelligens og skolebørns præstationer. Tal og statistiske metoder er vigtige redskaber. Måling og evaluering bliver synonymmer (ibid.26). Første Generations Evaluering praktiseres stadig (ibid.). Anden Generations Evaluering handler om beskrivelse (ibid.27-28). Det er beskrivelser af programmets styrker og svagheder i forhold til opstillede mål (ibid.28). Ifølge Guba og Lincoln kan Tredje Generations Evaluering betegnes som bedømmelse (1989:29-31). Evaluator foretager en bedømmelse i form af anbefalinger som fortæller, hvordan programmet kan blive bedre.

Med baggrund i Stakes Responsive Evaluering (ibid.38) og postmoderne konstruktivistisk tænkning (ibid.39) udvikler Guba og Lincoln (1989) en evalueringsteori, som de benævner FGE.

Guba og Lincoln gør op med troen på, at der findes en objektiv virkelighed, som evaluator med videnskabelige metoder kan kortlægge (1989:passim). De tror ikke på objektive facts, som kan isoleres fra subjektive værdier. At teori kan adskilles fra beskrivelser og observationer. At der både kan eksistere et teoretisk- og et observationsprog. De mener ikke, at evaluator kan gøre teoretiske begreber operationaliserbare og dermed udtale sig om årsags- og virkningsforhold. Guba og Lincoln (1989) gør op med traditionelle videnskabelige kvalitetskriterier, som validitet, reliabilitet og generalisering. De mener, at virkeligheden konstrueres af mennesker med henblik på at skabe mening. At der i moderne pluralistiske samfund er mange forskellige værdier, og derfor mange forskellige konstruktioner. Da mennesker er meningsdannende og handlende, tror de ikke på nomotetisk

samfundsvidenskab. Menneskers konstruktioner er knyttet til en bestemt kontekst i tid og rum. De mener, at ny information kan skabe nye konstruktioner.

På det praktiske område gør Guba og Lincoln op med den fokusering, der er på opdragsgiverens interesser og opfattelser af virkeligheden, når der evalueres (1989:passim). De mener, at traditionel evaluering sker på bekostning af andre interesser, som bliver ramt af evalueringen. Guba og Lincoln (1989) finder, at den myndighed eller ledelse, der bestiller en evaluering, har en særlig magt, som betyder, at den selv kan holde sig uden for evalueringen (ibid.32). De foreslår, at alle der påvirkes af evalueringen, skal inddrages i evalueringen. Interessenterne skal have mulighed for at udtrykke deres påstande, bekymringer og problemer i forhold til det, som evalueres (evaluanden). Evalueringen afsluttes med at interessenterne forhandler om det, de er uenige om med henblik på at skabe en fælles konsensusopfattelse af evaluanden.

Guba og Lincoln (1989:passim) vil skabe deltagelse og indflydelse hos interessenter, som normalt ikke bliver hørt. Evaluators rolle er bl.a. at støtte svage interessenter, men også at organisere og facilitere forhandlingerne mellem interessenterne. For Guba og Lincoln (1989) er evaluering en fortløbende proces, som aldrig bliver færdig, da der konstant udvikles nye konstruktioner. Guba og Lincolns model bidrager til læring. De mener, at forhandlingsprocessen mellem interessenterne skaber mere informerede og sofistikerede konstruktioner af virkeligheden.

1.1. Undring over Fjerde Generations Evaluering

Forfatteren undrer sig over, at mange evalueringer i Danmark, ifølge Guba og Lincolns terminologi (1989), stadig er Første, Anden og Tredje Generations Evalueringer, når det er mange år siden, at verden fik kendskab til Fjerde Generations Evaluering. Skyldes det, at der er problemer med Guba og Lincolns evalueringsteori? At det er en dårlig evalueringsteori?

Der findes ikke en samlet og aktuel opgørelse over de forskellige typer af gennemførte evalueringer i Danmark. Ifølge Hansen mener Hansson, at meget evaluering består af: *"empirisk baserede statistiske beskrivelser af forløb og processer"* (Hansen 2005:33). Det svarer til Guba og Lincolns Første Generations Evaluering, hvor fokus er på måling.

I en større, men ældre analyse af evaluering i Danmark, skriver Hansen, at den såkaldte målopfyldelsesmodel præger mange områder af evalueringen i Danmark (Hansen 2000:173). Det svarer til Guba og Lincolns Anden Generations Evaluering, hvor fokus er på beskrivelser af programmer i forhold til opstillede mål.

Dahler-Larsen har i 2002 foretaget en undersøgelse af amters og kommuners evalueringspraksis (Dahler-Larsen 2002A). Undersøgelsen viser bl.a., at de vigtigste hensyn ved iværksættelsen af evaluering er at vurdere effekter, undersøge brugertilfredshed, forstå indsats-virkning og effektiv ressourcefordeling (ibid.42). De mindst vigtige hensyn er bl.a. organisatorisk læring, personalets kompetenceudvikling, undersøge borgertilfredshed og høre interesser (ibid.43). Det er ejendommeligt, at hensynet til at høre interessenter ikke spiller en stor rolle. Dahler-Larsen (2000A) skriver: *"den beskedne prioritering af dette hensyn er interessant i lyset af, at det spiller en ganske væsentlig rolle i meget evalueringsteori og i udviklingen af en hel del evalueringsmodeller, som netop tilsigter at give stemme til brugere og andre interessenter, som ellers formodentlig ikke høres så meget i konventionelle styringsprocesser"*(ibid.44). Guba og Lincolns forslag om at inddrage interessenter i evalueringer spiller ikke nogen stor rolle i kommunal evalueringspraksis.

En Google søgning på Internettet i marts måned 2008 på danske hjemmesider viser under 10 hits på *"4.generations evaluering"/"Fjerde generations evaluering"*. I dansk sammenhæng har Krogstrup (2003) benævnt Guba og Lincolns FGE for interessentmodellen. En søgekombination af *"interessentmodellen"* og *"evaluering"* giver 100 hits. Oversættelsen til interessentmodellen fanger dog ikke hele budskabet i FGE. Interessentmodel anvendes også i andre betydninger end i en evalueringssammenhæng. En søgning på *"konstruktivistisk evaluering"* giver kun 5 hits. Til sammenligning giver *"virkningsevaluering"* 691 danske hjemmesider. Sammenfattende er der kun få danske hjemmesider om FGE.

Paradokset om manglende brug af den sidste nye generation af evalueringsteori (FGE) er udgangspunktet for masterafhandlingen. Hvis vi skal følge Guba og Lincolns logik, burde de tidligere generationers evalueringsteorier være udfaset. Det er mange år siden, de skrev bogen om FGE. Skyldes det videnskabsteoretiske og metodologiske problemer med FGE? At FGE er vanskelig eller umulig at anvende i praksis? I korthed, at det er en dårlig evalueringsteori?

1.2. Problemformulering, afgrænsning og metode

Afhandlingens problemformulering lyder:

Guba og Lincoln har beskrevet Første-, Anden- og Tredje Generations Evaluering i bogen *"Fourth Generation Evaluation"* fra 1989. I bogen har de udviklet en særlig Fjerde Generations Evaluering. Hvorfor er Fjerde Generations Evaluering ikke anvendt mere i praksis?

Problemformuleringen hævder, at der er tale om et paradoks (Buch et al. 2004:17) eller en anomali (Adolphsen & Qvist 1994:25). Vi burde forvente, at FGE fra 1989 vil være udbredt i 2008 i forhold til tidligere generationer af evalueringsteorier. I problemformuleringen påstås a priori, at FGE ikke er særlig udbredt i praksis. Påstanden er et aksiom, der ikke undersøges empirisk udover, hvad der dokumenteres i afsnit 1.1. og afsnit 3.5. om evalueringspraksis.

Det er et interessant problem, at FGE er en evalueringsteori, der lægger stor vægt på at inddrage interessenter, ikke anvendes mere i praksis jf. citat af Dahler-Larsen i afsnit 1.1. Afhandlingen beskæftiger sig ikke med tidligere generationers evalueringsteorier eller nyere teorier. Det undersøges ikke om manglende brug af FGE skyldes, at opdragsgivere, konsulentfirmaer eller evaluatorer ikke kender til FGE. Afhandlingen fokuserer i stedet på karakteristika ved FGE, som kan forklare manglende brug af FGE.

Afhandlingens disposition og analyse tager delvis udgangspunkt i Shadish et al.'s disposition og analyse af evalueringsteorier fra 1991. De skriver i bogen *"Foundations of Program Evaluation"* (1991) at: *"New theorists constantly emerge whose work could be included in a book like this, such as.....Lincoln and Guba...If a second volume like this were written, they would be prime candidates for inclusion"* (Shadish et al. 1991:65). Afhandlingen vil tage tråden op og forsøge at give et bud på en analyse af Guba og Lincoln (1989).

En del danske bøger om evaluering bygger på Shadish et al. jf. afsnit 2.1. Shadish et al. gør rede for, hvad der bør kræves af en god teori om evaluering. Ifølge Shadish et al. skal en god evalueringsteori sige noget om følgende fem forhold: 1) sociale programmer (genstandsfeltet), 2) viden, 3) værdier, 4) anvendelse og 5) evalueringspraksis. Skabelonen anvendes af Krogstrup (2003) og Dahler-Larsen (2003). Når sidstnævnte forelæser over evalueringsmodeller, konfronteres hver enkelt model med skabelonen. Hvad siger en evalueringsmodel om videns-, værdigrundlaget, anvendelse, evalueringsgenstanden og om evalueringspraksis? Hvis ikke en model kan forholde sig klart til de fem punkter, er det ifølge Dahler-Larsen ikke en god model (Dahler-Larsen 2006A). Shadish et al. mener, at det ikke er god evalueringsteori. De gør opmærksom på, at ingen af de syv teorier de analyserer, siger noget om alle fem forhold (1991:37). Bemærk, at Shadish et al. skriver om evalueringsteorier. Dahler-Larsen (2006A) taler om modeller. I kapitel 2 redegøres for, hvad der forstås ved evalueringsteori.

Hvis ikke Guba og Lincolns teori behandler de ovenfor nævnte fem forhold, er det ikke en god evalueringsteori. Det kan forklare den manglende brug af FGE i praksis. Afhandlingens tese lyder:

Fjerde Generations Evaluering anvendes kun lidt i praksis, fordi det er en dårlig evalueringsteori.

Det er ikke hensigten, hverken at verificere eller falsificere tesen i Poppers forstand (Knudsen 1990:99-111). Med inspiration fra analytisk generalisering (Kvale 1997:228-229), kommunikativ og håndværksmæssig validitet, transparens, heuristik og pragmatik (Dahler-Larsen 2002B), er ambitionen blot ved analyse og logisk argumentation at sandsynliggøre tesens plausibilitet.

Metoden til besvarelse af problemformulering og tese er analyse af litteratur om evaluering og FGE. Der genereres ikke selvstændig empiri. Centrale kilder er Guba og Lincolns forfatterskab i bøger og artikler, boganmeldelser af FGE og debat omkring FGE i internationale tidsskrifter. Der inddrages også artikler, der beskriver erfaringer fra gennemførte evalueringer med udgangspunkt i FGE. Vægten i afhandlingen er på en teoretisk analyse af, om FGE er en dårlig evalueringsteori. Hvis det er tilfælde forudsættes det aksiomatisk, at det kan forklare, hvorfor FGE ikke anvendes mere i praksis.

1.3. Disposition af afhandlingen

Afhandlingen er opdelt i fire kapitler. Et indledende kapitel (1.0.) som består af en kort introduktion til FGE, en undring (1.1.), problemformulering, afgræsning og metode (1.2.) samt en disposition (1.3.).

Afhandlingens tese påstår, at FGE er en dårlig evalueringsteori, men hvad er evalueringsteori? Derfor er det nødvendigt i kapitel 2.0. at præsentere en generel introduktion og analyse af evalueringsteori. Der redegøres kort for dansk evalueringsteori (2.1.), som er stærkt påvirket af udenlandsk evalueringsteori (2.2.). Shadish et al.'s metateoretiske begrebsramme præsenteres og problematiseres (2.2.1.). Det analyseres om evalueringsteori er sand (2.3.). Det påvises, at pragmatisme er sine qua non for evalueringsteori (2.3.1.). Det diskuteres om FGE overhovedet er evaluering (2.4.) ved at inddrage andre teoretikers definitioner af evaluering (2.4.1.). Kapitlet afsluttes med en opsamling (2.5.).

I kapitel 3.0. konfronteres FGE med shadish et al.'s begrebsramme med henblik på at afdække tesen om, at FGE er en dårlig evalueringsteori. Der lægges ud med en analyse af viden og FGE (3.1.), efterfulgt af værdier og FGE (3.2.) og sociale programmer og FGE (3.3.). Afsnittet er underinddelt i et afsnit om repræsentativt demokrati og FGE (3.3.1.), og magt og FGE (3.3.2.). Der er et afsnit om anvendelse og FGE (3.4.), hvorefter analysen afsluttes med evalueringspraksis og FGE (3.5.). Under evalueringspraksis inddrages de erfaringer, der i praksis er gjort med FGE.

Afhandlingen afsluttes med en konklusion (4.0.).

2.0. Evalueringsteori

Afhandlingens tese er, at FGE kun anvendes lidt i praksis, fordi det er en dårlig evalueringsteori. For at undersøge postulatet kræves en forståelse af, hvad evalueringsteori er, og hvornår en evalueringsteori er god eller dårlig. Der redegøres kort for dansk evalueringsteori (2.1.), som efterfølges af et afsnit om udenlandsk evalueringsteori (2.2.) uden dog at der er vandtætte skotter mellem de to afsnit. En metateoretisk begrebsramme præsenteres (2.2.1.) og evalueringsteori konfronteres med sandhedskriterier (2.3.). Der dvæles ved pragmatisme (2.3.1.), og det diskuteres om FGE er evaluering (2.4.) ved at inddrage definitioner af evaluering (2.4.1.). Kapitlet afsluttes med en opsamling (2.5.).

2.1. Dansk evalueringsteori

Evalueringsteori og evalueringsmodel synes i dansk og skandinavisk sammenhæng at blive anvendt synonymt.

Dahler-Larsen og Krogstrup (2003) skriver, at: *"en evalueringsmodel er et princip eller et hovedspørgsmål, som styrer en evaluering. Hovedspørgsmålet er afgørende for, hvad evalueringen handler om. I daglig tale omfatter ordet evalueringsmodel også en opskrift på, hvordan evaluering i praksis udføres for at besvare dette hovedspørgsmål"* (ibid.7). Det er uklart, hvad der forstås ved princip eller hovedspørgsmål. Et bud i forhold til FGE er, at princippet i FGE drejer sig om at identificere interessenternes: *"claims, concerns and issues"* (Guba & Lincoln 1989:39) (oversat af forfatteren til påstande, bekymringer og problemer) omkring evaluanden med henblik på at skabe enighed hos interessenterne. Senere i bogen står der, at: *"en meget vigtig funktion af en evalueringsmodel er at vejlede i, hvordan man træffer alle de valg, som et evalueringsforløb indebærer. Det handler f.eks. om emnemæssigt fokus, metode, evaluators rolle og inddragelse af interessenter"* (Dahler-Larsen & Krogstrup 2003:7). Sat på spidsen nærmer citaterne sig en tautologi. En evalueringsmodel skal vejlede i de valg, som et evalueringsforløb indebærer. Men er et evalueringsforløb ikke udtryk for, hvordan en evalueringsmodel defineres? I forhold til FGE redegør Guba og Lincoln (1989) meget nøjagtigt for metode, inddragelse af interessenter og evaluators rolle. De præsenterer altså en god evalueringsmodel.

I *"Håndbog i evaluering"* af Rieper (2004) fremgår det klart, at der ikke er enighed blandt evalueringsforskere i forhold til, hvad en evalueringsmodel er. Rieper skriver, at: *"faglitteraturen indeholder mange forslag til typologier over evalueringsmodeller"* (2004:18). I bogen *"Nye Veje i Evaluering"* af Dahler-Larsen og Krogstrup (2003) beskrives tre modeller. Rieper præsenterer *"ti evalueringsmodeller"* (2004:18). Han mener ikke, at man nødvendigvis skal

vælge én model. På trods af at Vedung (1998) ikke er dansker, inddrages han her. Han opstiller en typologi over forskellige evalueringsmodeller. Han skriver: *"Det råder næmligen utbredd oenighet om utvärderingars inriktning. Et dussintal ansatser har blivit så utarbetade att de återfinns i litteraturen. Helst borde dessa ansatser slås samman till ett fåtal modeller"*(ibid.49). Derefter opstiller Vedung en række utvärderingsmodeller (ibid.50). Rieper mener, at evalueringer må *"skræddersyes"* (2004:18). Modellerne kan give inspiration og danne rammer for diskussion af: *"hvordan man kan skræddersy konkrete evalueringer"*(2004:18). Altså, FGE kan vurderes med udgangspunkt i, hvor god modellen er til at igangsætte en diskussion i forhold til at designe en konkret evaluering.

Hansen opstiller evalueringsparadigmer (2003). Et evalueringsparadigme er et: *"...grundlæggende sæt af anskuelser, der præger evalueringspraksis"* (ibid.22). Guba og Lincoln (1989) mener, at FGE er et nyt paradigme, som de kontrasterer i forhold til et konventionelt paradigme. Hansen finder, at der sker et uheldigt spring fra abstrakte paradigmer til konkrete evalueringsmodeller eller opskrifter (2003:29). Hun mener, at forskellige evalueringsmodeller er udtryk for forskellige designs. Modellerne rejser: *"forskellige evalueringsspørgsmål og fastfryser forskellige evalueringskriterier"*(ibid.). Evalueringsmodeller er: *"byggeklodserne i enhver evaluering"*(ibid.). Modeller kombineres ofte i konkrete evalueringer.

Krogstrup indfører begrebet *"informeret evaluering"* (2003:29), og henviser til Shadish et al. Hun mener, at informeret evaluering må indeholde overvejelser om genstandsfeltet, viden, værdier og anvendelse. De fire elementer udgør informeret evalueringspraksis. Det er uklart, hvad hun præcist tænker på i forhold til begrebet *"informeret evalueringspraksis"*. Er informeret evalueringspraksis, det samme som god evaluering eller en god evalueringsteori? Måske skal det forstås sådan, at hvis Guba og Lincoln (1989) har overvejelser om de nævnte fire punkter, er der tale om informeret evaluering. I sin bog beskriver hun FGE som interessentmodellen. Den får et særskilt kapitel (Krogstrup 2003:103-109). På baggrund af Krogstrups fire komponenter er det vanskeligt at slutte, at FGE er en dårlig eller god evalueringsteori. Det gør det ikke lettere, at Krogstrup nævner, at der er mange andre komponenter, der kan indgå, når der vælges model eller design. Hun nævner f.eks. etablerede evalueringsopskrifter, tids- og personalemæssige ressourcer samt evalueringskompetence (ibid.30).

2.2. Udenlandsk evalueringsteori

Udenlandske teoretikere er i tvivl om, hvad der skal forstås ved evalueringsteori. Det gør det vanskeligt at undersøge afhandlingens tese, som påstår, at FGE er en dårlig evalueringsteori.

Donaldson og Lipsey (2006) skriver, at i evalueringslitteraturen henvises der ofte til teori. De fortsætter med at skrive: *"but what is meant by "theory" encompasses a confusing mix of concepts related to evaluators notions about how evaluation should be practiced, explanatory frameworks for social phenomena drawn from social science, and assumptions about how programs function or are supposed to function"*(ibid.57).

Som præsident for American Evaluation Association udtaler Shadish i 1998: *"Evaluation Theory Is Who We Are"*(Shadish 1998). Det er just ikke en præcis definition af evalueringsteori. Den er ikke anvendelig, når FGE skal vurderes i forhold til, om den er en dårlig evalueringsteori.

Shadish (1998) giver en række argumenter for, hvorfor evalueringsteori er vigtig. Han knytter argumentationen til professionen. Evalueringsteori giver en identitet og et fælles sprog. Donaldson og Lipsey (2006) skriver: *"Knowledge of evaluation theory can also help evaluators become better ambassadors for the profession of evaluation and educators of potential clients"*(ibid.61). Evalueringsteori tænkes som PR for en profession! Patton gør i den kendte debat om evalueringers anvendelse med Weiss (Patton 1990; Weiss 1990A & B) det også klart, at evaluering handler om en profession. Teori eller ikke teori, ifølge Patton skal kunderne have at vide, hvad de betaler for (Patton 1990:201). Pattons vision er en evalueringsprofession, som er forpligtet på at levere et produkt med: *"...intended uses by intended users..."*(ibid.).

Scriven gør i bogen *"Evaluation Thesaurus"* (1991A) status over evalueringsteorien. Han stiller spørgsmålet: *"How far is evaluation from being a new discipline? If we set out some criteria for a new discipline, and expand them in some detail for the case of evaluation, it will be clear that there is still a considerable distance to go"*(ibid.24). Scriven anvender stor energi på at lægge en afstand mellem evalueringsteori og normal samfundsvidenskab. Evaluering er ikke værdineutral og objektiv *"applied social science"* (ibid.4). Det handler ikke bare om at indhente data (ibid.7). Scriven mener, at evalueringsteorien skal frigøre sig fra de begrænsninger, der findes i samfundsvidenskaben (ibid.34). Shadish et al. skriver, som noget af det første i deres bog: *"We want this book to counter the belief that program evaluation is just applied science methodology"* (ibid.7). Det er omvendt hos Rossi, Lipsey og Freeman, som i bogen *"Evaluation"* (2003) lægger stor vægt på samfundsvidenskabelige metoder. De skriver: *"social research methods provide an especially appropriate approach to the task of describing program performance"*(ibid.16). Endvidere, at: *"...commitment to the rules of social research is at the core of the perspective on evaluation presented in this book"*...(ibid.).

Scriven mener, at evalueringsteori er mere end, hvad programteoretikere kan præsentere (1991A:22-23,285). Evalueringsteori er transdisciplinær, og bidrager til mange forskellige discipliner. Evaluering er: *"the process of*

determining the merit, worth and value of things"(ibid.1). Menneskeheden har altid foretaget evaluering. Det gælder f.eks. håndværkere, der vurderer deres arbejde og forbrugertests af f.eks. motorcykler. Scriven mener, at der også er tale om evaluering, når elever og studerende bliver bedømt, eller når folk vælges ud til stillinger. Andre teoretikere har en snævrere opfattelse. Shadish et al. beskæftiger sig med evaluering af sociale programmer (1991:19). Rossi, Lipsey og Freemans (2003) bog handler også om evaluering af programmer (ibid.3). Ved programevaluering forstår sidstnævnte: *"The use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizational environments and are designed to inform social action in ways to improve social conditions* (ibid.431).

Det er vanskeligt at tale om en fælles overordnet evalueringsteori, når der ikke er enighed om, hvad genstandsfeltet er. Derfor er det svært at vurdere om en særlig evalueringsteori, som f.eks. FGE er god eller dårlig. FGE er måske dårlig til at evaluere motorcykler, studerendes præstationer og kontraktstyring af køkkener (Dahler-Larsen & Krogstrup 2003:51-79). Teorien er måske god til at evaluere tiltag, som skal skabe en positiv udvikling af unge mennesker i et område (Huebner & Betts 1999). Om FGE er god eller dårlig, kan være afhængig af, hvad der skal evalueres.

At evalueringsteorien har vanskeligheder i forhold til sig selv som disciplin, men også til samfundsvidenskaben, kommer til udtryk i kapitel 2 i *"Handbook of Evaluation"* (Shaw et al.2006). Her foreslår Donaldson og Lipsey et kryds mellem evalueringsteori, samfundsvidenskabelig teori og programteori i begrebet: *"program theory-driven evaluation science"*(ibid.66). De mener, at krydset mellem de tre teorier vil bidrage til: *"social betterment"*(ibid.).

Evalueringsteorien giver ikke en entydig anvisning på, om der skal anvendes samfundsvidenskabelige kriterier og metoder, når vi skal vurdere, om f.eks. FGE er en dårlig evalueringsteori. Afhandlingens tese kan lettere besvares, hvis det blot handlede om at undersøge om FGE overholdt: *"rules of social research"* (Rossi, Lipsey & Freeman 2003:16). Scriven tror ikke på doktrinen om værdifri videnskab (1991A:2). Som Guba og Lincoln (1989:163-172) påstår også han, at doktrinen længe har været død blandt intellektuelle (Scriven 1991A:2): Det er således ikke nok, at undersøge om FGE producerer *"objektiv viden"*, og dermed er en god evalueringsteori. Tilgangsvinklen er i overensstemmelse med den tidlige Campbell (Shadish et al.1991:119-170).

Det er svært at vurdere, om FGE er en dårlig evalueringsteori, når teoretikerne ikke er enige om, hvad formålet er med evaluering. Shadish et al.(1991) skriver: *"Intrinsic to evaluation is an idealized problem-solving sequence for (a) identifying a problem, generating and implementing alternatives to reduce symptoms, (c) evaluating these alternatives, and then (d) adopting those that results suggest will reduce the problem satisfactorily"*(ibid.20-21). Opfattelsen

af evaluering er instrumental, forudsætter rationalitet og fokuserer på beslutningssituationer. Opfattelsen er i overensstemmelse med f.eks. Patton (1990) og Wholey (Shadish et al.1991:225-269). Scrivens vurderinger af *"merit, worth or value"*(Mark et al.2006:6) handler ikke kun om problemløsning. Weiss (1990A & B) advokerer for, at evalueringer ikke nødvendigvis skal kobles til en konkret beslutningssituation, men kan avendes konceptuelt til at skabe generel oplysning i samfundet. Guba og Lincoln (1989) ønsker handling ved, at interessenterne inddrages og forhandler om: *"claims, concerns and issues"* i forhold til evaluanden (ibid.39). At interessenter inddrages er i samsvar med Stakes Responsive Evaluering (Shadish et al.1991:270-314). Ud over at skabe handling (Guba & Lincoln 1989:249) er formålet også, at interessenterne i processen danner mere informative og sofistikerede konstruktioner af evaluanden og hinandens opfattelser. Guba og Lincoln (1989) mener: *"Evaluation is a teaching/learning process"*(ibid.263). Endvidere, at der er tale om en: *"empowering activity"*(ibid.264), af tidligere: *"disenfranchised stakeholders"*(ibid.247).

Evalueringsteoretikerne er således ikke enige om formålet med evaluering. Det gør det vanskeligt at skabe en overordnet evalueringsteori, men også at sammenligne teorierne. FGE er måske en god teori, hvis vi ønsker empowerment og demokratiske processer, men en dårlig teori, hvis vi ønsker at give beslutningstagere information om, hvad de skal gøre her og nu.

Shadish påpeger nødvendigheden af en unik vidensbase med henblik på at skabe en fælles identitet blandt evaluatorene (1998:1). Han skriver, at det vigtigste ved en selvstændig evalueringsteori er: *"what makes us different from other professions"*(ibid.). For nye evaluatorene, som skal lære professionen at kende, handler det om, at lære at: *"talk the talk"* (ibid.3). Senere i artiklen antyder han, at evalueringsteori er det, evaluatorene gør i praksis (ibid.4). Dette er ikke et stærkt udgangspunkt, når der skal foretages en teoretisk analyse af FGE. Analysen af FGE kan således ikke tage afsæt i evalueringsteori, men i om FGE anvendes i praksis! Afhandlingens tese bliver, hvad man i logikken kalder tautologisk. FGE er en dårlig evalueringsteori, fordi den ikke bliver anvendt i praksis. Den bliver ikke anvendt i praksis, fordi det er en dårlig evalueringsteori! Shadish's (1998) tankegang betyder, at f.eks. Virkningsevaluering (Dahler-Larsen & Krogstrup 2003:51-79) er en god teori, fordi den bliver anvendt i praksis. Eller astrologi er en god teori, fordi horoskoper findes i mange tidsskrifter, og læses af mange mennesker!

Donaldson og Lipsey refererer til en distinktion, hvor der skelnes mellem deskriptiv- og præsriptiv teori (2006:58). De mener, at de fleste evalueringsteorier er præsriptive. De citerer Alkin for at evalueringsteorier: *"offer a set of rules, prescriptions, prohibitions, and guiding frameworks that specify what good or proper evaluation is and how evaluation should be done"* (ibid.58). Evalueringsteori er konventioner om, hvordan vi bør evaluere. Ifølge Scriven er det et første skridt i retning af en selvstændig evalueringdisciplin

(1991A:11). Men Scriven mener, at en disciplin er mere end: *"bag of tricks and bricks"* (ibid.24). Shadish et al. konstaterer, at man indenfor forskellige sektorer har evalueret på forskellige måder med udgangspunkt i forskellige fagdiscipliner (ibid.28-30). De skriver: *"This diversity continues to the present day, making it difficult for evaluators to agree on what the practice of evaluation should be and why"*(ibid.30). Måske er FGE i god tradition med den måde pædagoger og lærere tænker evaluering på. Omvendt vil FGE næppe passe ind i den måde man evaluerer på indenfor det medicinske område. Her er vægten traditionelt på randomiserede eksperimenter og kvantitativ metode jf. den tidlige Campbell (Shadish et al.1991:119-170). Når afhandlingen stiller spørgsmålet, om FGE er en dårlig evalueringsteori, kan svaret være afhængig af den sektor, man identificerer sig med.

Hvis evalueringsteori er præskriptiv, kan vi ikke forvente, at få et klart svar på om FGE er en dårlig evalueringsteori. I stedet kan vi undersøge, om FGE lever op til de konventioner, der findes i forhold til, hvordan man bør evaluere. Da Guba og Lincoln (1989) klart giver udtryk for, at de ønsker at gøre op med traditionel evaluering, vil FGE logisk set blive en dårlig evalueringsteori. FGE lever ikke op til konventionerne indenfor evalueringsteorien.

Når Guba og Lincoln (1989) har udviklet et forslag til, hvordan man skal evaluere, er det vanskeligt at foretage en analyse med afsæt i evalueringsteori. Årsagen er, at evalueringsteori ikke er statisk. I Shadish et al.'s analyse af klassiske teoretikere sker der en inddeling i tre stadier. Første stadiet af teorier handler om opdagelse af sandhed. Fokus er på: *"effective solutions to social problems"*(ibid.67). Anden stadiet handler om, hvordan evaluering bliver anvendt, og hvordan offentlige organisationer fungerer (ibid.). Tredje stadiet er en integration af teorierne fra de to tidligere stadier (ibid.). Guba og Lincoln (1989) har selv inddelt evalueringsteoriene i fire generationer. Der er teoretikere, der har udviklet et: *"Evaluation Theory Tree"* (Donaldson & Lipsey 2006:60). Evalueringsteoretikerne kategoriseres i forhold til en træstamme, som består af: *"accountability, social inquiry"* (ibid.). Grenene på træet består af: *"use, methods and valuing"* (ibid.).

Vi bør være kritiske overfor kronologiske inddelinger. De bliver let til evolutionistiske inddelinger. Det kan få karakter af, at teorier på tredje stadiet, fjerde generation, eller dem højt i trætoppen er bedre og mere komplicerede teorier end dem på første stadiet, første generation eller i bunden af træet. I en dansk sammenhæng synes meget evaluering at tage udgangspunkt i teorier fra første og anden stadiet, eller første, anden, tredje generation jf. indledningen (Hansen 2000;2003;Dahler-Larsen & Krogstrup 2001). Derfor kan vi ikke slutte, at det er dårlig, anakronistisk og antikvarisk evaluering.

Når evalueringsteori er flygtig i tid (stadier og generationer), rum (sektorer) og indhold, er det vanskeligt at undersøge om FGE er en dårlig evalueringsteori. Hvad er det i forhold til? Alligevel er det et relevant spørgsmål for

opdragsgivere og evaluatorer. Vi skal praktisk og teoretisk kunne stille spørgsmålet, om en evalueringsteori er god eller dårlig.

Scriven er klar over, at det er problematisk for evalueringsteorien, hvis den ikke kan give anvisninger på, hvad der er en god eller dårlig teori (1991A:29). Han skriver: *"No discipline can restrict itself to relativistic claims, and certainly not to relativistic evaluative claims. For, if researchers in a field cannot distinguish what is actually – rather than relativistically – a good theory or interpretation from a bad one, the field would have no way to distinguish between discipline and drivel. It would be a double absurdity to have a discipline of evaluation which could not go beyond relativistic conclusions"*. (ibid.).

En vis relativisme i forhold til evalueringsmodeller findes hos Krogstrup (2003:9), som er omtalt under danske evalueringsteoretikere i afsnit 2.1. Udgangspunktet for hendes bog er: *"at ingen modeller er bedre end andre, men de kan noget forskelligt."*(ibid.9). Vi præsenteres for en række modeller, men hvilken model skal vi anvende og i hvilken situation? Er FGE en dårlig evalueringsteori? Det fremgår ikke i Krogstrups sammenfatning af, hvad hun kalder for interressentmodellen (ibid.109).

Shadish et al. ønsker en anden situation for evalueringsteorien ved at opstille et ideal. De skriver: *"The ideal (never achievable) evaluation theory would describe and justify why certain evaluation practices lead to particular kinds of results across situations that evaluators confront. It would (a) clarify the activities, processes, and goals of evaluation; (b) explicate relationships among evaluative activities and the processes and goals they facilitate; and (c) empirically test propositions to identify and address those that conflict with research or other critically appraised knowledge about evaluation"*(ibid.31).

Ifølge Donaldson og Lipsey (2006) giver Shadish et al.: *"one of the most fully developed perspectives available on the nature of evaluations theory"* (Donaldson & Lipsey 2006:60). Som omtalt i indledningen lægges dette perspektiv ned over FGE. Som det er vist i afsnit 2.1., anvendes perspektivet også i stor udstrækning af danske evalueringsforskere. Perspektivet skal nedenfor analyseres og problematiseres før det bringes i anvendelse i forhold til FGE.

2.2.1. En metateoretisk begrebsramme

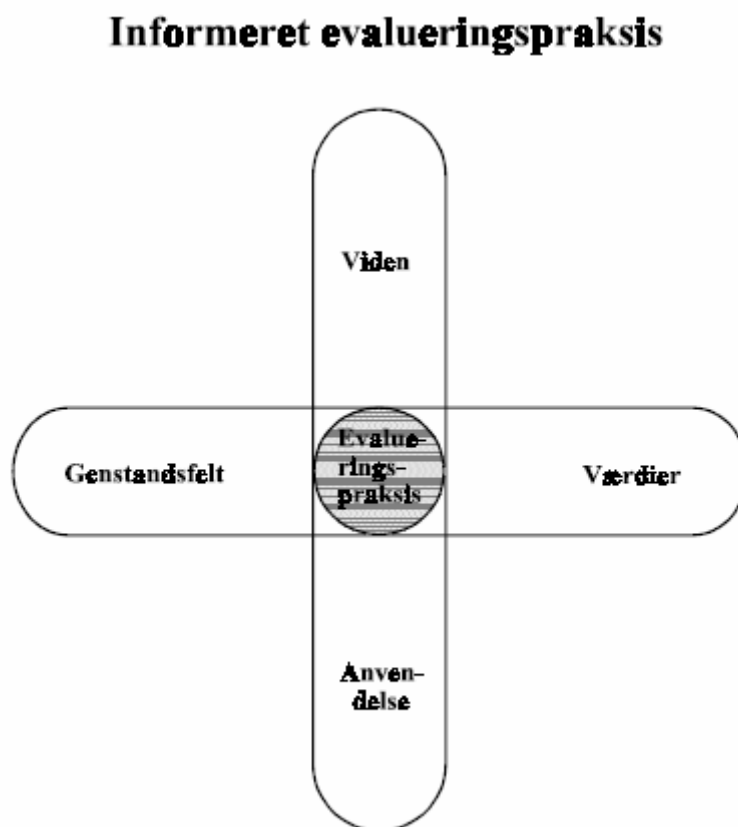
Shadish et al. mener, at en god teori om evalueringspraksis skal give en redegørelse om fem forhold. Evaluator får igennem de fem forhold svar på en række relevante spørgsmål. De er nedenfor søgt sammenfattet, forklaret og analyseret.

- 1) **Sociale programmer.** Det er teori om, hvordan sociale programmer udvikler sig, bliver bedre og ændre sig (Shadish et al.1991:32). Fokus er på sociale problemer, og hvordan sociale programmer kan skabe løsninger (Donaldson & Lipsey 2006:59). Hvad er det for vigtige sociale problemer programmet kan løse? Kan programmet forbedres? Er det værd at gøre dette? Hvad kan der ellers gøres?
- 2) **Viden.** Det er teori om, hvordan evaluator skaffer sig viden om, hvordan de sociale programmer fungerer, og personer agerer (Shadish et al.1991:32). Hvilke metoder skal evaluator anvende for at skabe troværdig viden (Donaldson & Lipsey 2006:59)? Hvordan kan evaluator vide, at evaluator ved det, evaluator har fundet frem til? Hvad er et troværdigt svar på det forrige spørgsmål?
- 3) **Værdier.** Det er teori om, hvordan sociale programmer vurderes (Shadish et al.1991:32). Hvordan foretager evaluator en bedømmelse? Hvad er et godt program, og hvad betyder godt (ibid.35)? Teorien skal give begrundelser for, at der kan foretages værdimæssige vurderinger.
- 4) **Anvendelse.** Det er teori om, hvordan den skabte viden anvendes til at modificere, ændre eller stoppe sociale programmer (Shadish et al.1991:32). Hvordan kan evaluator sikre sig, at resultaterne bliver anvendt hurtigt i forhold til programmet? Fordele og ulemper ved denne anvendelse? Kan evalueringen anvendes på andre måder?
- 5) **Evalueringspraksis.** Teori om taktikker og strategier samt begrænsninger evaluator skal følge i sit professionelle arbejde (Shadish et al.1991:32). I teorien kan evaluator få svar på spørgsmål som f.eks., hvordan gennemføres en evaluering i en situation, hvor evaluator har begrænsede færdigheder, tid og ressourcer (ibid.35)? Hvordan kan evaluator begrænse sine valg i en situation, hvor evalueringen kan løses på mange forskellige måder? Hvad er evaluators rolle? Lærer, metodologisk ekspert eller dommer? Hvilke spørgsmål skal evaluator stille, og hvilke metoder skal evaluator anvende?

Er de fem punkter ovenfor evalueringsteori? Donaldson og Lipsey skriver, at det er en ramme til at organisere teorier om evalueringspraksis (2006:59). Hvad skal rammen forklare? De forskellige teoretikers anbefalinger om, hvordan der skal evalueres? Om evaluatore kan hente en færdig opskrift, når de skal ud at evaluere? Skal rammen forklare evalueringspraksis? I så fald er det ejendommeligt, at evalueringspraksis selv indgår i de fem punkter, når den skal forklare evalueringspraksis! Shadish et al. finder, at teori om evalueringspraksis er den vigtigste komponent (ibid.57). De skriver: *"Theories of practice depend on the other four components of evaluation theory, because the content of the above list is drawn partly from the other components....Yet theory of practice is more than just the sum of content from the other components"*(ibid.58). Rammen indeholder fire punkter og så et femte punkt,

der bl.a. bygger på de fire punkter. Er de fire punkter lige vigtige, og bidrager de lige meget til det femte punkt, evalueringspraksis? Teoriens eller rammens uklarhed er måske årsagen til, at Krogstrup (2003:29-30) anvender begrebet "*informeret evalueringspraksispraksis*"(ibid.). Bemærk, at evalueringspraksis, som hos Shadish (Shadish et al.1991, Shadish 1998), er i centrum i hendes model jf. fig.1. Som Shadish et al. antyder, bygger evalueringspraksis på indholdet fra de fire andre punkter. Derfor sker der gentagelser og overlap i rammen. Som analysen af FGE i kapitel 3 viser, er det svært at vide, hvor f.eks. en magtanalyse skal placeres. Er det under sociale programmer, anvendelse eller evalueringspraksis?

Figur 1: Informeret evalueringspraksis



Kilde: Krogstrup 2003 p. 30.

Det komplekse i Shadish et al. 's ramme er, at der under de fire punkter, hhv. sociale programmer, viden, værdier og anvendelse, nævnes mange forskellige teorier. De kan f.eks. stamme fra politologi, organisationsteori, sociologi eller videnskabsfilosofi. Rammen låner og integrerer løst teorier fra mange forskellige discipliner. Opfattelsen er delvis i overensstemmelse med Scrivens ide om, at evalueringdisciplinen er transdisciplinær, forstået sådan at evalueringsteorien kan bidrage til andre discipliner, såvel som de kan bidrage til evalueringsteorien. Denne multidisciplinærhed, tværfaglighed eller

økumenisme gør det vanskeligt at vurdere, om Shadish et al.'s ramme og de enkelte evalueringsteorier er gode eller dårlige.

Kan rammen forklare om en evalueringsteori, som f.eks. FGE er dårlig eller god? Shadish et al. indrømmer: *"no theory we review deals explicitly and in detail with all five components"*(ibid.37). Betyder det, at vi i afhandlingen ikke skal forlange, at Guba og Lincoln (1989) redegør for alle fem punkter i rammen? Er teorien god, hvis der blot redegøres for to eller tre af punkterne? Hvordan er Shadish et al. i øvrigt kommet frem til rammen? Bygger den på teoretisk deduktion eller blot induktiv praktisk erfaring? Har evalueringsforskere en særlig forståelse for, hvad der skal forstås ved teori?

For Hellevik er teori: *"et deduktivt system av hypoteser"*(2003:472). Samme opfattelse findes hos Næss (1980:54). Inspireret af Poppers krav om falsificerbarhed (Knudsen 1990:99-102) skriver Næss: *"Når en teori eller et sett hypoteser er blitt utsatt for omfattende prøving uten at noen forutsigelser har slått feil, vil teorien ofte bli oppfattet som korrekt"* (Næss 1980:54). I en metodebog står, at: *"teorien er det redskab, der skal bruges til at forklare de beskrevne fænomener ud fra"*(Andersen & Gamdrup 1990:36).

Givet er det, at Shadish et al.'s såkaldte metaevalueringsteori eller begrebsapparat ikke kan udledes deduktivt. Shadish er sikkert klar over dette i 1998. Han skriver generelt om evalueringsteori: *"parts of evaluation theory are empirically based, but especially given the youth of our field, much more of it is hypothetical, conjectural, and unproven, still waiting to be tested"*(Shadish 1998:2). Bemærk, at Shadish ikke er fremmed for, at evalueringsteori kan testes empirisk. Kan vi teste dele af FGE med henblik på at finde ud af, om det er en dårlig evalueringsteori? Er det rigtigt, som Guba og Lincoln hævder, at FGE betyder større anvendelse af evalueringsresultater (1989:10)?

Shadish indrømmer i 1998, at: *"with few exceptions, evaluation theory is neither concise nor axiomatic; and it is not a single theory but rather a set of diverse theoretical writings held together by the common glue of having evaluation practice as their target"*(ibid.2). Rammen af Shadish et al. er heller ikke aksiomatisk og deduktivt udledt.

I 1991 skriver Shadish et al, at der ikke findes én bestemt opfattelse af, hvad teori er. De mener, at: *"theory connotes a body of knowledge that organizes, categorizes, describes, predicts, explains, and otherwise aids in understanding and controlling a topic"*(ibid.30). Schwandt (2001) nævner i hans opslagsværk fire forskellige opfattelser af teori (ibid.252-253). I *"Leksikon i Statskundskab"* er teori bl.a.: *"...et ord som i faglig sammenhæng har en positiv klang, og som enkelte gange bruges synonymt med en "god ide", andre gange om en oversigt over vigtige begreber inden for et fagområde..."*(Statskundskab 2004:30). Shadish et al.'s ramme og gennemgang af syv forfatterskaber kan betragtes som en god ide, og indeholder vigtige begreber indenfor et

fagområde. I forhold til FGE bliver spørgsmålet, om FGE er en god ide jf. definition af teori ovenfor.

Shadish har i 1998 et afsnit, som handler om problemer ved evalueringsteori. Han nævner fire forhold. For det første har de fleste evalueringsteoretikere ikke udviklet argumenter for, hvornår situationen tilsiger, at man netop skal træffe de valg og ikke andre valg (ibid.5). For det andet er der blandt teoretikerne en generel mangel på en fælles vision om, hvilken rolle evaluering skal spille (ibid.). For det tredje mangler der en generel accepteret metateoretisk nomenklatur, som kan anvendes til at klassificere en evalueringsteori, og som kan fortælle, hvad en given teori gør og ikke gør (ibid.6). For det fjerde mangler der komparativ teori af evalueringsteorier, der anvender den manglende metateoretiske nomenklatur til at sammenligne relative styrker og svagheder ved de enkelte evalueringsteorier (ibid.6).

Tesen i afhandlingen handler om, at FGE kun anvendes lidt i praksis, fordi det er en dårlig evalueringsteori. Både ud fra en teoretisk og praktisk vinkel må det være relevant, at stille spørgsmålet om en given evalueringsteori er god eller dårlig. Det er den indgangsvinkel Shadish et al. anvender. Det er dog en problematisk tese, når Shadish i 1998, som præsident for AEA skriver, at der ligefrem mangler en fælles nomenklatur til at beskrive de forskellige evalueringsteorier herunder FGE. At der mangler forskning, hvor man sammenligner teoriernes styrker og svagheder. Afhandlingen vil dog alligevel forsøge at give et svar på baggrund af den nomenklatur Shadish et al. udviklede i 1991.

Afhandlingens analyse vanskeliggøres ved at der er trade-offs i alle evalueringsteorier i forhold til forskellige mål (Shadish 1998:6). Vi kan f.eks. hypotetisk sige at Responsive evalueringsteorier, der lægger vægt på participation foretager et trade-off i forhold til design, metode og viden. Guba og Lincoln går så langt som til at sige, at et FGE evalueringsdesign er emergent og kontingent (1989:254-255). Det udvikler sig under processen, hvor interessenterne inddrages. Dette er i modsætning til teoribaseret evaluering, hvor der på forhånd opstilles programteorier, og fokus er på kausalitet (Bickman 2000; Dahler-Larsen & Krogstrup 2003:96-112). Det sker måske på bekostning af den løbende deltagelse af interessenter. Hvad er vigtigst: Deltagelse af interessenter i demokratiske evalueringsprocesser eller flotte programteorier, der undersøges ved stringent samfundsvidenskabelig metode? Shadish et al.'s ramme giver os ikke et svar.

Det centrale problem er: *"our lack of a common metatheoretical language and framework within which we can categorize and debate diverse positions"* (Shadish 1998:8). Shadish mener, at det giver problemer, hvis vi f.eks. skal forstå nye evalueringsteorier (ibid.). Scriven advokerer også for udviklingen af en metateori, som er nødvendig for at evaluering kan udvikle sig til en selvstændig videnskabelig disciplin (Scriven 1991A:24-26). Han opstiller krav

som f.eks. en kompleks taksonomi, fælles bevidsthed om noget nyt, en metodologi som giver berigende og anvendelige resultater, opbygning af vidensdatabase, teorier, lovmæssigheder og generalisationer (ibid.).

Som påvist ovenfor har danske og udenlandske evalueringsforskere hentet støtte i rammen, som Shadish et al. præsenterede i 1991. Rammen eller nomenklaturen anvendes også som skelet i afhandlingens analyse af FGE i kapitel 3. Shadish er klar over, at det kun er en nomenklatur, som ikke giver mulighed for at sige, at én teori eller ét valg er bedre end andre teorier eller andre valg. Han skriver: "*This metatheoretical nomenclature is not intended to dictate which option is best*" (Shadish 1998:8). Hensigten er i stedet at kategorisere de forskellige indfaldsvinkler til evaluering (ibid.).

2.3. Er evalueringsteori sand?

Evalueringsteorierne er på et stadie, hvor det er vanskeligt at undersøge om en evalueringsteori er god, dårlig eller ligefrem sand. En sand teori antages a priori at være en god teori. En teori kan videnskabsfilosofisk underkastes tre sandhedskriterier (Adolphsen 1995:73; Lübcke 1983:381-383).

Det første er korrespondenskriteriet. Teorien er sand, hvis den er i overensstemmelse med virkeligheden (Adolphsen 1995:73). Det virker ikke ulogisk at undersøge, om rammen udviklet af Shadish et al. er i overensstemmelse med virkeligheden. Optager de fem punkter professionelle evaluators? Giver de et dækkende billede af eksisterende teorier om evaluering? Korresponderer Scrivens teori med virkeligheden, når han siger, at evaluering handler om: "*determining the merit, worth and value of things*: (1991A:1)? Et spørgeskema kan sendes ud til professionelle evaluators. Resultatet kan vise, om der er overensstemmelse mellem udsagnet og virkeligheden.

Det er ikke muligt at undersøge, om FGE korresponderer med virkeligheden, da FGE er en præskriptiv teori om, hvordan vi bør evaluere. Som påvist gælder det de fleste evalueringsteorier herunder Scriven. Alligevel kan dele af en evalueringsteori testes i forhold til teoriens konkrete påstande om virkeligheden jf. overfor.

En alvorligere sag i forhold til korrespondenskriteriet er, at Guba og Lincoln (1989) ikke anerkender, at der eksisterer en virkelighed uafhængig af menneskers konstruktioner (1989:passim). De accepterer ikke distinktionen mellem teori og empiri. De skriver: "*theories and facts are not independent*" (ibid.100). De mener, at det er umuligt at opretholde: "*separate theoretical and observational languages*" (ibid.). Af ontologiske grunde vil de ikke godtage, at FGE sættes i forhold til korrespondenskriteriet. Popper og de kritiske rationalister (Beyme:1984:32-50; Jacobsen et al.1999:149-152) vil sige, at

FGE er en dårlig teori, hvis den ikke lader sig falsificere jf. Poppers falsificerbarhedskriterium (Knudsen 1990:100). En teori om, at virkeligheden kun består af menneskelige konstruktioner, er en ikke falsificerbar teori. Ifølge Popper vil en falsifikation kræve, at der findes blot et tilfælde af en klump virkelighed uden for menneskers konstruktioner. Vil det ikke kræve et menneske til at erkende klumpen af virkelighed udenfor menneskers konstruktioner? Det er absurd og et *contradictio in adjecto*. Hvad med den intersubjektive overførbarhed i forhold til falsifikationen? De vil sige, at FGE er: *"unvereinbar mit wissenschaftlichem Denken"* (Beyme 1984:34). I en dansk introduktion til videnskabsteori står der: *"netop dette at kunne falsificeres er et af kendetegnene på en videnskabelig teori"* (Jacobsen et al.1999:51). Klart er det, at FGE eller Shadish et al.'s begrebsramme ikke er gode videnskabelige teorier i Poppers forstand.

Det andet sandhedskriterium er koherenskriteriet (Adolphsen 1995:74). Ifølge Adolphsen menes her: *"en teori er sand, hvis den hænger sammen med eller er logisk konsistent med – kohererer – med andre teorier"* (ibid.74).

Begrebsrammen fra Shadish et al. kan hypotetisk konfronteres med andre begrebsrammer indenfor evalueringsteorien. Da rammen synes at stå alene, og er metaevalueringsteorien par excellence, kan der ikke undersøges for koherens i forhold til andre rammer. Shadish et al.'s begrebsramme er sandsynligvis dannet induktivt ved, at de har læst og analyseret syv evalueringsteoretikere. Derfor kohererer rammen i en eller anden udstrækning med teoretikernes værker. Koherensen gør begrebsrammen sand og dermed til en god metaevalueringsteori. Når Shadish et al. analyserer teoretikerne, er de selv inspireret af en eller anden form for koherens. Hænger det sammen for den enkelte teoretiker? Er der modsigelse og inkonsistens? De ser f.eks. på sammenhæng mellem en teoretikers ontologi, epistemologi og metodologi. De undersøger, hvordan en teoretiker logisk kommer frem til værdidomme. De konfronterer også den enkelte teoretiker med andre teoretikere (Shadish et al. 1991:passim).

Kohererer FGE med andre værker om evaluering? Berører Guba og Lincoln (1989) de punkter, som Shadish et al. har foreslået i deres ramme? Dette vil analysen i kapitel 3 vise. Er FGE logisk konsistent med, hvad andre klassikere har skrevet om evaluering? Da FGE, ifølge Guba og Lincoln (1989), er Fjerde Generations Evaluering, er det næppe interessant at undersøge, om FGE hænger sammen med tidligere generationers evalueringsteori. Guba og Lincoln (1989) påstår jo, at FGE er et opgør med tidligere generationer og det konventionelle paradigme. De ønsker ikke, at deres teori kohererer med andre evalueringsteorier. De skriver, at FGE: *"...stands in stark relief to earlier generations of evaluation...evaluation must now be viewed in very different ways"* (ibid.252). I forhold til koherenskriteriet er FGE en usand teori. Hellevik vil kritisk sige, at FGE ikke medvirker til en deduktiv videnskakkumulation

(Hellevik 2003:65), fordi FGE skiller sig markant ud fra andre evalueringsteorier. Dermed er det dårlig teori.

For det tredje findes et pragmatisk sandhedskriterium. En teori anses som sand: *"hvis den er frugtbar eller nyttig"* (Adolphsen 1995:74). Hvis teorien giver ny erkendelse eller beriger praksis er den sand. Man kan hævde, at rammen udviklet af Shadish et al. er sand, fordi den i stor udstrækning er nyttig og bliver anvendt af danske og udenlandske evalueringforskere. Spørgsmålet kan afklares empirisk. Det samme gælder i forhold til professionelle evaluatore. Finder de rammen nyttig? Kan FGE med fordel anvendes i praksis? Spørgsmålet er i overensstemmelse med Shadish et al.'s vægt på evalueringpraksis. Det pragmatisk kriterium virker anvendeligt i forhold til de rudimentære evalueringsteorier herunder metateori om evaluering.

Der er problemer med alle tre sandhedskriterier. Hvad er det for en virkelighed, som teorien skal korrespondere med? Konstruktivister hævder, at virkelighed er noget, der konstrueres af mennesker, velvidende at der er forskellige konstruktivismen (Wenneberg 2002; Jacobsen et al. 1999:153-159; Gergen 1985). Guba og Lincolns (1989) ontologi behandles i afsnit 3.1.

Koherenskræteriet udelukker ikke, at teorien er forkert. Det kræver dog, accept af, at noget er rigtigt og forkert. At man tror på korrespondens. Tyge Brahe udviklede et udviklet astronomisk system, som kohererede med det ptolemæiske system (Theilgaard 2006:24). I Brahes solsystem er jorden stadig i centrum (ibid.). På trods af at teorien kohererede med andre teorier, er teorien fejlagtig. At Shadish et al.'s metateoretiske ramme kohererer med de klassiske evalueringsteoretikere udelukker ikke, at rammen er forkert. Shadish et al. indrømmer, at ingen teoretikere passer perfekt ind i rammen, fordi de ikke forholder sig til alle de punkter Shadish et al. har sat op.

Det pragmatisk kriterium er heller ikke uden problemer. Påstanden er, at teorien er sand, hvis den er nyttig og fungerer i praksis. Nyttig for hvem og for hvad? En praksisteori som siger, at jorden er flad vil være tilstrækkelig for de fleste mennesker i dagligdagen. At noget fungerer i praksis kan være forkert og dybt konservativt.

2.3.1. Pragmatisme

Pragmatismens filosofiske hjemland er USA (Lübcke 1982A:9-14; Mortensen 1996). Nogle mener, at det er den amerikanske nationalfilosofi (Lübcke 1982A:10). De førende pragmatikere var bl.a. James og Dewey. Pragmatismen kulminerer i starten af det 20. århundrede (ibid.9). Fokus er på handling og praktiske konsekvenser. Sandhed er forbundet med succes og nytte (ibid.). Lübcke skriver: *"det sande er det, der har nyttige konsekvenser"* (ibid.). Han mener i øvrigt, at det er vulgærpragmatisme og et kuriøst fortidslevn (ibid.).

Pragmatismen kommer klart til udtryk hos mange evalueringsteoretikere. Patton (1990,1997) fokuserer på anvendelse (utilization). Pattons polemik med Weiss er fyldt med pragmatiske udsagn, som: "*our worth is to be judged by use*"(Patton 1990:192); "*evaluation is going to be useful*"(ibid.194) og: "*what worked and what didn't work*"(ibid.198). Hos Scriven findes lignende pragmatiske citater. Idealet på gode evalueringer er forbrugertests af varer (Scriven 1991A). Evaluering handler om: "*determining the merit, worth and value of things*"(ibid.1). Shadish (1998) mener, at evalueringsteori er drevet af praksis. Netop praksis er centralt i rammen af Shadish et al. Begrebsrammen synes at være inspireret af det pragmatiske kriterium. Det er hensigtsmæssigt at se på en evalueringsteoretiker i forhold til opfattelsen af sociale programmer, viden, værdier, anvendelse og evalueringspraksis. På trods af at Guba og Lincoln (1989) mener, at FGE er revolutionerende, og er et opgør med tidligere generationer af evaluering, skriver de sig ikke ud af den amerikanske pragmatisme. Faktisk er de endnu mere pragmatiske! De mener, at FGE sikrer, at evaluering virkelig bliver anvendt og skaber handling og løsninger (ibid.258). De skriver: "*But fourth generation evaluation, both because of its open stance that is empowering and enfranchising and because its formulations lead to constructions in which shared responsibility it featured, is catalytic in producing action. Evaluation and action are part and parcel of the same process...*"(ibid.259).

Mange debatter indenfor evalueringsteori handler om metodologi og anvendelse (Dahler-Larsen & Krogstrup 2001:passim). Det bærer præg af pragmatisme. Pragmatismen er et grundlæggende karakteristikon ved evalueringsdisciplinen i forhold til andre humanistiske- og samfundsvidenskabelige discipliner. Bemærk bogtitler som "*Handbook of Evaluation*" (Shaw et al.2006) eller "*Håndbog i evaluering*" (Rieper 2004) med undertitlen "*metoder til at dokumentere og vurdere proces og effekt af offentlige indsatser*". Der tales om evalueringsmodeller jf. afsnit 2.1. En bog giver en slavisk gennemgang af de enkelte modeller (Krogstrup 2003). I en anden "*håndbog*" af Dahler-Larsen og Krogstrup (2003) står der, at formålet er: "*at udvikle og afprøve evalueringsmodeller, som amter og kommuner kan bruge i fortløbende evaluering af deres egen indsats på det sociale område*" (ibid.6). Evalueringsmodel bliver her til: "*en opskrift på, hvordan evaluering i praksis udføres*"(ibid.7).

Filosoffer har svært ved at slå sig til tåls med pragmatismen. At noget virker eller ikke virker i praksis, kan ikke tages som udtryk for sandhed ifølge Lübcke (1982A:9-14). Der kan også rejses etiske indvendinger. At en nazistisk koncentrationslejr virker effektiv, gør den ikke sand. Det, som ikke virker, kan for nogen være meget nyttigt og dermed virke! I særlige sammenhænge kan det være nyttigt at lyve. Det virker! At en særlig evalueringsmodel, opskrift eller teori anvendes i praksis gør den ikke sand. Måske end ikke nyttig.

Pragmatismen forhindrer, at der udvikles gode evalueringsteorier herunder en fælles metaevalueringsteori. Årsagen er, at der bliver mindre at være fælles om blandt teorierne, når udgangspunktet er, at en sand eller god evalueringsteori blot skal være nyttig og fungere i en praksis. At Pawson og Tilley (1997) har udviklet Realistisk Evaluering med CMO-mekanismer (ibid.217) for at tage hensyn til konteksten, løser ikke problemerne. Det er snarere udtryk for en teoriudvikling, der er endnu mere pragmatisk! Teorien tager højde for selv lokale kausaliteter. For dem handler det om at kortlægge: *"which works for whom in what circumstances"*(ibid.) jf. Virkningsevaluering på dansk (Dahler-Larsen & Krogstrup 2003:51-79).

Evalueringsteoriens situation antyder, at det er pragmatismen, der vanskeliggør en teoretisk udvikling, der distancerer sig fra en praksis med uendelige mængder af konkrete situationer. En god teori om staters adfærd i international politik behøver ikke at give anvisninger på, hvordan stater bør optræde i konkrete situationer. En god medicinsk teori behøver ikke at give anvisninger på, hvordan lægen skal behandle i alle situationer. En teori om solsystemet behøver ikke at give svar på, hvordan man skal bedrive astronomi i praksis. Rammen fra Shadish et al. kræver, at teorierne skal give anvisninger på, hvordan man skal evaluere i alle kontekster med alle genstandsfelter. De krav som Shadish et al. opstiller, er overvældende. Skal en evalueringsteori virkelig omkring teorier om sociale programmer, videnskonsstruktion, værdier, anvendelse og evalueringspraksis?

Evalueringsteoretikere kan hente inspiration i Occams barberkniv (Bateson 2000:89,244), som også kaldes Lex parsimoniae. Princippet er, at den teori, som har færrest antagelser, enheder og postulater er at foretrække frem for en konkurrerende teori med flere antagelser, enheder og postulater. De pragmatiske evalueringsteorier bliver ofte meget omfattende. De bliver landkort, som forsøger at ligne landskabet i et og alt. Umiddelbart tror man, at Lex parsimoniae gør FGE til en god evalueringsteori. Guba og Lincoln (1989) skriver i forordet: *"...we hope that no one will read this book with the intent of finding a formula for doing evaluation. We do not wish anyone to conclude that now, finally, we know what to do and how do it"*(ibid.17). De skriver om emergent evalueringsdesigns og kontingente situationer. Faktum er dog, at bogen er fyldt med *"what to do"* lister. I bogens kapitel syv præsenteres et detaljeret flow-diagram. Guba og Lincoln (1989) fortæller om FGE processen i detaljer herunder om indgåelse af og indhold i kontrakter. Rossi, Lipsey & Freemann (2003) skriver i indledningen, at deres bog ikke er: *"intended to be a cookbook of procedures for conducting evaluation"* (ibid.IX) på trods af, at bogen er fyldt med praktiske anvisninger.

Sammenfattende giver de tre sandhedskriterier ikke et klart svar på afhandlingens tese om, at FGE er en dårlig evalueringsteori. Af grunde nævnt ovenfor er korrespondens- og koherenskriteriet svært anvendeligt på FGE. Tilbage står det pragmatiske kriterium. Om FGE er en dårlig teori, er afhængig

af teoriens nytte, men nytte for hvem, for hvad og hvornår? Det er problematisk at antage, at den evalueringsteori, der har flest praktiske anvisninger og fungerer i praksis, er den bedste evalueringsteori. Flint blev anvendt som materiale til redskaber gennem årtusinder. Og omvendt, at en teori som FGE, der næsten ikke bliver anvendt i praksis, er en dårlig evalueringsteori.

Sandhedskriterierne giver end ikke klarhed over, om Shadish et al.'s metateori om evaluering, er den sande metateori i forhold til at bedømme en enkelt evalueringsteori som FGE. Rammen forekommer pragmatisk sand. Den er nyttig, og vil i mangel af bedre blive anvendt i kapitel 3 til at analysere FGE.

2.4. Er FGE evaluering?

Inden selve analysen af FGE i kapitel 3, er det relevant kort at undersøge, om FGE er en teori, som handler om evaluering og ikke andre ting.

Guba og Lincoln (1989) skriver i begyndelsen af bogen, at spørgsmålet om, hvad evaluering er, er et forkert spørgsmål (ibid.21). De skriver: "*We take definitions of evaluation to be human mental constructions, whose correspondence to some "reality" is not and cannot be an issue. There is no answer to the question, "But what is evaluation really?" and there is no point in asking it*"(ibid.)(Understregning i citater betyder her og fremover, at det er kursiveret i original teksten). De mener, at der ikke findes en rigtig måde at definere evaluering på (ibid.). Dette udgangspunkt gør det endog meget vanskeligt at foretage en analyse af FGE, og undersøge om FGE er dårlig eller god evalueringsteori. Afhandlingen skal altså analysere et fænomen, som ikke defineres og afgrænses i tid, rum og indhold. Evaluering kan være alt.

Det er uklart, hvilken type af konstruktivisme (Wenneberg 2002;Heap 1995; Gergen 1991) de taler om, når de påstår, at en definition af evaluering blot er en menneskelig mental konstruktion. Hvad forstår Guba og Lincoln (1989) ved menneskelige konstruktioner? Er det kognitivkonstruktivisme eller socialkonstruktivisme (Heap 1995)? Findes begrænsningerne og problemerne i sproget eller i det, at der ikke eksisterer en virkelighed uafhængig af menneskets konstruktioner? Spørgsmålet tages op i afsnit 3.1.

Når Guba og Lincoln (1989) bekender sig til konstruktivisme, må det være interessant for dem at undersøge, hvorfor, hvordan, hos hvem og hvor der opstår en mental-/socialkonstruktion, som betegnes "*evaluating*". Som læser ønsker man en sådan analyse, når de arrogant undlader at give en definition på evaluering. De skriver sig dermed også ud af den metaevalueringsteori, der måtte være. Hvordan bedømmer vi noget, vi ikke ved, hvad er?

Guba og Lincoln (1989) kunne have præsenteret en diskursanalyse af udviklingen af evalueringsdefinitioner igennem tiden. De kunne have ladet sig inspirere af Foucaults vidensarkæologi og genealogi (Andersen 1999:28-63). I forhold til evalueringsbegrebet kunne de: *"problematisere de diskurser og praktikker, vi har i dag, ved at vise dem tilbage til de herredømmeforhold, de er etableret under, herunder udpege de brud i terræn, indenfor hvilket strategier, institutioner og praktikker er formet. Det drejer sig om at opløse samtidens selvfølgeligheder ved hjælp af historien"*(ibid.56-57). Som udgangspunkt kunne Guba og Lincolns (1989) påstand om, at evaluering blot er en mental konstruktion være starten på en interessant Foucault analyse. Andersen skriver videre, at: *"I stedet for på forhånd at definere, hvad vi undersøger, må vi undersøge hvordan vor genstand er konstrueret historisk på vidt forskellige måder og i vidt forskellige sammenhænge"*(Andersen 1999:58). En sådan analyse er ikke relativisme, men perspektivisme (ibid.62). Denne tråd tager Guba og Lincoln (1989) kun delvist op. I dansk sammenhæng findes der antaster til diskursanalyse af begrebet evaluering (Dahler-Larsen 2001A:9-23;2006B:27-36). Guba og Lincolns (1989) fragmentariske gennemgang af Første-, Anden- og Tredje Generations Evaluering er langt fra en diskursanalyse af evalueringsbegrebet. De reducerer de tre generationer til at handle om hhv. måling, beskrivelse og bedømmelse (ibid.22-31). Det bliver til en kort stereotyp omtale af påståede metoder, der bliver anvendt af de forskellige generationer. Guba og Lincoln (1989) giver ingen analyse af herredømmeforhold, og hvorfor evalueringsbegrebet netop har det indhold, som hører til de forskellige generationer. De forklarer ikke, hvorfor, hvordan, hvem og hvor der opstår brud mellem generationerne. De moderer også opdelingen ved at sige, at de forskellige generationer er indlejrede i hinanden og stadig findes på samme tid. Altså måling, beskrivelse og bedømmelse praktiseres stadig.

På trods af Guba og Lincolns (1989) konstruktivistiske præambel, glider de igennem bogen over til pragmatiske anvisninger på, hvordan man skal evaluere i forhold til FGE. De skriver sidst i bogen: *"The purpose of evaluation is some form of action/or decision making"*(ibid.249). De falder altså selv i fælden, og giver en slags definition af evaluering. Det vurderes som en mangel, at de ikke forholder sig eksplicit, kvalificeret og analytisk til den mentale konstruktion, som de har døbt *"evaluering"*.

Guba og Lincoln (1989) mener, at FGE selv er en konstruktion, som er: *"more informed and sophisticated than previous constructions have been"*(ibid.22). De mener, at de tre tidligere evalueringsgenerationer har tre alvorlige fejl (ibid.31-32). For det første en tendens til managerialisme. Evaluering er fokuseret på styring og opdragsgiverens behov og ønsker. For det andet en manglende tilpasning til værdipluralismen i samfundet. For det tredje en forpligtelse på det traditionelle positivistiske videnskabsparadigme. I forhold til måling, beskrivelse og bedømmelse karakteriserer de Fjerde Generations

Evaluering ved begrebet forhandling (ibid.8). Evaluering handler om forhandling mellem interessenter, der bliver berørt af evalueringen.

I afhandlingen foretages der ikke en diskursanalyse af evalueringens udvikling, men i stedet konfronteres FGE nedenfor med andre definitioner af evaluering. På den måde undersøges om FGE overhovedet er en teori om evaluering.

2.4.1. Evalueringsdefinitioner

Som det er fremgået, er der ikke enighed om, hvad evaluering er. Scriven mener, at evaluering er: *"the process of determining the merit, worth and value of things"*(1991A:1). For Guba og Lincoln (1989) handler evaluering om, at interessenter i en løbende proces tilkendegiver, forhandler og stræber efter konsensus omkring: *"claims, concerns, and issues about the evaluand"* (ibid.39). I overensstemmelse med Stakes Responsive Evaluering (Stake 2004;Shadish et al.1991:270-314) er det ikke i FGE evaluators opgave, at bedømme, hvad der er godt og skidt (Larsen 2007A:12-14). Ifølge Scrivens definition er FGE ikke evaluering. I FGE er fokus på formativ evaluering (Guba & Lincoln 1989:190;Scriven 1991B:18). Guba og Lincoln (1989) skriver: *"Constructivist evaluators may work in each arena but are more likely to be interested in formative rather than summative efforts"*(ibid.190).

Som påvist forstår Rossi, Lipsey og Freeman (2003) programevaluering som: *"...the use of social research methods to systematically investigate the effectiveness of social intervention programs in ways that are adapted to their political and organizations environments and are designed to inform social action to improve social conditions"*(ibid.16). De forpligter sig ikke til at anvende en bestemt samfundsvidenskabelig metode (ibid.16). Guba og Lincoln (1989) er også åben overfor både kvalitative og kvantitative metoder. De skriver: *"But there will be times when quantitative methods - test or other measurement instruments, or numeric displays - will be, and should be, used"* (ibid.25).

Det kan diskuteres om FGE vil få problemer i forhold til Rossi, Lipsey og Freemans (2003) definition, som kræver, at evaluering skal være en systematisk undersøgelse af resultaterne af sociale programmer.

Guba og Lincoln (1989) mener, at en evaluand som f.eks. et socialt program, er en mental konstruktion (ibid.188). De skriver: *"The evaluand is no more "real" than any other mental construction"*(ibid.188-189). Dette ontologiske udgangspunkt er afsættet for Guba og Lincolns (1989) metodologiske overvejelser. Grupper af personer som har en interesse i evaluanden skal inddrages i evalueringen, da evaluanden ikke eksisterer, som en *"ding an sich"* i kantiansk forstand (Hartnack & Sløk 2000 passim;Lübcke 1982B:21). Inddragelse af interessenter betyder, at Guba og Lincoln (1989) taler om

emergens i forhold til evalueringsdesigns. Det er ikke muligt på forhånd at afgøre, hvilke mentale konstruktioner, herunder påstande, bekymringer og problemer, interessenterne oplever omkring evaluanden. Det er heller ikke muligt på forhånd at sige, hvilke informationer og samfundsvidenskabelige metoder, der skal anvendes. Interessenternes forskellige mentale konstruktioner kan betyde, at der skal indhentes yderligere information. Det kan være faglitteratur, andre evalueringsrapporter eller samfundsvidenskabelige undersøgelser igangsæt af evaluator.

Guba og Lincoln (1989) er åben overfor traditionel samfundsvidenskabelig metode, der kendes fra Første Generations Evaluering. De skriver: *"Some information may require the construction of new instrumentation which may include paper-and-pencil or even brass instrumentations; similarly, existing instrumentations, such as published tests, may be used..."*(ibid.216-217). Det kræver dog, at der er interessenter, der ønsker deres påstande, bekymringer og problemer belyst med kvantitative metoder. Der er ingen tvivl om, at Guba og Lincoln (1989) foretrækker kvalitative metoder som observation og åbne spørgsmål. De mener, at interessenters mentale konstruktioner af evaluanden er i et dialektisk forhold til tid og kontekst. De mentale konstruktioner bestemmes af tiden og konteksten, men også at tiden og konteksten bestemmes af de mentale konstruktioner (ibid.177). Derfor foretrækker de ikke generaliserbar viden, der er af ideografisk karakter. De skriver: *"The use of a human instrument suggest that the method employed should be primarily those that are congenial to humans – qualitative methods"*(ibid.). I forhold til traditionel samfundsvidenskab opstiller Guba og Lincoln (1989) parallelle kriterier for, hvad god videnskab er. De taler om: *"Trustworthiness"*(ibid.233) i stedet for kriterier som validitet, reliabilitet og generaliserbarhed.

I FGE afsluttes processen nødvendigvis ikke med en traditionel evalueringsrapport, som viser resultaterne af undersøgelsen af det sociale program jf. definitionen af evaluering hos Rossi, Lipsey og Freeman (2003). I overensstemmelse med Stake anbefaler Guba og Lincoln (1989) et case-studie, som ikke blot omhandler resultater og *"facts"* ved evaluanden (ibid.233). De skriver: *"The report cannot simply be about the evaluand and its context, but must enable readers to see how the constructors make sense of it, and why"*(ibid.223). Læseren af rapporten skal få en følelse af: *"walk in the shoes of the local actors"*(ibid.).

Rossi, Lipsey og Freemans (2003) resultater af sociale programmer, er hos Guba og Lincoln (1989) blevet til en systematisk forhandling og dokumentation af interessenters mentale konstruktioner af et socialt program i en specifik spatiotemporal konfiguration. Ifølge Guba og Lincoln (1989) er de mentale konstruktioner emergente, og gælder kun for en kort tid (ibid.226). De skriver, at FGE: *"...never stop; they merely pause"*(ibid.). Bemærk, at FGE i sine procedurer for gennemførelse virker meget systematisk og lineær (ibid.184-

227). Dette er på trods af, at det er en kritik forfatterne retter mod tidligere generationer af evaluering.

Ifølge Rossi, Lipsey og Freemans (2003) definition af evaluering, er FGE ikke evaluering. FGE er ikke en systematisk undersøgelse af sociale programmets resultater, ud fra traditionel samfundsvidenskabelig metode med henblik på at forbedre de sociale betingelser.

FGE er ikke i overensstemmelse med Pattons definition af evaluering. Han skriver, ifølge Mark et al.(2006), at: *"Program evaluation is the systematic collection of information about the activities, characteristics, and outcome of programs to make judgements about the program, improve program effectiveness, and/or decisions about future programming"*(ibid.6). Som påvist, handler FGE om forhandling mellem interessenter omkring deres påstande, bekymringer og problemer ved en evaluand. Der er ikke tale om systematisk og dokumenteret indsamling af information, som evaluator anvender til at give en bedømmelse af programmet.

FGE er ikke forenelig med Vedungs (1998) definition af evaluering. Han skriver: *"Utvärdering er noggrann efterhandsbedömning av utfall, slutprestationer eller förvaltning i offentlig verksamhet, vilken avses spela en roll i praktiske beslutssituationer"*(ibid.20). FGE lægger vægt på formativ evaluering og dermed ikke en snæver: *"efterhandsbedömning av utfall, slutprestationer eller förvaltning i offentlig verksamhet"*(ibid.). Det er ligefrem et mål for FGE ikke at fokusere på at skabe resultater til en ledelse, der skal træffe en beslutning. FGE er imod managerialisme. Som påvist, hævder Guba og Lincoln (1989) alligevel, at det er en styrke, at FGE kan skabe resultater i praksis, dvs. i Vedungs: *"praktiske beslutssituationer"*(1998:20).

Der er andre definitioner af evaluering, hvor FGE klart går hen og bliver til evaluering. Da Guba og Lincoln (1989) er inspireret af Stakes Responsive Evaluering (ibid.38-39), kan det ikke undre, at FGE er i god overensstemmelse med Stake opfattelse af evaluering (Stake 2004;Shadish et al.1991:270-314).

Stake foreslår også casestudiet. Evaluering skal foregå lokalt og tæt på praksis. Det er i samsvar med det kontekstuelle og ideografiske i FGE. Ifølge Stake, er der mange virkeligheder. Et program har forskellig værdi, for forskellige personer. Stake skriver: *"a program has no single true value...the value of an arts-in-education program will be different for different people, for different purpose"* (Shadish et al.1991:307). Det er næsten i overensstemmelse med interessenters forskellige mentale konstruktioner hos Guba og Lincoln (1989). Stake anvender en vifte af kvalitative metoder, som også indgår i FGE. Det er f.eks. åbne interviews, observationer og læsning af dokumenter. I tråd med FGE foretrækker Stake at beskrive forskellige personers opfattelse af virkeligheden. Stake rejser også tvivl om, der skal skrives en afsluttende evalueringsrapport (Shadish et al.1991:281). Der kan skrives forskellige

rapporter til forskellige målgruppers behov og interesser (ibid.283). Guba og Lincoln (1989) foreslår også, at der kan skrives forskellige casestudier til forskellige interessenter. Det er vigtigt for Stake, at det er læserne af evalueringsrapporten, der selv fortolker resultaterne og drager konklusionerne (Shadish et al.1991:285-286). Pointen overtages af Guba og Lincoln (1989). Evaluators rolle i FGE synes delvist at ligne den rolle, som Stake omtaler. For Stake er evaluator facilitator mere end en person, der spreder indsigt.

En klar forskel på Stake og Guba og Lincoln (1989) er, at FGE har en konstruktivistisk indpakning. Guba og Lincoln har et stort forfatterskab bag sig. Det er tydeligt, at de i FGE fra 1989 inddrager ideer og fragmenter fra bøger i 1981 (Guba & Lincoln) og fra 1985 (Lincoln og Guba). I 1989 bliver ældre pointer, fra tidligere bøger, til FGE og konstruktivistisk evaluering. De henviser f.eks. til bogen: *"Effective Evaluation"* (1981), hvor de så i 1989 moderer deres distinktioner, og giver det en konstruktivistisk fernis (1989:189-190). I 1989 skriver de, at det ikke længere hedder: *"naturalistic inquiry"* (1989:19,note 1). Åbent erkender de: *"...we feel those terms are no longer appropriate. Further, we have come to appreciate that the central feature of our paradigm is its ontological assumption that realities, certainly social/behavioural realities, are mental constructions. Thus we have elected to use the terms constructivism and constructivist..."*(ibid.19).

Stake vil antageligvis kalder FGE for evaluering. Bemærk dog, at Stake tager afstand fra konstruktivisme. Kushner skriver: *"Stake concludes that denying a world of concrete realities is not at good way of "crossing a busy street"'*(Kushner 1996:196).

I *"Handbook of Evaluation"*(Shaw et al. 2006:5-6) mener redaktørerne, at: *"systematic evaluation is conceptualized as a social and politicized practice that nonetheless aspires to some position of impartiality or fairness, so that evaluation can contribute meaningfully to the well-being of people in that specific context and beyond"*(ibid.5-6). Definitionen af evaluering som en social- og politisk praksis kan forenes med Guba og Lincoln (1989), som mener, at begrebet evaluering blot er en mental konstruktion. Altså, en social- og politisk praksis. Logisk får vi en tautologi. Evaluering er det, som kaldes for evaluering!

På trods af at Guba og Lincoln (1989) ikke giver en definition af, hvad evaluering er, viser definitionerne ovenfor, at FGE både kan være evaluering, men også ikke evaluering. Det er afhængigt af, hvilke teoretikeres definitioner der anvendes. I forhold til afhandlingens tese, er FGE en dårlig evalueringsteori, hvis der ikke er tale om en teori om evaluering. Men som påvist, er der teoretikere, som antageligvis vil kalde FGE for evaluering. FGE bliver også omtalt i bøger om evaluering (Shaw et al.2006:passim;Krogstrup 2003:103-110). Dele af bogen fra 1989 af Guba og Lincoln inddrages som

obligatorisk litteratur på et masterstudie i evaluering på Syddansk Universitet. Med andre ord forbindes FGE med evaluering.

I flere af definitionerne af evaluering indgår kravet om systematik uden, at det er klart, hvad der menes. Som påvist, er det pragmatiske et karakteristikon ved evaluering. Et andet karakteristikon er det systematiske. Problemet er, at det ikke er klart, hvad det er for en systematik, der tales om for at noget kan betegnes som evaluering. Er det nok, at der er en systematisk fremgangsmåde eller proces? F.eks. flowdiagrammet for FGE (Guba & Lincoln 1989:186-187)? I så fald kan næsten alt være evaluering. Er det i stedet kravet om systematisk indsamling af information? Så kan alt statistik blive til evaluering. Er det en systematisk anvendelse af samfundsvidenskabelig metode? Albæk og Rieper (2001) går videre end traditionel forskning og samfundsvidenskabelig metode. For dem er evaluering: "*systematisk indsamling af data*"(ibid.130) uden systematik bliver defineret. Er det snarere en identifikation af en logisk systematik mellem høstede erfaringer og de slutninger der drages? Det er uklart, hvad det er for en systematik, der kræves for at noget bliver til evaluering. Under alle omstændigheder er FGE pragmatisk og anbefaler en systematisk proces, som i sig selv gør FGE til evaluering. Det konkluderes, at processen i FGE kan betegnes som evaluering. Derfor kan det undersøges, om FGE er en dårlig evalueringsteori.

2.5. Opsamling

Sammenfattende kan der skelnes mellem teori om evalueringspraksis (metateori) og teori om, hvordan der bør evalueres i en konkret praksis. Altså, der er teori om praksis og teori i praksis. Sammenfattende består metaevalueringsteorien af kronologiske, begrebs- og indholdsmæssige inddelinger af forfattere, som har skrevet om, hvordan der bør evalueres.

Gennemgangen har vist, at metaevalueringsteorien er på et tidligt og uudviklet stadium. Ikke mindst fordi der er mange pragmatiske og pekuniære argumenter for en selvstændig evalueringsteori, hvor det burde have været teoretiske argumenter. Derfor er det vanskeligt at bedømme en evalueringsteori, som FGE. Det bedste bud på en metateori er begrebsrammen fra Shadish et al. Den anvendes også af danske evalueringforskere. Den vil blive anvendt i den efterfølgende analyse af FGE. Rammen er blevet problematiseret. Den er induktiv, og kan ikke give et overbevisende svar på, om FGE er en dårlig evalueringsteori. Rammen er ikke logisk, da der er tvivl, om den består af fire eller fem punkter. Dens krav til en god evalueringsteori er store. Der er ikke klare teoretiske argumenter for, at en given evalueringsteori netop skal indeholde de mange forskellige teorier, som Shadish et al. kræver.

Det er påvist, at både metateori om evalueringspraksis og evalueringsteori i praksis er præget af amerikansk pragmatisme. Netop pragmatismen gør det vanskeligt at etablere en fælles teoretisk platform, hvorfra FGE kan bedømmes. Evaluering bliver det man anvender i praksis og kalder for evaluering. Det er en tautologi.

Ved hjælp af forskellige definitioner af evaluering kan der advokeres for, at FGE er en evalueringsteori. Generelt er det ikke enighed om, hvordan evaluering skal defineres. Guba og Lincoln (1989) underlader helt at definere begrebet evaluering.

Det er valgt at analysere FGE teoretisk. Da evaluering og evalueringsteori ifølge Shadish (1998) er et: "*practice driven field*"(ibid. 10), og i væsentlig grad fremskyndet af konkrete evaluators behov og opfindsomhed (ibid.) kunne vi i stedet have spurgt evaluators, hvorfor de ikke anvender FGE. Denne tilgang er fravalgt, da mange danske evaluators næppe har kendskab til eller erfaring med FGE.

3.0. Er FGE en dårlig evalueringsteori?

Kapitlet indeholder en analyse af FGE med udgangspunkt i den metateoretiske ramme udviklet af Shadish et al. Rammen anvendes som disposition for analysen. Alle hovedafsnit indledes med en kort neutral gennemgang af Shadish et al. En problematisering er foretaget i kapitel 2.0.

Kapitlet indledes i afsnit 3.1. med en undersøgelse af FGEs konstruktion af viden. Afsnit 3.2. indeholder en analyse af værdier i FGE. Afsnit 3.3. omhandler FGEs opfattelse af sociale programmer. Afsnittet er underinddelt i et afsnit om repræsentativt demokrati og FGE (3.3.1.), og magt og FGE (3.3.2.). Der er et afsnit om anvendelse og FGE (3.4.) samt endelig et afsnit 3.5. om evalueringspraksis og FGE. I afsnittet om evalueringspraksis inddrages erfaringer fra praksis med FGE.

3.1. Viden og FGE

Shadish et al. kræver, at en god evalueringsteori gør rede for, hvordan viden konstrueres. Det er teori om, hvordan evaluator skaffer sig viden om, hvordan sociale programmer fungerer og personer agerer (ibid.32). Der skal redegøres for, hvilke metoder der skal anvendes for at fremskaffe troværdig viden (Donaldson & Lipsey 2006:59). Der skal gøres rede for, hvordan den viden, der fremskaffes af evaluator, er forskellig fra anden viden (Shadish et al.1991:42).

Shadish et al. kræver, at en god evalueringsteori redegør for sit ontologiske, epistemologiske og metodologiske udgangspunkt (ibid.42). Det ontologiske handler om spørgsmål vedrørende virkeligheden eller det værende. Eksisterer der en virkelighed? Hvilken natur har den? (ibid.42-43). Hvad er det værende? Som Olsen (2002) skriver: *"Hvad betyder det, at noget – fx "sociale fænomener" eksisterer eller findes?"*(ibid.50). Shadish et al. skriver: *"ontology concerns common and pervasive attributes of being, such as whether things we experience are real or not"*(ibid.43). Det epistemologiske vedrører en redegørelse for, hvorfor noget er sand viden. I forhold til epistemologi skriver Olsen (2002), at det handler om: *"hvad er mulighederne for "sand" erkendelse om "det værende"?"*(ibid.50). Shadish et al. skriver: *"epistemology is the study of the characteristics of and standards for knowledge"*(ibid.43). Metodologi handler om, metoder til at konstruere viden. Olsen (2002) stiller her spørgsmålet om: *"hvordan erkendes "det værende"?"*(ibid.50). Med andre ord: *"procedurer for generering af erkendelse"*(ibid.).

Shadish et al. finder, at der er en tendens til, at evalueringsteoretikerne fokuserer meget på metodologiske forhold. De mener, at en god evalueringsteori også inddrager ontologiske og epistemologiske spørgsmål (ibid.43). Det er ikke et spørgsmål om den rigtige ontologi eller epistemologi,

men mere om der er en argumentation og teori. Det er også et spørgsmål om, der er en logisk sammenhæng mellem evalueringsteoriens ontologiske, epistemologiske og metodologiske argumentationer.

For en umiddelbar betragtning er FGE en god evalueringsteori, fordi Guba og Lincoln (1989) anvender meget plads på at redegøre for deres ontologiske og epistemologiske udgangspunkt. En boganmelder skriver, at FGE er et: *"...iconoclastic attack on the scientific method"* (Bond 1991) med henblik på at skabe et nyt paradigme indenfor evaluering (ibid.). En anden forsker mener, at Guba og Lincoln (1989) tegner en karikatur af videnskab og positivisme med sigte på at: *"build up a "straw-man" argument"* (Kushner 1996:195). Bond ser bogens pointer som: *"treatise on epistemology"* (Bond 1991). I forhold til det metodologiske skriver Bond, at: *"For an ostensibly practical text, it is unnecessarily polemical, repetitive, and abstract"* (ibid.). En anmelder i et canadisk tidsskrift registrerer: *"What cannot be gained from this book, good as it is, is a felt sense of the power, and of the costs, of this methodology. Such a sense comes only from experience"* (Canadian Journal of Communication (1991). En anmelder af bogen hæfter sig ved, at der: *"...is not even one sample study that the authors present in enough detail to demonstrate in actuality the practical value of their model"* (Fishman 1992:268). Han mener, at der mangler konkrete teknikker i forhold til arbejdet med de mange interessenter, som skal inddrages i FGE-processen (ibid.). Citaterne viser, at det måske snarere er på det metodologiske område, at FGE har problemer. Dette behandles i afsnit 3.5. om evalueringspraksis.

Det er ikke hensigten at afvise ontologien og epistemologien bag FGE, men snarere at påvise dårligt beskyttede flanker i Guba og Lincolns bog (1989). Afhandlingen giver ikke en analyse af den store paradigmedebat indenfor samfundsvidenskaben (Guba 1990; Guba & Lincoln 1994). Der vil heller ikke være en snæver konstruktivistisk analyse af FGE for konstruktivistiske feinsmeckere. Analysen holder sig snævert til Guba og Lincolns bog fra 1989.

Som påvist, er det ontologiske udgangspunkt bag FGE, at der ikke findes en objektiv virkelighed, som kan erkendes uafhængig af observatøren. Guba og Lincoln (1989) mener, at FGE er meget forskellig fra tidligere generationer af evaluering (ibid.12). De skriver: *"The posture of fourth generation evaluation is quite different, It begins with the assumptions that realities are not objectively "out there" but are constructed by people, often under the influence of a variety of social and cultural factors that lead to shared constructions. But socially shared constructions are not equivalent to the positivist`s "reality"; there is no reality except that created by people as they attempt to "make sense" of their surrounds"* (ibid.12-13).

Det er ikke klart, hvor radikal Guba og Lincolns (1989) konstruktivisme er. Konstruktivisme dækker over flere forskellige opfattelser (Wenneberg 2002; Rasmussen 2004:241-274). Wenneberg skelner mellem fire forskellige

konstruktivismen. Der kan også skelnes mellem en socialkonstruktion af viden om den sociale virkelighed, og en socialkonstruktion af viden om den fysiske virkelighed (Wenneberg 2002:112). I citatet ovenfor antyder Guba og Lincoln (1989), at det eneste, der eksisterer, er menneskers konstruktioner i forsøg på at danne mening af deres omgivelser. Dette betegnes her i afhandlingen som radikal konstruktivisme. Bemærk dog, at det ikke er entydigt, hvad radikal konstruktivisme dækker over. Wenneberg skriver om radikal epistemologisk konstruktivisme (2002:109,151). Qvortrup har en lidt anden opfattelse af begrebet radikal konstruktivisme (2004:78). Opfattelsen i afhandlingen af radikal konstruktivisme er mere i overensstemmelse med Rasmussens anvendelse i hans beskrivelse af Glasersfelds radikale konstruktivisme (Rasmussen 1997:109-113). Altså, at der ikke eksisterer en virkelighed udenfor mennesket.

Guba og Lincolns bog (1989) indeholder mange eksempler fra naturvidenskaben, hvor naturvidenskabsmænd erkender, at naturvidenskab er konstruktioner, der bygger på modeller skabt af mennesker. Som eksempler fra "*hard science*" (ibid.93) nævner de studiet af lys (ibid.91), Einsteins relativitetsteori (ibid.) og atomers opbygning (ibid.). I note 1 (ibid.19) kan der spores en vis tvivl hos Guba og Lincoln (1989) i forhold til, hvor radikal deres konstruktivisme er. De skriver: "*...realities, certainly social/behavioral realities, are mental constructions*" (ibid.19). De er ikke sikre på, at der ikke eksisterer en fysisk virkelighed, når det handler om naturen.

Bemærk, i de to citater ovenfor taler Guba og Lincoln (1989) i det første citat om: "*socially shared constructions*", og i det andet citat om: "*mental constructions*". Det er ikke ligegyldigt, om det er sociale fælles konstruktioner eller individuelle mentale konstruktioner. Sætningen "*mentale konstruktioner*" anvendes mange steder i bogen. Der synes at være tale om en individuel perception og fortolkning af virkeligheden. Denne form for konstruktivisme findes bl.a. hos Piaget (Labinowicz 1980 cf. Rasmussen 1997:109-113) og Dewey (Fink 2000). De påstår ikke, at der ikke ontologisk findes en verden uden for mennesket. Andre steder får man indtryk af, at Guba og Lincoln (1989) tænker på sociale konstruktioner à la Vygotsky (Moll 1990:passim), og visse steder bliver de radikale, og ender i solipsisme (Rasmussen 1997:125-131).

Heap (1995) sætter sig for at analysere Guba og Lincolns (1989) konstruktivisme. Heaps udgangspunkt er, at der er en uoverensstemmelse mellem deres ontologi og metodologi. Han har sympati for Guba og Lincolns demokratiske ideer med at inddrage interessenter i evaluering (Heap 1995:51). Han finder, at det nye i FGE er metodologien, som handler om at skabe konsensus blandt interessenter. Heap (1995) skriver: "*The heart and valued of Guba and Lincoln`s contribution lie in their innovative consensual methodology for evaluation. That is where practitioners of the evaluative arts and science ought give their attention*" (ibid.52).

Heap (1995) mener, at der kan skelnes mellem to hovedretninger indenfor konstruktivisme, som hhv. er kognitiv- og socialkonstruktivisme. Der skal ikke redegøres nærmere for udviklingen af de to typer af konstruktivismer (jf.Heap 1995). Kognitiv konstruktivisme er inspireret af Piaget og er mindre radikal. I analysen vælges ordet konstruktivisme frem for konstruktionisme. Førstnævnte betegnelse er knyttet til Piaget og det kognitive, og sidstnævnte mere til det sociale, fælles konstruktioner og sproget. Konstruktivisme og konstruktionisme anvendes synonymt, da det er tilstrækkeligt i forhold til analysen af Guba og Lincoln (1989).

Heap (1995) simplificerer forskellen mellem de to konstruktivismer ved at skrive: *"Social constructionism is concerned with object collectively constructed. Cognitive constructionism is concerned with objects cognitively constructed"*(ibid.53). Ofte er de to retninger blandet sammen og forudsætter hinanden. Sociale konstruktioner forudsætter kognitive konstruktioner hos det enkelte individ og omvendt. De to retninger fokuserer på forskellige forhold.

Heap (1995) foretager en underinddeling af de to konstruktivismer i forhold til monisme og dualisme (cf.Rasmussen 2004:236-237). Descartes mener, at der kan skelnes mellem bevidsthed og stof, subjekt og objekt, *res cogitans* og *res extensa* (Qvortrup 2004:75-79,99). Det er udtryk for dualisme (cf.Rasmussen 2004:236-237). Monisme hævder, f.eks. som idealisterne, at alt er bevidsthed eller som marxisterne, at alt kan reduceres til materie. Hos Wenneberg (2002) er monisme en mere radikal konstruktivisme.

Ifølge Heap (1995) er Berger og Luckmann de første, som taler om en social konstruktion af virkeligheden (ibid.52). Det handler om, hvordan mennesker konstruerer deres virkelighed i dagligdagen (jf.Schutz 2005). Berger og Luckmann påstår ikke, at der ikke findes en objektiv virkelighed udenfor hverdagens konstruktioner (jf. Berger & Kellner 1982). De er repræsentanter for en socialkonstruktivisme, der accepterer dualisme. Piaget og Dewey er kognitive konstruktivister, der også accepterer en dualisme. Der er en objektiv virkelighed udenfor børnenes mentale konstruktioner (Labinowicz 1980:passim).

Heap (1995) mener, at Guba og Lincoln (1989) blander det hele sammen på en inkonsistent måde. Han skriver: *"I shall note deep inconsistencies between the rhetoric and practice of fourth generation evaluation"* (Heap 1995:56). Der er mange citater fra Guba og Lincoln (1989), hvor de entydigt giver udtryk for, at der ikke eksisterer en virkelighed. Alt er sociale konstruktioner. Det gælder også deres egen teori (ibid.264). Ifølge Heap (1995) er det socialkonstruktivisme med monisme.

Andre steder moderer Guba og Lincoln (1989) deres synspunkt, og skriver om blød videnskab. I forhold til *"soft"* science stiller de spørgsmålet: *"Can terms*

like organization, school, personality, academia, patient care, safety net, curriculum, task, neurosis, and the like, stand for anything "real" or are they more fruitfully viewed as constructions that can and should change with time?" (ibid.94). Guba og Lincoln (1989) redegør ikke for processerne bag de sociale konstruktioner. Hvordan er konstruktionen skole skabt, og hvordan ændrer den sig? De mangler at beskrive motoren og brændstoffet bag en social konstruktion. Som påpeget, kunne Guba og Lincoln (1989) have anvendt Foucaults diskursanalyse.

Når Guba og Lincoln (1989) skriver om mentale konstruktioner og redegør for den metodik, der skal anvendes i FGE, glider de ubesværet over i dualisme. Kongeeksemplet er deres uklare distinktion mellem informationer og konstruktioner. De skriver: *"Any given construction can be no better than the information on which it is based and the sophistication the constructor can bring to bear in appreciating and understanding that information"*(ibid.258). De skriver om, hvordan interessenter kan danne fælles konstruktioner: *"if enough information is available and if appropriate levels of sophistication can be reached"*(ibid.216). De skriver om, hvordan: *"information may have been insufficient to lead to reconstruction"*(ibid.216). Det er et krav, at repræsentanter for interessenterne kan forhandle og have adgang til samme information (ibid.246). Evaluator kan igangsætte kvalitative og kvantitative undersøgelser for at få tilvejebragt information om stridspunkter vedrørende evaluanden. Guba og Lincoln (1989) skriver: *"The evaluator's task is to identify the information needed"*(ibid.73). Information er altså det fænomen, begreb eller panacea, der kan afgøre en konkurrence mellem konstruktioner. De skriver: *"...competing constructions...can be changed only through new information or increased sophistication"*(ibid.73).

Er informationer ikke konstruktioner? Hvad er forskellen på information og konstruktion? Igennem bogen får man indtryk af, at information er noget objektivt, der kan skabe mere sofistikerede konstruktioner. Der er en dualisme mellem information og konstruktion. Der er konstruktioner, mentale og sociale, og så er der informationer, som kan medvirke til mere sofistikerede konstruktioner. Er informationer data fra en mere objektiv virkelighed, som kan påvirke interessenternes konstruktioner? Guba og Lincoln (1989) fortæller, at mennesker har forskellig adgang til informationer, og har forskellige færdigheder i at behandle informationer (ibid.144). Det minder om Piagets teori om, hvordan børn lærer (Labinowicz 1980).

Det ligger uden for afhandlingen at drøfte, hvad information er, da det er en stor teoretisk diskussion (Qvortrup 1998:215-225; Rasmussen 1997:150-156). Information kan også forstås konstruktivistisk jf. Luhmanns operative konstruktivism (Luhmann 2000:79; Rasmussen 1997:133-156; Rasmussen 2004:247-255) og Batesons definition, hvor information er en forskel, der gør en forskel (Bateson 2000:315,459). Pointen er, at Guba og Lincoln (1989) ikke

redegør for, hvornår noget er hhv. informationer og konstruktioner samt definerer information.

Den manglende ontologiske klarhed viser sig også, når Guba og Lincoln (1989) introducerer begreberne etic og emic (Schwandt 2001:65-66). Etic er forskerens eller evaluators begreber, og emic er interessenternes. Hvis virkeligheden kun består af individuelle mentale konstruktioner, hvorfor da indføre distinktionen mellem etic og emic? Alle begreber bliver til emic konstruktioner, da verden kun består af unikke og individuelle menneskelige konstruktioner!

Forvirringen er total, når det ikke er klart om Guba og Lincolns (1989) konstruktivisme bygger på det enkelte individs mentale konstruktioner eller det er sociale konstruktioner. Hvis udgangspunktet er det enkelte individs unikke mentale konstruktioner, kan vi ende i solipsisme (Laughlin & Broadbent 1996:437-438; jf. Rasmussen 1997:125-131).

Det bliver vanskeligt at finde frem til grupper af interessenter i FGE, hvis hvert individ har unikke mentale konstruktioner af evaluanden. Forvirringen fremgår af følgende: *"For ontologically, it denies the existence of an objective reality, asserting instead that realities are social constructions of the mind, and that there exist as many such constructions as there are individuals (although clearly many constructions will be shared)"* (Guba & Lincoln 1989:43). Er det mentale- eller sociale konstruktioner? Kognitiv- eller socialkonstruktivisme? Hvorfor og hvordan er der opstået fælles konstruktioner? Hvilke processer gør sig gældende? Er det fælles konstruktioner, der skaber interessenter omkring en given evaluand? Hvor mange konstruktører er der? Guba og Lincoln (1989) skriver: *"There are as many constructions as there are constructors"* (ibid.144). Det er tautologisk. Der er lige så mange vandløb i Danmark, som der er vandløb!

Det virker som om, at Guba og Lincoln (1989) rundt omkring har opsamlet fragmenter af konstruktivisme. I forhold til den kognitive dualistiske konstruktivisme får man indtryk af, at de som uddannelsesfolk har hørt om Piaget. Spredt rundt i bogen anvender de Piagets væsentligste begreber, som f.eks. *"accommodation"* og *"assimilation"* (Labinowicz 1980; Piaget 2000) uden kildeangivelse eller at de har sat sig ind i begrebernes betydning. Eksempelvis: *"The case study helps readers to assimilate the joint construction into their own constructions..."* (Guba & Lincoln 1989:224). Assimilation er bl.a. udtryk for en lineær læreproces (Illeris 2000:27). Et andet sted skriver de: *"Sometimes such accommodation must be very radical; it may change the existing construction to such a degree as to deserve the appellation "revolution"* (ibid.213). Her er det en akkomodativ læreproces i Piagets forstand, men de blander det sammen, da akkomodative læreprocesser ofte er revolutionære (Illeris 2000:27). Inspirationen fra Piagets ligevægtsteori kommer til udtryk i

distinktionerne omkring informationsbehandling i bogens figur 5.1.(ibid.145-146), men uden kildehenvisning til Piaget.

Hvis Guba og Lincoln (1989) tilslutter sig en monistisk konstruktivisme, hvorfor da bruge tid på at forhandle og skabe konsensus mellem interessenter? Heap (1995) konkluderer, at Guba og Lincolns: *"social practices for developing a joint (linguistic) construction" are unintelligible in monist cognitive terms*"(ibid.58). Alle menneskelige konstruktioner er lige rigtige og legitime, da der ikke eksisterer en virkelighed per se. Men Guba og Lincoln (1989) mener, at der også kan være forkerte konstruktioner! De skriver: *"Of course, malconstructions are possible. For example, constructions may be incomplete, simplistic, uniformed, internally inconsistent, or derived by an inadequate methodology"*(ibid.143). Som omtalt, er vidundermidlet overfor forkerte konstruktioner ny information. Det tilvejebringes ved forhandling mellem interessenterne, men også iværksatte undersøgelser af evaluator. Men et sted i bogen kommer Guba og Lincoln (1989) med en ny løsning. De skriver: *"But constructions can be properly judged only by criteria appropriate to the paradigm out of which the constructor operates"* (ibid.143). Som læsere svæver vi i forhold til, om konstruktioner kan være sande eller usande. Guba og Lincoln (1989) skriver: *"All constructions must be treated as meaningful unless evidence to the contrary can be adduced"*(ibid.144). Hvad er det for et bevis, der kan sige, at en konstruktion ikke er meningsfuld? Citatet ovenfor antyder dualisme. At der eksisterer en objektiv virkelighed uafhængig af konstruktioner.

I Guba og Lincolns (1989) Hermeneutiske Dialektiske Proces, hvor interessenter forhandler om konstruktioner omkring evaluanden, kan grupper udelukkes, fordi deres konstruktioner er forkerte! Guba og Lincoln (1989) nævner børn, mentalt handicappede og folk med psykologiske vrangforestillinger: *"psychotics"*(ibid.150). I sidstnævnte tilfælde er det, fordi det er: *"self-deluded personalities"*(ibid.). De skriver videre: *"Constructions are self-delusional when they arise not from an interaction but solely from within the constructor's own mind"*(ibid.) Gode mentale konstruktioner kræver åbenbart interaktion, men interaktion med hvad og med hvem? Nærmer de sig en dualistisk socialkonstruktivisme?

At der er en virkelighed uafhængig af konstruktioner antydes kraftigt, når Guba og Lincoln (1989) skriver om deres metodologi. I forbindelse med evaluators undersøgelser nævner de, at man kan foretage observationer. De skriver: *"observational data relevant to the emerging constructions, for instance, that test them in practice, may be introduced"*(ibid.154). Her signalerer de ligefrem positivisme, når de foreslår, at observationer skal testes i praksis. Det fortsætter, når Guba og Lincoln (1989) skriver om sampling i forbindelse med, at evaluator skal opstarte FGE processen med interview af respondenter (ibid.178). Guba og Lincoln (1989) fanges af traditionel positivistisk samfundsvidenskabelig jargon. F.eks.: *"as each sample element is*

selected, each datum recorded..."(ibid.180). Her er det ikke konstruktioner, men data! De taler også om: "*item of information is identified...*"(ibid.182). Der er en diskrepans mellem FGEs ontologiske, epistemologiske og metodologiske anvisninger. Ontologien er monistisk, men metodologien dualistisk.

Heap (1995) giver flere eksempler på, hvordan Guba og Lincoln (1989) i deres kriterier for at bedømme god Fjerde Generations Evaluering glider over i en dualisme, hvor de forudsætter, at der eksisterer en objektiv virkelighed. De skriver, at evaluator skal foretage interview af interessenter som efterfølgende analyseres og dokumenteres. Interviewene bliver til data og objektiveres. Heap (1995) mener ikke, at Guba og Lincoln (1989) argumenterer for springet fra den ontologiske monisme til den metodologiske dualisme (Heap 1995:57).

Heaps (1995) kritik er hård. Han mener, at Guba og Lincolns (1989) monistiske konstruktivisme betyder, at der ikke findes en evaluand, men kun sociale konstruktioner af evaluanden. Vi skal så evaluere på en evaluand, der ikke eksisterer! Heap (1995) finder: "*evaluation inquiries necessarily are dualist*"(ibid.59). Kan vi evaluere på sociale konstruktioner med udgangspunkt i andre sociale konstruktioner? Hvilke kriterier skal vi anvende for at undersøge, om det er en god evaluering af en social konstruktion? Heaps (1995) konklusion er: "*extreme relativist, subjectivist rhetoric of monist constructionism should be dropped entirely from evaluation inquiry*"(ibid.59).

I forhold til diskussionen om en social konstrueret virkelighed er det et faktum, at selvom virkeligheden måske er socialt konstrueret, har konstruktionerne virkelige konsekvenser (Dahler-Larsen 2001B:334). Dette er Thomass teorem. Hvis mennesker definerer situationer som virkelige, er de virkelige i deres konsekvenser (Heap 1995:52). Derfor er en ontologisk diskussion i forhold til vurdering af en evalueringsteori måske mindre vigtig, og har kun akademisk værdi.

Det kan konkluderes, at der er uoverensstemmelser mellem ontologiske, epistemologiske og metodologiske overvejelser i FGE. Endvidere er det et problem, at det ontologiske udgangspunkt ikke beskrives klart. Det er u hensigtsmæssigt, at Guba og Lincoln (1989) blander flere konstruktivismer sammen. Det betyder, at FGE er en dårlig evalueringsteori ifølge Shadish et al. Det er positivt, at Guba og Lincoln (1989) som evalueringsteoretikere, har overvejelser om ontologi, epistemologi og metodologi. Andre evalueringsteoretikere undlader dette jf. Shadish et al. Det er sandsynligt, at FGE kun anvendes lidt i praksis, fordi det ontologiske udgangspunkt er for abstrakt, uklart og vanskeligt at anvende i praksis jf. afsnit 3.5. om evalueringspraksis.

3.2. Værdier og FGE

Ifølge Shadish et al. indeholder en god evalueringsteori overvejelser om, hvordan evaluanden skal vurderes (ibid.32). Hensigten er at give evaluatoren et teoretisk udgangspunkt, når der skal foretages en bedømmelse. I de værditeoretiske overvejelser kan evaluatoren få svar på, hvad et godt program er, og hvad godt betyder (ibid.35). En god evalueringsteori giver begrundelser for værdimæssige konklusioner.

I værdidiskussionen indfører Shadish et al. tre aspekter, når de analyserer evalueringsteoretikere. Det er hhv. meta-, præskriptiv- og deskriptiv teori. Ved metateori undersøges om teoretikerne har overvejelser omkring essensen af værdier, værdiers natur, hvorfor vi har værdier og hvordan værdier skabes (ibid.48). Præskriptiv teori er spørgsmålet, om teoretikerne advokerer for specielle værdier. Deskriptiv teori handler om, at teoretikerne beskriver værdier uden at de påstår, at nogle værdier er bedre end andre. Shadish et al. plæderer for deskriptiv teori.

I analysen af de syv evalueringsteoretikere finder Shadish et al. frem til, at det kun er Scriven, der har en metateori (ibid.48). Guba og Lincoln (1989) præsenterer ikke en metateori i FGE. Der findes ikke teori om værdier og deres rolle for vurderingen af en evaluand. Den konstruktivistiske ontologi udelukker selvstændige betragtninger omkring værdier, da værdier reduceres til sociale eller mentale konstruktioner. I stedet udvikler Guba og Lincoln (1989) Den Hermeneutiske Dialektiske Proces (ibid.149-155), hvor interessenter inddrages og forhandler med hinanden. I forhandlingerne om påstande, bekymringer og problemer vedr. evaluanden er målet konsensus. De fælles konstruktioner bliver indikator for, hvad der er viden, og hvad der er godt og skidt. Evaluatoren kommer ikke med selvstændige anbefalinger. Ifølge Shadish et al. vil en sådan deskriptiv værditeori fungere fint i vestlige pluralistiske samfund. Guba og Lincoln (1989) skriver også, at deres forhandlingsmodel er en tilpasning til værdipluralismen i samfundet (ibid.32).

Der kan argumenteres for, at det er problematisk, at Guba og Lincoln (1989) ikke har en præskriptiv teori, som giver argumenter for, hvorfor noget er godt eller dårligt. At der er konsensus blandt interessenter, gør ikke en vurdering økonomisk, politisk, juridisk, ideologisk, moralsk eller etisk forsvarlig. Det bliver en procedure etik, hvor noget bliver rigtigt og godt, hvis blot proceduren i Den Hermeneutiske Dialektiske Proces er overholdt. Guba og Lincoln (1989) begrundet ikke proceduren i f.eks. demokratiteori (Held 1993), Habermass teori om Kommunikativ Handlen (Andersen 1996:356-360; Bøndergaard 2001:124-132) eller i andre teorier. Disse aspekter behandles i afsnit 3.3.1. og 3.3.2.

Bortset fra Scriven er værdirelativisme et generelt problem for evalueringsteorien. Teorierne udelukker ikke, at vi som evaluatoren kunne

have evalueret for nazisterne i Det Tredje Rige. At vi kunne undersøge om en given koncentrationslejr levede op til målet eller undersøgte effekter og bieffekter. En FGE evaluering ville dog have inddraget jøderne som interessenter.

I artiklen *"Anything goes"* (Adelman 1996:291) behandles problematikken omkring relativisme. I forhold til Guba og Lincoln (1989) skriver han: *"These authors leave us with an indefinite relativism"*(Adelman 1996:293). Han mener, at det er uforeneligt at tilslutte sig postmodernisme og konstruktivism og så være evaluator (ibid.291). For Adelman handler evaluering om, at bedømme en evaluand i forhold til lighed og retfærdighed. For Scriven handler det bl.a. om at undersøge brugeres behov og bedømme om de tilgodeses (Shadish et al.1991:85-87). Som det påvises nedenfor, tager Guba og Lincoln (1989) ikke konsekvensen af deres ontologiske udgangspunkt, da de indirekte fokuserer på værdier. I forhold til Guba og Lincoln (1989) er det ikke helt korrekt, når Adelman (1996) skriver: *"Postmodernist discourse make no social judgements"*(ibid.299).

I 1992 skriver Guba en artikel med overskriften: *"Relativism"*. Tidligt i artiklen begår han et kunststykke ved at skrive: *"I specifically exclude consideration of moral, ethical, and cultural relativism, choosing instead to focus on that kind of relativism which relates to the disciplined acquisition of knowledge..."*(ibid.18). Relativisme er altså ikke i forhold til moral, etik og kultur, men i forhold til erhvervelse af viden.

Heuristisk kan der opstilles en firefeltstabel jf. figur 2. Spørgsmålstegnene i figuren er udtryk for, at det netop er heuristisk. I artiklen taler Guba (1992) kun om relativisme i forhold til viden. Han udelukker ikke, at der er absolutte værdier uden dog at forholde sig til spørgsmålet. I artiklen synes Guba (1992) at erkende den traditionelle distinktion mellem værdier og viden. En distinktion som Guba og Lincoln tog afstand fra i 1989, da den blev tilskrevet det konventionelle videnskabelige paradigme. I 1992 tager Guba afstand fra, hvad han kalder for radikal relativisme og hvor: *"anything goes"*(ibid.20). Han mener også, at fordi vi er usikre på vores viden, betyder det ikke, at vi ingen viden har (ibid).

Figur 2: Værdier og viden

	Absolut viden	Relativ viden
Absolutte værdier	Den klassiske religiøse naturvidenskabsmand (Isaac Newton)	Guba (1992)?
Relative værdier	Den moderne naturvidenskabsmand?	Guba & Lincoln (1989)

Det ligger udenfor afhandlingen filosofisk og videnskabsteoretisk at undersøge, om det er holdbart at tilslutte sig en konstruktivistisk relativ opfattelse af viden, og en mere absolut opfattelse af værdier. Det forekommer dog svært at advokere for, at viden er en social konstruktion, men værdier herunder moral, etik og kultur ikke er det. Det er ikke i overensstemmelse med den ontologi, der præsenteres i FGE fra 1989.

På det overordnede plan indeholder FGE ikke en metateori eller en sammenhængende præskriptiv teori i forhold til, hvordan en evaluand skal bedømmes. Paradoksalt er Guba og Lincolns (1989) evalueringsprocedure præskriptiv og begrundes indirekte i værdier. De mener, at FGE ikke har en: *"tendency toward managerialism"*(ibid.31). De finder, at de tre tidligere evalueringsgenerationer beskytter opdragsgivere og ledere (ibid:32) på bekostning af andre grupper. For at undgå dette skal alle interessenter inddrages i evalueringen. Det er både dem, der nyder godt af et program, men også dem som er ofre (ibid.14). I Den Hermeneutiske Dialektiske Proces er det vigtigt at alle deltager. Guba og Lincoln (1989) skriver: *"In a participative process a major role of the evaluator must be to attain the condition of participation for all – a political role, it should be noted"*(ibid.260-261). Guba og Lincoln (1989) advokerer for en partipatorisk model, uden at de knytter det op på en metateori om værdier.

Hvorfor er det vigtigt at alle interessenter deltager, når sandhed og værdier er sociale/mentale konstruktioner? Det er jo ikke for at komme tættere på sandheden eller finde frem til de rigtige vurderinger. Proceduren i FGE bygger på, at interessenter skal kende og fremføre egne interesser. Det ligger i begrebet *"Stakeholders"*. Fokus er ikke på sandhed og gode værdier, men egeninteresser i evaluanden. Guba og Lincoln (1989) skriver: *"Evaluation, like democracy, is a process that, to be at its best, depends upon the application of enlightened and informed self-interest"*(ibid.253).

Guba og Lincoln (1989) tilslutter sig implicit en eller anden lighedsopfattelse (jf.Rasmussen 1981). I forhandlinger mellem interessenter er det vigtigt, at interessenternes repræsentanter er lige dygtige, og har lige adgang til information. Hvis ikke det er tilfældet, skal evaluator gå ind og hjælpe. De skriver: *"When it is necessary, the evaluator will act as adviser and educator to the less skilled. We are aware that this appears to be an advocacy role, which some will resist, but we have already argued earlier that the proper and appropriate province of the evaluator is the empowerment of previously disenfranchised stakeholders..."*(ibid.247). Det handler om at de interessenter, der ikke tidligere blev hørt, nu skal høres. Det er en legitim opgave for evaluator, at hjælpe svage interessenter med at forstå deres egne interesser (ibid.247). Det virker som om, at Guba og Lincoln (1989) er klar over, at på trods af deres konstruktivistiske ontologi, eksisterer der en virkelighed, der har virkelige konsekvenser for virkelige mennesker jf. Thomas teorem afsnit 3.1.

Analysen viser en uoverensstemmelse mellem videns- og værdiopfattelse i FGE.

Målet med såkaldt empowerment går videre end blot, at interessenter skal høres. Guba og Lincoln (1989) skriver: *"Empowerment means the sharing of power tools and the sharing of leadership"*(ibid.227). De opregner en række fordele ved empowerment. De hævder, at det betyder øget tilskyndelse til at tage risikoer, udløser energi, stimulerer kreativitet, giver stolthed, skaber forpligtethed, medfører at man tager ansvar, og giver en følelse af ejerskab (ibid.226-225). Guba og Lincolns (1989) empowerede mennesker behøver ikke: *"...a kind of bureaucratic hierarchy that sees to it that each lower level does things "by the book". The person-organization distinction disappears"* (ibid.227).

Guba og Lincoln (1989) præsenterer ikke en værdimæssig metateori eller samfundsvidenskabelige forskningsresultater i forhold til empowerment. Det er kuriøst, at Guba og Lincoln igennem hele bogen fra 1989 tager afstand fra kausalitet, generalisering og nomotetiske påstande, men pludselig giver generelle bud på, hvad empowerment betyder for medarbejdere og organisationer. Pawson (1996) har i en kritisk artikel om konstruktivistisk evaluering forholdt sig kritisk til forskning som: *"communicatively securing the authentic voice of the subjugated"*(ibid.218).

I forhold til kravene hos Shadish et al. er det ikke entydigt, om FGE er en god eller dårlig teori, når det handler om værdier. Det, som taler for en god teori, er FGEs åbenhed overfor pluralisme. Ifølge den deskriptive værditeori beskrives interessenternes værdier. Evaluator lægger sig ikke fast på en given værdi. Det, som taler for en dårlig teori er, at der hos Guba og Lincoln (1989) mangler en værdimæssig redegørelse for den metodologi, de anbefaler, dvs. Den Hermeneutiske Dialektiske Proces. Guba og Lincoln (1989) kommer selv med værdidomme, hvor de ønsker empowerment af grupper, som normalt ikke bliver hørt. De mener, at evaluator skal hjælpe og uddanne svage interessenter til at blive bedre til at forstå deres egne interesser og fremføre disse. Det er prisværdigt, men de mangler at koble det til teorier, der hæver deres værdier ud over egen ideologi. Det kunne være demokrati- (Held 1993) eller velfærdsstatsteori (Esping Andersen et al.2002;Christiansen et al.2006:197-214). Guba og Lincoln (1989) problematiserer ikke konsekvenserne ved en metodologi, der påskønner interessenters eller brugers egne snævre interesser (Andersen 1998:61-64;Hoff 1993). Ej heller at interessevaretagelse kan ses som et nulsumsspil (Østerud 1995:44), eller at aktører handler ud fra snævre quid pro quo betragtninger.

Grundlæggende er der en uoverensstemmelse mellem videns- og værdiopfattelsen i FGE. En uoverensstemmelse mellem en konstruktivistisk ontologi og en normativ metodologi. Sidstnævnte begrundes indirekte i normative udsagn om lighed.

3.3. Sociale programmer og FGE

Ifølge Shadish et al. omhandler en god evalueringsteori viden om, hvordan sociale programmer og politikker bliver bedre og udvikler sig (ibid.32). Det er viden om sociale problemer og om, hvordan sociale programmer kan skabe løsninger (Donaldson & Lipsey 2006:59). Evaluators skal i evalueringsteorien få svar på en række spørgsmål (ibid.35), som f.eks., hvad er det for vigtige sociale problemer programmer kan løse? Kan programmer forbedres, og er det værd at gøre dette? Shadish et al. (1991:37) opstiller tre elementer, som evalueringsteorien skal berøre for at være en god teori. For det første, hvordan er den interne struktur i programmer, og hvordan fungerer de? For det andet, overvejelser om eksterne begrænsninger, der skaber og påvirker programmer. For det tredje, overvejelser om, hvordan sociale forandringer sker, hvordan forandringer sker i programmer, og hvordan programændringer medvirker til sociale forandringer.

I Guba og Lincolns bog fra 1989 findes der kun få overvejelser om sociale programmer. Det er i samsvar med bogens ontologiske, epistemologiske og metodologiske udgangspunkt. Programmer er sociale konstruktioner, som fortolkes forskelligt af forskellige interessenter i forskellige kontekster. Derfor kan Guba og Lincoln (1989) ikke skrive noget generelt om, hvordan programmer virker og ændrer sig. Som påvist tager Guba og Lincoln (1989) afstand fra kausalitet og generalisering. Metodologisk foretrækker de ideografiske casestudier frem for nomotetiske kvantitative analyser. I forhold til overvejelser om programmer står evaluator uoplyst efter at have læst bogen fra 1989. Det vurderes som positivt at Guba og Lincoln (1989) fremhæver det kontekstuelle i forhold til et program (evaluand). Det er dog problematisk, at de ikke gør sig generelle og abstrakte overvejelser omkring den kontekst et program (evaluand) indgår i, når det handler om pluralistiske vestlige demokratier. De påstår jo, at FGE er bedre egnet i pluralistiske samfund. Pawson (1996) rammer plet, når han skriver, at Guba og Lincolns konsensusforhandlinger mellem interessenter: *"belong to the cosy suburban world of the meeting between school governors, teachers and PTA (Parent-Teacher Association) in which they thrash out homework policy for the fifth grade"* (ibid.217).

Der skal fremdrages to forhold, som Guba og Lincoln (1989) burde have haft med i bogen om FGE. Der burde være en diskussion om, hvordan FGE fungerer i en styreform, der bygger på repræsentativt demokrati jf. afsnit 3.3.1. Der burde være en diskussion om magt i forhold til FGE jf. afsnit 3.3.2.

3.3.1. Repræsentativt demokrati og FGE

Repræsentative demokratier er karakteriseret ved, at der er folkestyre. I politologisk sammenhæng tales der ofte om Den Parlamentariske Styringskæde. Det er en idealtypisk model, der er udviklet af den norske

politolog Johan P. Olsen (Christensen et al.2006:29-30;Østerud 1995:67). Systemet består af fire institutioner. Et suverænt folk, parlament, regering og en forvaltning.

Modellen går i korthed ud på, at magten og legitimiteten udgår fra et suverænt folk, der igennem frie og lige valg mellem konkurrerende partier sammensætter et parlament. I Danmark er der et negativt parlamentarisk princip (1901) som betyder, at en regering ikke må have et flertal imod sig (Sørensen 1973:117-120). En regering har flere opgaver, men i idealiseret form er det dens hovedopgave at sørge for at love, der er vedtaget af parlamentet føres ud i livet. Der findes en række procedurer, som et parlament kan anvende til at kontrollere regeringen. Til sin rådighed har regeringen en forvaltning. Dens opgave er, i idealiseret form, at få implementeret lovene i praksis. I princippet er der et over-underordningsforhold mellem regering og forvaltning. Der hersker en hierarkisk organisationsform i forvaltningen med henblik på at sikre at de love, der er vedtaget af parlamentet føres korrekt ud i livet. Der er et legalitetsprincip (Garde 1981:131-132), som betyder at forvaltningens afgørelser skal have hjemmel i lov. Principielt findes modellen også i kommuner.

Den Parlamentariske Styringskæde udbygges ofte med et principal-agent-perspektiv (Christensen et al.2006:31;Sørensen 2005). Relationer mellem styringskædens aktører, og internt i forvaltningen som delegation fra et niveau til et andet, kan forstås som relationer: *"mellem en principal, der ønsker en bestemt aktivitet gennemført, og den agent, som principlen vælger at lade udføre aktiviteten"*(Christensen et al.2006:31). Problemstillingen med at nogle skal stå til ansvar overfor højere myndigheder eller politikere sammenfattes i begrebet: *"accountability"* (Day & Klein 1987:4-31).

Det er ikke hensigten at gå nærmere ind i Den Parlamentariske Styringskæde og dens utilstrækkelighed i forhold til virkeligheden (jf.Kelstrup 1995). Det samme gælder en diskussion af principal-agent perspektivet og accountability. I stedet er det hensigten at kaste lys over den overordnede kontekst, som mange sociale programmer indgår i. Det er ofte programmer, som er vedtaget af et parlament eller en kommunalbestyrelse. Da det ofte er programmer, som kræver mange skatteyderkroner, har programmerne målsætninger. Omkring de sociale programmer er der ofte opbygget et net af retskilder som love, bemærkninger til love, bekendtgørelser, vejledninger, tjenestemandsbefalinger, domme osv.(Jensen 1981:89ff.). Sociale programmer bliver ikke uden videre implementeret (Winter 1994), men under alle omstændigheder, er der en række begrænsninger i handlefriheden for Guba og Lincolns (1989) interessenter. Evaluanden (programmet) er ofte spundet ind i et regelgrundlag, som giver begrænsninger i empowerment og varetagelse af interessenters interesser. Det vurderes, at det normale er, at det ikke er muligt at handle over natten og ændre et program. Det skal dog

lige være, hvis det er evaluering af omfanget af lektier i 5. klasse på den lokale skole (Pawson 1996:217).

Med baggrund i Den Parlamentariske Styringskæde kan der argumenteres for, at interessenter allerede er repræsenteret via de politikere, de har stemt ind i parlamentet. Hvis der f.eks. er brugere, der er utilfredse med et program, har de via deres ytrings-, forenings- og forsamlingsfrihed samt stemmeret mulighed for at påvirke eller stemme andre politikere ind i parlamentet. Disse politikere kan så arbejde for at ændre programmet. Det Guba og Lincoln (1989) kalder for managerialisme i tidligere generationer af evaluering, kan forklares ud fra styringskæden og principal-agent perspektivet. Det er den offentlige leder i et hierarki, som står til ansvar (accountability), og er agent overfor et højere forvaltningsniveau, lovgivningen og politikere (principalen). Lederen er forpligtet på, at loven overholdes og målene i et program indfries. Derfor kan der argumenteres for, at lederen, forvaltningen eller politikerne har stor indflydelse på, hvordan en evaluering designes og gennemføres. Agenter og principaler har større indflydelse som interessenter end f.eks. brugere. En leder af et program kan også tænkes at have større interesse i en evaluand end en perifer bruger jf. afsnit 3.5. om evalueringspraksis. Det omvendte kan man dog også forestille sig.

Guba og Lincoln (1989) skriver flere steder, at det er forkert at tillægge enkelt personer et ansvar, fordi årsags-virkningsforhold ikke kan udredes jf. deres ontologi. De skriver: *"accountability is a characteristic of a conglomerate of mutual and simultaneous shapers, no one of which nor one subset of which can be uniquely singled out for praise or blame"*(ibid.110). Guba og Lincoln (1989) mener, at FGE i stedet vil føre til fælles ansvar og empowerment. De skriver: *"The second consequence of following the fourth generation model is that accountability yields to shared responsibility"*(ibid.256). De finder, at traditionelle evalueringer og organisationers fokus på accountability er et sygdomstegn (ibid.257). Accountability kommer til at handle om skyld (ibid.).

Guba og Lincolns (1989) bemærkninger om accountability er revolutionerende. De kommer med et eksempel på, at lærere holdes ansvarlige for at børn ikke lærer noget i skolen. Hvis de i stedet havde overvejet andre evaluander end et curriculum eller et lærebogssystem, ville de måske have set det problematiske i deres synspunkt. Hvis evaluanden er USAs militære invasion i Vietnam eller Irak, er det vel fornuftigt, at nogle kan holdes ansvarlige? Tænk, hvis evaluanden er et sygehus, hvor der er mange fejloperationer. Et sygehus, hvor der er snavset og fyldt med stafylokokker. Et plejehjem, hvor beboerne mistrives pga. dårlig pleje. Et domstolssystem, hvor der er korrupsion og magtmisbrug. Vil Guba og Lincoln i disse tilfælde mene, at fokus skal være på et fælles ansvar, fordi fælles simultane processer skaber verdenen jf. citatet ovenfor?(ibid.257). Disse fælles simultane processer er ofte finansieret af borgere, som over skatten afgiver forbrugsmuligheder og nytte. I et principal-agentforhold med fokus på accountability, er det vanskeligt at forestille sig, at

der skal være: *"equal positions of power"*(ibid.246) hos alle interessenter. I Den Parlamentariske Styringskæde er love vedtaget af politikere mere legitime end interessenters sociale konstruktioner af et program.

Kritikken af Guba og Lincolns (1989) opfattelse af accountability skal ikke forstås som en normativ kritik, da der kan siges mange gode ting i forhold til at inddrage interessenter. Det er en kritik af, at de ikke har redegjort for deres revolutionerende og spredte ideer i forhold til et repræsentativt demokrati. De henviser selv til, at FGE passer bedre til pluralistiske samfund. De kunne med fordel have præsenteret andre demokratimodeller og styreformer (Held 1993), hvor fokus f.eks. er på selvforvaltning og anarkistiske idealer om et samfund uden statsstyre (Lauritzen 1972:231-265).

I forhold til teorier om sociale programmer jf. Shadish et al., er FGE en dårlig teori, da en række centrale problematikker ikke berøres. Guba og Lincolns manglende forståelse for den politiske og juridiske kontekst kan være årsag til, at FGE kun anvendes lidt i praksis. Opdragsgivere er ofte i et principal-agentforhold, hvor der er brug for, at der kan placeres et ansvar mere end at mange forskellige interesser bliver inddraget og føler sig empowerede. Nogle kan have brug for at få undersøgt og legitimeret, at programmet (evaluanden) lever op til vedtagne målsætninger jf. teori om evalueringers anvendelse (Vedung 1998:209-233; Dahler-Larsen 2001A:passim). Anvendelse behandles i afsnit 3.4.

3.3.2. Magt og FGE

Som anført er evaluander normalt indlejret i politiske, juridiske og bureaukratiske strukturer, som sætter grænser for indflydelse og ændringer i en evaluand. Et forhold som Guba og Lincoln (1989) ikke beskriver nærmere. I bogen er der en manglende eksplicit omkring magt. Vi kan slutte negativt, at Guba og Lincoln (1989) er klar over magtforhold i traditionel management orienteret evaluering. Det er derfor, de foreslår FGE, som en alternativ evalueringsmetode. FGE fokuserer ikke på opdragsgivere og ledelser, men inddrager alle interessenter herunder opdragsgiver og ledelse, hvor alle i forhandlingerne skal have: *"approximately equal positions of power, not just in principle but also in practice"*(ibid.247). I forhandlingerne, skal repræsentanter for interessenterne forsøge at nå til enighed om deres påstande, bekymringer og problemer ved evaluanden. Guba og Lincoln (1989) er klar over, at det er en vanskelig proces, og at enighed ikke altid opnås. Selve processen kalder de for Den Hermeneutiske Dialektiske Proces.

Det er hensigten at foretage en kort teoretisk magtanalyse med udgangspunkt i Habermas (Andersen 1996;Bøndergaard 2001) og Foucault (Andersen 1999; Lindgren 1996), som fokuserer på kommunikation og diskurs. Der kunne anlægges andre mere politologiske perspektiver på magt, men de er fravalgt, da fokus for analysen er Den Hermeneutiske Dialektiske Proces. Vi kunne have

udført en analyse af magtens forskellige ansigter i forhold til FGE og sociale programmer (jf. Christensen & Jensen 1986; Thomsen 2000).

Det er et problem, at Guba og Lincoln (1989) ikke gør sig teoretiske overvejelser omkring magtforhold i deres foreslåede konsensusforhandlinger mellem interessenter. Som påvist, mener Guba og Lincoln (1989), at evaluator skal gå ind og hjælpe svage interessenter. Hvad med distinktionen mellem etic og emic? Er det ikke evaluators etic konstruktioner, der kommer til at dominere over svage gruppers emic konstruktioner? Vil nogle gruppers konstruktioner af virkeligheden ikke være stærkere end andres?

Guba skriver i 1992, at: *"The construction to be "believed" is that one which, in the opinion of those best able to make such a judgment, is the most informed and sophisticated"* (ibid.20). Når der ikke findes en objektiv virkelighed at referere til, kan man med lidt fantasi og politologisk viden forestille sig, at der åbner sig en pandora æske af muligheder for manipulation fra magtfulde interessenter. Findes der manipulation og løgn, når virkeligheden blot består af sociale konstruktioner? O'Neil (1995) stiller kritisk spørgsmålet: *"Who has the right amongst equal stakeholders to determine which constructions are less sophisticated than others?"* (ibid.13). O'Neil finder det naivt at tro, at man i en forhandlingsproces mellem interessenter kan suspendere magtforskelle (ibid.).

Guba og Lincolns (1989) terminologi omkring processen er uklar og uforenelig med konstruktivisme. De taler om en Hermeneutisk Dialektisk Cirkel (ibid.152) og en Hermeneutisk Dialektisk Proces (ibid.149-156). Der er flere forståelser af hermeneutik, men den generelle pointe er, at man ved fortolkning nærmer sig en sandhed (Lübcke 1983:184-185). Begrebet *"dialektik"* har også betydet flere forskellige ting igennem historien, men i Hegel og Marxs forstand er der ikke tale om konstruktivisme (Lübcke 1983:88-89). O'Neil (1995) skriver i forhold til Guba og Lincolns (1989) brug af begrebet *"hermeneutisk cirkel"*, at: *"It probably would have been wiser to have coined a more distinctive term for their concept"* (ibid.19).

Det er oplagt at udbygge Guba og Lincolns (1989) teori om en hermeneutisk dialektisk proces med Habermass teori om Kommunikativ Handlen. Dette foreslås af Laughlin og Broadbent (1996). De skriver: *"Habermas provides "benchmarks"...for conducting open discourse between actors about both understanding and any consequent actions"* (ibid.444). De mener, at Habermass model kan medvirke til en vigtig videreudvikling af FGE (ibid.445). Habermass model fokuserer på, at der ved hjælp af kommunikative regler kan skabes konsensus mellem aktører med forskellige interesser (Flyvbjerg 2001:88-98). Til forskel fra Guba og Lincoln (1989) er Habermas ikke konstruktivist, men tror på, at sandheden kan nås i en diskurs, hvor der opstår enighed på baggrund af det bedste argument (Flyvbjerg 2001:90). Flyvbjerg citerer Habermas for at skrive: *"the unconstrained, unifying, consensus-*

bringing force of argumentative speech" er "*central experience*"(ibid.90) i menneskers liv. Mennesker defineres som demokratiske væsner (ibid.). Den form for diskurs Habermas advokerer for, er ikke strategisk og bygger ikke på manipulation og kompromisser mellem forskellige interesser. Dette er delvis i overensstemmelse med Guba og Lincolns (1989) tanker om forhandlinger mellem interessenter omkring evaluanden. Bemærk dog pointen i afsnit 3.2, hvor Guba og Lincoln (1989) mener, at forhandlingerne skal tage udgangspunkt i interessenternes egne interesser. Som Guba og Lincoln (1989), forholder Habermas sig heller ikke til det indholdsmæssige i diskurserne, men kun til det institutionelle og processuelle.

Ifølge Flyvbjerg (2001) er problemet ved Habermas, at der er langt mellem ideal og virkelighed, mellem intentioner og praktisk implementering (ibid.92). Flyvbjerg (2001) mener, at Habermas lider af: "*insufficient conception of power*"(ibid.93). Habermas mangler en konkret forståelse af magtrelationer, som er nødvendig, hvis der skal skabes politiske forandringer (ibid.). Den samme kritik kan rettes mod Guba og Lincolns (1989) Hermeneutiske Dialektiske Proces og dermed FGE.

Ifølge Flyvbjerg (2001) mener Nietzsche, Foucault, Derrida og mange andre, at kommunikation mellem mennesker altid er penetreret af magt (ibid.93). Foucault mener: "*Power is always present*"(ibid.). Det er derfor meningsløst at operere med et kommunikationsbegreb, hvor magt er fraværende. Magtteoretikere mener: "*communication is more typically characterized by rhetoric and maintenance of interests*"(ibid.94). Foucault tror ikke på den kraft, der findes i det bedre argument (ibid.101). Flyvbjerg (2001) skriver: "*resistance, struggle, and conflict, in contrast to consensus, are for Foucault the most solid bases for the practice of freedom*"(ibid.102). For Foucault er Habermas en utopist (ibid.). Habermas betragter Foucault som en kyniker og relativist (ibid.). Efter gennemgangen af de to teoretikere konkluderer Flyvbjerg (2001): "*social conflicts are the true pillars of democratic society*"(ibid.108). Og videre: "*Political consensus cannot be expected to neutralize particular group obligations, commitments, and interests*"(ibid.). Denne viden er ikke ny og findes bl.a. hos Dahl og pluralistisk teori (Ham & Hill 1993:26-30). Flyvbjerg (2001) mener, at Habermas' ide om fornuftige argumenter og "*zwanglos*" (ibid.106) kommunikation er sociologisk naivt både empirisk og normativt.

Guba og Lincolns (1989) vægt på, at interessenter skal nå til en fælles konsensus konstruktion af påstande, bekymringer og problemer omkring evaluanden, er diskutabel jf. kritikken af Habermas. Som påvist kritiserer Guba og Lincoln (1989) tidligere generationer af evaluering for, at de ikke lever op til værdipluralismen i moderne samfund. Men hvorfor skal værdipluralismen betyde, at der skal opnås konsensus mellem interessenter med forskellige værdier og interesser? Når der ikke er en objektiv virkelighed som referencepunkt, er det endnu mere tvivlsomt, at interessenter bliver enige.

Flere steder i bogen antyder Guba og Lincoln (1989), at man ikke altid skal forvente konsensus. De ser konsensus som et ideal (ibid.221). De skriver: *"there is no guarantee that consensus in reconstruction will in fact occur"*(ibid.255). Udfaldet af evalueringsprocessen kendes ikke. De skriver i slutningen af bogen, at det er i orden, at: *"agree to disagree"*(ibid.253). Et værste tilfælde må være, at forhandlingerne medfører, at interessenterne for alvor bliver uenige om påstande, bekymringer og problemer omkring evaluanden, og at processen må stoppe pga. uenighed eller manglende tid og ressourcer. Situationen er måske værre end, da evalueringsprocessen startede. Dette scenario omtaler Guba og Lincoln (1989) ikke. De er klar over, at der kan være magtforskelle, siden evaluator skal uddanne og hjælpe svage grupper, men de undlader at problematisere magtaspektet.

FGE er en dårlig evalueringsteori, fordi der ikke gives en redegørelse for den kontekst sociale programmer normalt indgår i jf. afsnit 3.3.1. Det er problematisk, at Guba og Lincoln (1989) ikke problematiserer de magtforhold, der kan optræde i den evalueringsprocedure, de foreslår. De kobler ikke deres Hermeneutiske Dialektiske Proces til relevant samfundsteori som f.eks. Habermas. Den manglende og urealistiske opfattelse af kontekst og magt kan være en af forklaringerne på, at FGE ikke anvendes mere i praksis jf. afsnit 3.5. om evalueringspraksis.

3.4. Anvendelse og FGE

Ifølge Shadish et al. indeholder en god evalueringsteori overvejelser om, hvordan samfundsvidenskabelig viden anvendes til at ændre politikker og sociale programmer (ibid.32). Det er overvejelser om, hvordan evaluator kan sikre, at resultaterne bliver anvendt i forhold til programmet, men også fordele og ulemper ved forskellige anvendelser. De mener, at evaluatorer: *"need a theory to tell how, when, where, and why they can produce useful results"*(ibid.52). Bemærk pragmatismen jf. afsnit 2.3.1. En god evalueringsteori ekspliciterer og retfærdiggør, de valg en evaluator må tage i forhold til anvendelse.

Shadish et al. går ud fra, at en evalueringsteori grundlæggende tror på, at den er anvendelig til at løse sociale problemer. Alligevel noterer de, at: *"some theorists have no explicit theory of use at all"*(ibid.54). Ifølge Shadish et al. har teorier om anvendelse tre elementer. For det første, en beskrivelse af mulige anvendelser af evalueringen. For det andet, en beskrivelse af tidsrammen for anvendelse af evalueringen. Og endelig for det tredje, en forklaring på, hvad evaluator kan gøre for at facilitere anvendelse af evalueringen (ibid.52).

Shadish et al. giver en distinktion mellem tre anvendelsesformer (jf. Vedung 1998:209-233; Dahler-Larsen 2001A:passim). Den første er instrumentel anvendelse. Her anvendes evalueringsresultater til at: *"making direct decisions*

abort changing programs"(Shadish et al.1991:53). Den anden anvendelse er, når evalueringer påvirker, hvordan folk tænker over en given evaluand. Weiss taler om: "*enlightenment*"(Weiss 1977;Shadish et al. 1991:53). Andre taler om: "*conceptual use*"(Shadish et al.1991:53). Endelig nævner Shadish et al.: "*persuasive*" anvendelse (ibid.), hvor det handler om, at evalueringer anvendes til at overbevise og legitimere forhold.

En god evalueringsteori har ifølge Shadish et al. også en distinktion om anvendelse på kort- og langt sigt. Hvis en evalueringsteori ikke har overvejelser omkring anvendelse, er der risiko for, at evaluator producerer information, som ikke kan eller bliver anvendt. Det ligger udenfor afhandlingen at redegøre for avancerede overvejelser om anvendelse, som f.eks. evalueringers konstitutive virkninger (Dahler-Larsen 2001A:passim).

I forordet skriver Guba og Lincoln (1989), at FGE i større udstrækning vil blive anvendt i forhold til tidligere generationer af evaluering (ibid.10). Årsagen er, at FGE er handlingsorienteret og inddrager interessenter. De mener: "*Very often the evaluation "product" is a set of recommendations that suits only the purpose of the evaluator and the client*"(ibid.). Skævheden betyder, at evalueringen ikke bliver anvendt og skaber handlinger. Processen i FGE betyder, at interessenter: "*...find a reason to support it, work at it, and feel good about it*"(ibid.10), og dermed at evalueringen bliver anvendt.

Fokus på handling i FGE betyder, at der er en instrumentel opfattelse af anvendelse. Oplysning og konceptuel anvendelse udelukkes implicit, fordi Guba og Lincoln (1989) ikke tror på generalisering og nomotetiske forhold. De fokuserer på den aktuelle kontekst og aktuelle konstruktioner af evaluanden. De skriver: "*Fourth generation evaluations never stop; they merely pause*"(ibid.226). De mener, at en FGE-proces i princippet aldrig bliver færdig. De skriver, at FGE-processen rejser flere nye spørgsmål end der besvares. Der opnås ikke enighed hos interessenterne om alle påstande, bekymringer og problemer omkring evaluanden. Ligeledes, at: "*the emergent construction is likely to hold only for a brief time*"(ibid.226). Dermed er anvendelse af evalueringer kortsigtet. Polemisk kan vi stille spørgsmålet om, hvor afgrænset en kontekst skal være i tid og rum. Hvis konteksten defineres til situationen kl.14.43, den 1. april 2008 på Sennels skole, hvor repræsentanter for lærere, ledelse og forældre diskuterer en evaluand, kan konstruktionerne da anvendes kl. 8.32, den 2. april 2008 på hele skolen?

Anvendelsesspørgsmålet er også afhængigt af, om det er summativ eller formativ evaluering (Scriven 1991B). Guba og Lincoln (1989) lægger vægt på formativ evaluering (ibid.190). Anvendelse sker i selve processen. Derfor lægger de ikke vægt på afsluttende evalueringer og skriftlige anbefalinger. Når fokus er på formativ evaluering, er det mindre vigtigt, om evalueringen på længere sigt kan skabe oplysning eller begrebslig udvikling.

FGE kan forstås som en kritik af evalueringer, som anvendes til at overbevise og legitimere uden at relevante interessenter inddrages, og at der skabes handlinger.

Guba og Lincoln (1989) er eksplicitte i forhold til overvejelser om anvendelse. Alligevel er der tale om: *"one size fits all"*. Det er en omfattende proces, der skal i gang for at skabe den nødvendige anvendelse jf. næste afsnit om evalueringspraksis. Guba og Lincoln (1989) giver ikke eksempler på instrumentel anvendelse. De blander den instrumentelle anvendelse med fokus på handling sammen med normative overvejelser om empowerment. De siger intet om, at FGE kan anvendes til at skabe social udvikling eller effektivitet. De giver ikke empirisk belæg for at tidligere generationer af evaluering ikke anvendes udover, at de henviser (ibid.10) til debatten mellem Patton (1990) og Weiss (1990A & B).

Som det er påvist i afsnit 3.3.2 om magt, kan konsensus omkring påstande, bekymringer og problemer vedrørende evaluanden forekomme naive. Uenighed kan måske betyde, at evalueringsresultater ikke anvendes eller ikke skabes. I forbindelse med anvendelse skriver Shadish et al.: *"one obstacle is that evaluative results often threaten entrenched interests. Evaluation always has an adversarial relationship with some parties"*(ibid.54). Dette aspekt har Guba og Lincoln (1989) ikke behandlet.

Et filosofisk argument for at evalueringsresultater i FGE vil blive anvendt er, at ved at inddrage interessenter er der større chance for, at der bliver produceret phronetisk viden. Guba og Lincolns (1989) konstruktivisme går dog hinsidig en anerkendelse af phronetisk viden, som blot bliver en konstruktion. Ifølge Flyvbjerg (2001) kan Aristoteles begreb om phronesis forstås som: *"prudence or practical wisdom"*(ibid.2). Viden som umiddelbart kan skabe handling (jf. Schön 2001; Dreyfus & Dreyfus 1991). I forhold til den tidligere magtdiskussion, mener Flyvbjerg (2001), at phronesisbegrebet skal videreudvikles til også at inddrage magt (ibid.3). Interessenters hverdagsviden (jf. Schutz 2005:80-82) kan således være meget pragmatisk, konservativ og bygge på snævre egen interesser. Desværre udelukker Guba og Lincolns (1989) ontologi et fornuftigt argument om, at processen med inddragelse af interessenter i forhandlinger om en evaluand kan producere anvendelig og praktisk viden. I stedet bliver det relative og tidsmæssige begrænsede: *"joint constructions"*(ibid.221). Guba og Lincoln (1989) skriver: *"Negotiations end when some consensus is reached on each previously unresolved CC&I (påstande, bekymringer og problemer). In effect, a new joint construction is reached by consensus that supersedes previously developed constructions"*(ibid.221). Koblingen mellem phronetisk viden og evaluering findes hos Schwandt (Schwandt 2005:102; Schwandt & Burgon 2006).

I forhold til, hvad evaluator kan gøre for at facilitere anvendelse af evalueringen skriver Guba og Lincoln (1989), at det er vigtigt at der indgås en

kontrakt med opdragsgiveren (ibid.188-195). Denne skal forpligte sig på FGE og Den Hermeneutiske Dialektiske Proces, hvor der forhandles med alle relevante interessenter (ibid.191). Opdragsgiveren skal skrive under på: *"willingness to share power, willingness to change in the face of persuasive negotiations, willingness to reconsider value positions as appropriate, and a willingness to make the commitments of time and energy that be required"* (ibid.191).

I forhold til kravene hos Shadish et al. gør Guba og Lincoln (1989) sig overvejelser om anvendelse af deres evalueringresultater. De giver detaljerede anvisninger på, hvad kontrakten med opdragsgiveren skal indeholde. Teoretisk er FGE en god evalueringsteori, fordi den beskæftiger sig med anvendelsesproblematikken. Den fokuserer på instrumentel anvendelse på kort sigt. Men der er ingen eksempler fra praksis. Det er problematisk at evalueringresultater, ifølge Guba og Lincolns (1989) ontologi, ikke kan anvendes udenfor en afgrænset kontekst og er meget midlertidige. Det kan måske forklare, hvorfor FGE ikke anvendes meget i praksis. Det er ærgerligt, at FGE ikke ontologisk og epistemologisk kan værdsætte interessenternes phronetiske viden.

3.5. Evalueringspraksis og FGE

Afsnittet går videre end den teoretiske analyse hos Shadish et al., og inddrager praktiske erfaringer med FGE. På den måde bliver det belyst, om FGE er realistisk.

Som det er fremgået, er evalueringspraksis den vigtigste komponent hos Shadish et al. Som det ses af figur 1 side 16, bidrager de fire andre komponenter, viden, værdier, sociale programmer (genstandsfelt) og anvendelse til evalueringspraksis. Shadish et al. skriver: *"Theories of practice – the most essential component in theories of evaluation"*(ibid.57). De nævner en række praktiske spørgsmål, som evalueringsteorien skal hjælpe med at give svar på. Eksempelvis, tilsiger situationen, at der skal foretages en evaluering? Hvad skal formålet være med evalueringen? Hvilken rolle bør evaluator spille? Hvilke spørgsmål skal stilles? Hvilke evalueringsdesign skal vælges? Hvilke aktiviteter skal gennemføres for at lette evalueringens anvendelse?

Teorier om evalueringspraksis skal indkludere: *"advice about what to do and how to do it"*(Shadish et al.1991:60). Gode teorier om evalueringspraksis skal være dækkende, men også realistiske (ibid.61). Shadish et al. omtaler praksisteorier, som er vanskelige at implementere i praksis, fordi evaluatorene arbejder i situationer, der indeholder en række begrænsende faktorer.

Mange af spørgsmålene under evalueringspraksis er behandlet i de forrige afsnit jf. kritik af begrebsrammen i afsnit 2.2.1. Fokus under

evalueringspraksis er, om evalueringsteorien gør det muligt at foretage en prioritering i den konkrete situation, hvor tid, ressourcer, færdigheder, interesser og bemanding er begrænsede (ibid.61). Praksisteorier skal hjælpe evaluator med at foretage valg og begrunde og legitimere disse valg.

Shadish et al. indfører en distinktion mellem fokusovervejelser og kontingentovervejelser (ibid.62). Praksisteorier kan tage udgangspunkt i, at der er særlige aspekter evaluator skal fokusere på i en situation med begrænsninger i tid, ressourcer og færdigheder. Det kan betragtes som minimumskrav til en evaluering. Kontingentovervejelser er mere nuancerede, da der her tages hensyn til forskellige situationer. Det er anvisninger på, hvad evaluator skal gøre i forskellige situationer.

På trods af at Guba og Lincoln (1989) skriver, at et evalueringsdesign skal være åben og udvikle sig igennem processen indeholder FGE ikke en kontingent praksisteori. At Guba og Lincoln (1989) holder design, proces og resultatet åben, hjælper ikke evaluator i forhold til at træffe valg og foretage prioriteringer. Der opstilles ikke forskellige kontingente situationer med forskellige anbefalinger til evalueringspraksis. Er det f.eks. altid nødvendigt at inddrage alle interessenter? Er det nødvendigt altid at gennemføre Den Hermeneutiske Dialektiske Proces? En proces, hvor interessenterne først internt klargøre deres konstruktioner i forhold til påstande, bekymringer og problemer i forhold til evaluanden. Herefter vælger de repræsentanter, som forhandler med udgangspunkt i en dagsorden opstillet af evaluator med de punkter, hvor der ikke er enighed.

Man får indtryk af *"one size fits all"* uanset evaluand og kontekst. En kritik der kan rettes mod mange andre evalueringsteorier jf. kapitel 2.0. I forhold til Guba og Lincoln (1989) er det som om, at der altid skal bygges en supertanker, hvor vi i situationer måske kunne nøjes med en jolle. Guba og Lincoln (1989) skriver: *"It may be a difficult and cumbersome process, and may take much longer than we are accustomed to thinking evaluations should take. More resources may be required. But in the end, it works, and when it works, things happen"*(ibid.227). Her imod slutningen af bogen indrømmer Guba og Lincoln (1989), at FGE ikke har en lang historie i praksis. De skriver: *"We ourselves are just beginning to explore this form of evaluation, and our experience is nearly as limited as the readers"*(ibid.262)(sic). En praksisteori som antyder, at evalueringsprocessen stopper, når der ikke er flere ressourcer tilbage, vurderes her som svag og vanskelig at sælge overfor en opdragsgiver.

Der findes flere artikler som foreslår, at FGE skal anvendes og evt. modificeres. Det er karakteristisk, at en fuldstændig FGE-proces lige efter bogen ikke synes gennemført ifølge artiklerne. Nævnes kan artiklerne: *"Redesigning Fourth Generation Evaluation"*(Laughlin & Broadbent 1996) og: *"Beyond measurement: fourth-generation evaluation in nursing"*(Koch 1994).

I 2007 skriver Lay og Papadopoulos en artikel i tidsskriftet: "*Evaluation*" med titlen: "*An Exploration of Fourth Generation Evaluation in Practice*". De går desværre ikke systematisk igennem, hvad der er publiceret om FGE i praksis. De nævner f.eks. ikke Kochs studier indenfor sundhedsområdet (1994;2000). De skriver, at der kun er: "*few published studies describing the application of FGE methods to an actual evaluation*"(Lay & Papadopoulos 2007:497). Derfor er det vanskeligt at sige noget om, hvordan FGE fungerer i praksis. Om praksisteorien i FGE er tilstrækkelig. Omvendt bestyrker det afhandlingens påstand om, at FGE kun bliver anvendt lidt i praksis jf. kapitel 1. Lay og Papadopoulos (2007) mener, at FGE ikke er særlig kendt eller accepteret. FGE er ikke en del af mainstream indenfor evalueringsmetodologien.

Det er slående, at Lay og Papadopoulos (2007) i deres evaluering af udsatte børn ikke slavisk gennemfører Den Hermeneutiske Dialektiske Proces (ibid.503). De er dog begejstret for FGE herunder inddragelsen af interessenter. De mener, at FGE sandsynligvis har medført et: "*high commitment to the change process*"(ibid.503). De konkluderer, at FGEs: "*value in facilitating a responsive, empowering and educational evaluation process was clear*"(ibid.503). De anbefaler, at opdragsgivere og ledere anvender FGE (ibid.). Ikke mindst fordi de i overensstemmelse med Guba og Lincoln (1989:15) mener, at konventionel evaluering er moralsk og etisk forkert, men også politisk naiv og begrebslig snæver. Lay og Papadopoulos (2007) giver ingen klar dokumentation for, at deres FGE har været en succes. De citerer ikke udsagn fra interessenter herunder ledere og opdragsgivere.

Lay og Papadopoulos (2007) henviser til en artikel af Huebner og Betts fra 1999. Den har titlen: "*Examining Fourth Generation Evaluation*". Huebner og Bett skriver i 1999: "*A recent review of the literature revealed no published studies documenting FGE as an evaluation strategy*"(ibid.348). De ser også bort fra Kochs (1994) studier indenfor sundhedsområdet. Huebner og Bett (1999) har selv udført FGE i forbindelse med et projekt om positiv udvikling af unge mennesker. Også Huebner og Bett (1999) mener, at det er konstruktivt, at interessenter inddrages. Dog finder de det vanskeligt at definere og afgrænse, hvem interessenterne er i det aktuelle program (ibid.350). Ligeledes iagttager de, at interessenterne har forskellige evner, og ikke er lige dygtige til at formulere sig. De finder det vanskeligt at dokumentere, at interessenterne lærer noget af processen (ibid.351). Det samme gælder dokumentation for, at der er sket empowerment hos svage interessenter (ibid). I forhold til læring og empowerment mener Huebner og Bett (1999), at Guba og Lincoln er uklare, og at de: "*provide little guidance*", og: "*offer little advice*"(ibid.351). Deres erfaring er, at det er urealistisk og for idealistisk, at interessenter skal blive enige om en enkel fælles konstruktion omkring evaluanden (ibid.350). De konkluderer, at FGE har været anvendelig og god, fordi interessenterne blev inddraget. Hvis der sker modifikationer i forhold til de kritikpunkter, de har rejst af FGE, mener de: "*FGE could become a powerful method in any program dependent upon community/stakeholder support*"(ibid.352). FGE sikrer, at

interessenternes forventninger bliver hørt, og at et program får støtte fra starten af. De mener: *"support is vital to insure program success"* (ibid.352).

Hverken Lay og Papadopoulos (2007) eller Huebner og Bett (1999) forholder sig til de videnskabsteoretiske overvejelser ved FGE jf. afsnit 3.1. De problematiserer ikke ontologien og epistemologien bag FGE. At interessenter støtter et givet program kan næppe være et mål per se, hvis programmet objektivt er dårligt jf. 3.1.

Som nævnt har Lay og Papadopoulos (2007) ikke fundet frem til en artikel af Koch (2000) fra Australien med titlen: *"Having a say: negotiation in fourth-generation evaluation"*. Koch henviser til tre evalueringer indenfor sundhedsområdet, hvor FGE er anvendt. Koch dokumenterer forskellige praktiske problemer ved FGE.

I det første studie handler det om evaluering af den pleje ældre personer får, som er kommet på hospitalet (Koch 2000:120). Hun iagttager, at patienterne er tøvende i forhold til at fortælle, hvordan de oplever negative erfaringer med den pleje, de modtager på hospitalet. I interviewene skal de ældre konstant bekræftes i, at interviewene er anonyme. Det er i modstrid med FGE, hvor de forskellige interessenter skal forhandle om deres konstruktioner af evaluanden. Koch (2000) skriver: *"Patients feared exposure and victimization by nursing staff. Their involvement in interviews had the potential to make things worse (dis-empowered) rather than better (empowered)"*(ibid.120). De ældre patienter ønsker ikke at deltage i en hermeneutisk dialektisk proces med sygeplejersker eller hospitalledelsen. Forfatteren har oplevet lignende processer i forbindelse med evaluering af et leve-bo-miljø (Larsen 2006;2007B).

I den anden evaluering Koch (2000) nævner, handler det om at udvikle et plejehjem. Koch (2000) skriver: *"Whilst all nursing home staff were engaged in the interview and negotiation process, only seven residents (22%) were able to articulate their concerns. The other 78% had varying degrees of dementia"* (ibid.121). Koch konkluderer: *"Although we believed that the FGE approach was feasible in nursing homes, it was shown to have serious limitations"*(ibid. 121).

Det er en svaghed, at Guba og Lincoln (1989) ikke redegør for, hvad man nærmere skal gøre, hvis det er meget svage personer, der er interessenter, og som ikke ønsker at deltage i forhandlinger. Endnu værre er det, hvis det f.eks. er demente plejehjemsbeboere. Kan evaluator med etic konstruktioner varetage deres interesser i forhandlingerne? Kan demente forhandle sig frem til en fælles konstruktion med plejepersonalet i forhold til den pleje, de får?

Guba og Lincolns FGE er åbenbart ikke lige velegnet til alle evaluander. Den implicitte strategi om *"one size fits all"* har problemer, når konteksten er et

plejehjem eller et hospital. Guba og Lincolns forhandlingsmodel, som har konnotationer til forhandlinger på arbejdsmarkedet, virker sandsynligvis kun, hvis interessenterne er stærke og velartikulerede. Men så er det ikke længere empowerment af svage grupper.

Den tredje evaluering Koch (2000) omtaler, handler om et projekt vedrørende sygeplejerskers uddannelse og karriere i Sydaustralien. Uden at gå i detaljer er der 70 interessenter. Koch (2000) iagttager også her nogle praktiske problemer ved Den Hermeneutiske Dialektiske Proces, hvor interessenter skal forhandle sig frem til en fælles konstruktion af påstande, bekymringer og problemer ved evaluanden. Nogle interessenter har større interesse i evaluanden end andre. Det betyder, at der er interessenter, som ikke altid møder frem til møder og fælles forhandlinger (ibid.123). Koch (2000) finder, at nogle grupper dominerer processen og er mere: *"outspoken"* (ibid.). Omkring magt skriver Koch (2000): *"The evaluators observed the power differentials and there were quite striking differences in the negotiating strength of the various stakeholders....this meant that some voices were not as well heard as others."* (ibid.123). Koch mener, at spørgsmålet om interessenters forskellige magt er meget kritisk i FGE (ibid.124). Hun nævner, hvordan en stærk fagforening er en hindring for Den Hermeneutiske Dialektiske Proces og ændringer i et program. I de tre eksempler på evalueringer opstår der ikke konsensus mellem interessenterne (Ibid.124).

På baggrund af erfaringer fra de tre evalueringer giver Koch (2000) en række anbefalinger, som kan gøre Guba og Lincolns teori mere kontingent. Koch (2000) mener, at man skal overveje, hvad det er for omgivelser FGE skal anvendes i (ibid.125). Er det en mindre organisation, et hospital eller et nationalt projekt? Det er vigtigt at styre processen med interessenter, som ikke må udgøre mere end ca. 40 til 50 personer (ibid.). Man skal sikre sig en forpligtelse hos alle interessenter i forhold til, at de reelt ønsker at forbedre situationen (ibid.). Man skal finde artikulerede interessenter, som kan formulere deres påstande, bekymringer og problemer (ibid.). Man skal forhandle sig frem til en fornuftig tidsramme. Endelig skal man anvende dygtige evaluatore, som er i stand til at styre magtforskelle mellem interessenter (ibid.). Koch (2000) slutter artiklen med at skrive: *"I still believe that FGE has much to offer health and human service evaluation"* (ibid.125).

På baggrund af en evaluering af et naturvidenskabeligt uddannelsesprojekt skriver O'Neil i 1995 en artikel med titlen: *"Implementation Frailties of Guba and Lincolns Fourth Generation Evaluation Theory"*. Projektet handler i korthed om, at lærere i Sydafrika skal blive bedre til at undervise i naturvidenskab. Projektet er ikke færdigevalueret, da O'Neil skriver artiklen. O'Neils artikel nævnes ikke af Lay og Papadopoulos (2007) eller Huebner og Bett (1999). Der er flere grunde til, at man vælger FGE. O'Neil (1995) skriver bl.a.: *"The attractiveness of the theory lies in its promises of widespread stakeholder empowerment, equal valuing of the constructions of all stakeholders,*

consensual constructions on the "claims, concerns and issues" of most stakeholders, and guaranteed usage of the evaluation process and product" (1995:5-6).

O'Neil (1995) nævner, at det var umuligt at etablere: "*conceptual parity*" (ibid.8) mellem interessenterne. Grupperne havde ikke samme forståelse af FGE. O'Neil (1995) mener, at det er vanskeligt at etablere en forståelse for konstruktivisme hos interessenter (ibid.10). Derfor er der interessenter, der ikke forstår processen, og derfor ikke deltager med engagement. De fleste interessenter tænker: "*Yes, but when is the real evaluation going to begin?*"(ibid.10). De interessenter, der ikke forstod Den Hermeneutiske Dialektiske Proces, havde ikke ejerskab til evalueringen (ibid.11). Det var svært for de ellers veluddannede interessenter, at forstå den mystiske specialist jargon og ontologi bag konstruktivisme. Interessenterne blev usikre ifølge O'Neil (1995).

Repræsentanter fra regeringen og NGO'er blev ofte skiftet ud eller mødte ikke op til forhandlingerne i Den Hermeneutiske Dialektiske Proces. O'Neil (1995) skriver: "*The levels of participation of this entire group continued to be lower than I had hoped for*"(ibid.8). Som nævnt påpeger Koch (2000) også problemer med fremmødet hos visse interessenter. O'Neil (1995) skriver: "*Many of our stakeholders were passive and nonchalant towards the process from beginning to end*"(ibid.14). O'Neil (1995) finder det naturligt, at der er interessenter, for hvem evaluanden kun udgør en lille del af deres liv. Derfor er deres viden og engagement i processen begrænset (ibid.14). I projektet var forhandlingerne mellem interessenterne ikke gennemført. Årsagen var, at processen med formulering af interessenternes påstande, bekymringer og problemer har været for krævende for deltagerne. O'Neil (1995) skriver: "*A break was necessary if evaluation fatigue was not to make a serious impact on the process*"(ibid.9). O'Neil (1995) finder, at Guba og Lincolns (1989) tekst omkring interessenters deltagelse er urealistisk og uhensigtsmæssig.

Som Koch (2000) antyder ovenfor, ser O'Neil (1995) også problemer ved afgrænsningen af interessenter. De har ikke samme interesse i evaluanden og derfor ikke samme viden og engagement. Endvidere er det vanskeligt at tale om interessenter med fælles konstruktioner som gruppe. Snarere er der tale om individuelle personer med forskellige konstruktioner af påstande, bekymringer og problemer ved evaluanden. O'Neil (1995) oplever større forskelligheder af konstruktioner internt hos en af interessenterne (repræsentanter fra regeringen) end mellem interessenterne. Logisk bliver afgrænsningen af interessenter endnu mere uklar, når Guba og Lincolns ontologi, som påvist i afsnit 3.1. tangerer til solipsisme. Hvert individ danner sine egne mentale konstruktioner. Hvordan kan vi da have grupper af interessenter med fælles konstruktioner? Guba og Lincoln (1989) refererer ikke til teori om grupper, stratifikation og sociale klassers dannelse (jf. Haralambos 1980;Giddens 1993).

På trods af O'Neils (1995) praktiske erfaringer med FGE, er han positiv overfor Guba og Lincolns (1989) ide om at inddrage konstruktivisme i evalueringsteori (ibid.20). Han skriver: "*a place where is has an obvious application*"(ibid.).

I 1991 skriver Lai en kort artikel med titlen: "*Field-Based Concerns about Fourth-Generation Evaluation Theory*", som heller ikke er omtalt af Lay og Papadopoulos (2007) eller Huebner og Bett (1999). På baggrund af den umiddelbare elegance, hvormed FGE er formuleret forsøgte universitetet på Hawaii (CRDG) at anvende nogle af Guba og Lincolns procedurer i en række kontraktevalueringer (ibid.3).

Lai (1991) skriver, at der opstod en række praktiske problemer. Der er en række formelle krav, der skal overholdes, når det handler om evalueringer finansieret af det føderale niveau i USA. De tager ikke udgangspunkt i FGE. Ofte ønskes hårde kvantitative data (ibid.3). Lovgivere har begrænset tid, og kan ikke indgå i tidskrævende interessentforhandlinger. I virkelighedens verden lytter lovgiverne til dem, som kan producere simple, korte og definitive evalueringer. Det er dem, som får evalueringsopgaverne ifølge Lai (ibid.4).

Lai (1991) finder det vanskeligt at anvende FGE, når der er andre stærke interessenter, som tilslutter sig positivisme og andre evalueringsmodeller. Lai (1991) mener, at FGE er en: "*...academic exercise that is basically not applicable in the real world*"(ibid.6).

Sammenfattende viser de empiriske artikler, at der er problemer med FGE i forhold til evalueringspraksis. Guba og Lincolns praksisteori er ikke kontingent uanset at de lægger vægt på kontekst. Evalueringen på et plejehjem i Australien med mange demente beboere er forskellig fra en evaluering af et naturvidenskabeligt curriculum i Sydafrika, som igen er forskellig fra en føderal evaluering i USA. Sammenfattende kan der konstateres problemer med Den Hermeneutiske Dialektiske Proces i praksis. Det gælder afgræsningen af interessenter, deres deltagelse og indsigt. FGE viser sig i praksis at kræve meget tid. Ingen af eksemplerne fra FGE i praksis har stringent overholdt den fulde procedure i Den Hermeneutiske Dialektiske Proces.

På baggrund af terminologien i Shadish et al. kan det konkluderes, at der er mangler i FGEs praksisteori. Empirien antyder, at FGE er en dårlig evalueringsteori. Problemerne i praksis kan sandsynligvis forklare, hvorfor FGE ikke anvendes mere i praksis.

4.0. Konklusion

Udgangspunktet i afhandlingen er en undring over, hvorfor FGE ikke bliver anvendt mere i praksis, når der er tale om såkaldt Fjerde Generations Evaluering. Afhandlingens tese er, at årsagen skyldes, at FGE er en dårlig evalueringsteori.

Der findes ikke en fælles metaevalueringsteori, som giver et entydigt svar på, hvad der er god eller dårlig evalueringsteori. Der er tale om en ny disciplin. Der hersker forvirring, og der mangler en fælles begrebsramme. Der er ikke enighed om, hvad evaluering er, og hvordan evaluering skal defineres. Der mangler en fælles vision om, hvad evaluering skal bidrage med i forhold til samfundsudviklingen. Derfor er det vanskeligt at afgøre om FGE er en dårlig evalueringsteori.

Evalueringsteori er præget af amerikansk pragmatisme. Det skal fungere i en verden, hvor der er sælgere og købere af evaluering. Sidstnævnte forventer resultater. Da verden er mangfoldig, er det vanskeligt at etablere en flot forkromet fælles metaevalueringsteori, hvorfra vi kan bedømme de enkelte evalueringsteorier. Problemet løses pragmatisk ved at definere evaluering, som det evaluatorene fortager sig. Ud fra traditionel videnskabelig tænkning er det en "slap" tautologisk definition. FGE bliver til en god evalueringsteori, hvis FGE anvendes i praksis af evaluatorene! Dårlig evalueringsteori, er teori som ikke anvendes i praksis. Da FGE kun anvendes lidt i praksis, er det logisk en dårlig teori.

På det videnskabsteoretiske område er der ikke enighed om, hvad der kræves af en god teori. Der findes forskellige sandhedskriterier og forskellige opfattelser af teori. Derfor er det ikke kun evalueringsteori, der har problemer med at definere, hvad en god eller dårlig teori er.

Shadish et al. er det nærmeste vi kommer udviklingen af et begrebsapparat, hvormed det er muligt, at bedømme om en evalueringsteori er god eller dårlig. Begrebsapparatet er ikke udledt deduktivt, men induktivt og pragmatisk. Derfor hænger begrebsapparatet ikke logisk sammen. Der indgår mange forskellige teorier i apparatet. Der er tale om normative overvejelser over, hvad en evalueringsteori bør indeholde for at være god. Det er ikke en metateori, men en løs begrebsramme, som ikke prætenderer at give forklaringskraft. Kravene til en god evalueringsteori er store. Ingen evalueringsteoretiker har kunnet leve op til alle krav. Derfor gør det ikke FGE dårlig, at den kun lever op til nogle af kravene. I mangel af bedre er begrebsapparatet anset som velegnet til at vurdere FGE.

På vidensområdet har FGE store problemer i forhold til den konstruktivistiske ontologi. Guba og Lincoln (1989) præsenterer en uklar konstruktivisme, hvor

de blander kognitiv- og socialkonstruktivisme sammen. Det er uklart, hvor radikal deres konstruktivisme er. Der er en alvorlig uoverensstemmelse mellem det ontologiske, epistemologiske og metodologiske niveau. På det metodologiske niveau tager de udgangspunkt i en ikke radikal konstruktivisme, hvor metoderne ikke er meget forskellige fra traditionel positivistisk samfundsvidenskab. På vidensområdet er det sammenfattende tale om en dårlig evalueringsteori, som kan være årsag til, at FGE ikke anvendes mere i praksis. Evaluatorene som har anvendt FGE nævner, at det er vanskeligt at få interessenter til at acceptere ontologien bag FGE. Afhandlingen konkluderer ikke, at konstruktivisme ikke har nogen plads i evaluering, men blot at Guba og Lincoln (1989) teoretisk ikke har formået at gøre det konsekvent og overbevisende.

På det værdimæssige område opstår der også problemer med konstruktivismen. Analysen viser, at Guba og Lincoln (1989) flere steder i bogen havner i relativisme og solipsisme. Desuagtet at de ønsker empowerment af svage grupper og lighed mellem interessenter. De redegør ikke for nogen præsriptiv værditeori. I FGE er der en diskrepans mellem videns- og værdiområdet. Der er tale om dårlig evalueringsteori, som kan være årsag til, at FGE ikke anvendes mere i praksis. I eksemplerne fra praksis er der ingen, der nævner problemer med relativisme og manglende præsriptiv værditeori. Afhandlingen konkluderer ikke, at det ikke er godt, at inddrage interessenter og støtte svage grupper i en evaluering.

På baggrund af deres konstruktivisme mangler Guba og Lincoln (1989) teori om, hvordan sociale programmer virker, da programmer reduceres til konstruktioner. Det vurderes som en mangel, at de ikke forsøger, at forstå den kontekst sociale programmer virker i. De erkender ikke de begrænsninger, som følger af et pluralistisk repræsentativt demokrati, hvor folkevalgte politikere har magten til at bestemme. De undlader at diskutere vigtige forhold, der anskueliggøres i modellen, Den Parlamentariske Styringskæde. De overser forhold som illustreres af Principal-agent modellen. Paradoksalt mener Guba og Lincoln (1989), at netop deres generation af evalueringsteori er bedre tilpasset pluralistiske samfund. En praktiker nævner, at det er et problem at anvende FGE, når flertallet på det føderale niveau har opsat retningslinier, hvor andre typer af evalueringer ønskes.

De manglende overvejelser omkring den kontekst sociale programmer indgår i, betyder, at Guba og Lincoln (1989) får en naiv opfattelse af magt. Begrebet behandles stedmoderligt. De kunne have inddraget Habermass teori om Kommunikativ Handlen for på den måde at få teori bag deres Hermeneutiske Dialektiske Proces. De tror på en åben, lige og konsensusføgende forhandlingsproces mellem interessenter. Med baggrund i Foucault viser analysen, at kommunikation ikke er fri for magt. I forhold til manglende teori om sociale programmer vurderes FGE til at være en dårlig teori. Flere evaluatorene, som har afprøvet FGE, hæfter sig ved, at Den Hermeneutiske

Dialektiske Proces ikke er fri for magt. At interessenter er forskellige, har forskellige begreber og indsigt, og at nogle lyttes der mere til end andre. Der er endog interessenter, som er så svage, at de ikke ønsker at deltage i Den Hermeneutiske Dialektiske Proces. Den manglende teori om sociale programmer kan sandsynligvis forklare, hvorfor FGE ikke anvendes mere i praksis.

Guba og Lincoln (1989) mener, at evalueringresultater fra FGE i større udstrækning vil blive anvendt end resultater fra tidligere generationer af evaluering. De har en instrumentel opfattelse af anvendelse på kort sigt. Guba og Lincoln (1989) advokerer for formativ evaluering. Evalueringresultater skabes og anvendes i processen. Det er midlertidige konstruktioner i en given kontekst. Vi kan kritisk spørge, hvor lang holdbarhed resultaterne har, og hvor smal er konteksten? Kan resultaterne anvendes i morgen? Deres konstruktivisme betyder, at de må koble sig fra den trend i moderne samfundsvidenskab, som fokuserer på phronetisk viden. Det er netop et karakteristikon, at denne viden skabes af praktikere (interessenter), og kan anvendes i praksis. Det at Guba og Lincoln (1989) udelukker generalisering af evalueringresultater kan måske forklare, hvorfor FGE ikke anvendes mere i praksis. En enkelt praktiker nævner det som et problem i forhold til at vinde kontrakter hos en føderal myndighed. Som udgangspunkt vil en central opdragsgiver gerne have resultater, som kan anvendes ud over det sted, hvor evalueringen er foretaget.

Teori om evalueringpraksis er det vigtigste ifølge Shadish et al. Guba og Lincoln (1989) giver detaljerede anvisninger på det praktiske uden, at de kommer med konkrete eksempler. De fortæller, at de ikke har meget mere erfaring med FGE end læseren af deres bog. Indtrykket er, at deres teori kan anvendes på alle evaluander og i alle situationer. De opstiller ikke en kontingent teori. Litteratursøgning i internationale tidsskrifter og databaser viser, at der ikke er foretaget meget evaluering efter principperne i FGE. Gennemgående har praktikerne mødt de samme problemer med FGE. Det er problemer med at afgrænse og definere interessenter. At få dem til at forstå FGE og deltage aktivt i Den Hermeneutiske Dialektiske Proces. Det er problemer med manglende tid og ligefrem evalueringstræthed. Det er problemer med interessenternes forskellige magt, grad af interesse i evaluanden og evne til at formulere sig. Ingen praktikere har anvendt FGE stringent, efter de principper Guba og Lincoln (1989) har opstillet. Omvendt, ser stort set alle praktikere af FGE et fremtidigt potentiale i FGE. De praktiske problemer og den manglende empiri omkring FGE kan være en forklaring på, at FGE ikke anvendes mere i praksis.

Sammenfattende viser analysen med udgangspunkt i Shadish et al., at FGE ikke er en god evalueringsteori, men snarere en dårlig teori. Det kan forklare, hvorfor FGE ikke anvendes mere i praksis. Konklusionen udelukker ikke, at FGE kan udvikles og anvendes i fremtiden.

Litteratur

- Adelman, C. (1996): "*Anything Goes: Evaluation and Relativism*", *Evaluation* 2 (3), 1996, pp.291-305.
- Adolphsen, Jes; Qvist, Palle (1994): ABC i problemformulering, problemløsning og projektskrivning, Édition Edupax.
- Adolphsen, Jes (1995): Problemer i videnskab. En erkendelsesteoretisk begrundelse for problemorientering, Aalborg Universitetsforlag.
- Albæk, Erik; Rieper, Ole (2001): "*Evaluering i Danmark: Effektevaluering, monitorering og formativ evaluering*", pp.130-140 i Dahler-Larsen, Peter; Krogstrup, Hanne Kathrine (red.)(2001): Tendenser i evaluering, Syddansk Universitet.
- Andersen, Heine (1996): "*Jürgen Habermas*" pp.349-366 i Andersen, Heine; Kaspersen, Lars Bo (1996): Klassisk og moderne samfundsteori, Hans Reitzels Forlag.
- Andersen, Jørgen Goul (1998): Politik og samfund i forandring, bind 2., Columbus.
- Andersen, Niels Åkerstrøm (1999): Diskursive analysestrategier. Foucault, Koselleck, Laclau, Luhmann, Nyt fra Samfundsvidenskaberne.
- Andersen, Vilmer; Gamdrup, Peter (1990): "*Om problemformulering og projektarbejde*" pp.25-56 i Andersen, Heine (red.) (1990), Videnskabsteori og metodelære, bind 1, 3. udgave, Samfundslitteratur.
- Bateson, Gregory (2000): Steps to an ecology of mind, The University of Chicago Pres, Chicago and London.
- Berger, Peter L.; Kellner, Hansfried (1982): Nytolkning af Sociologien. Et essay om metode og engagement, Lindhardt og Ringhof.
- Beyme, Klaus von (1984): Die Politischen Theorien Der Gegenwart, R. Piper & CO Verlag, München, Zürich.
- Bickman, Leonard (2000): "*Summing Up Program Theory*", *New Directions for Evaluation*, No.87, fall 2000, pp.103-112.
- Bond, Gary R. (1991): "*Book Reviews. Fourth Generation Evaluation*", *Psychosocial Rehabilitation Journal*, 1991, Vol. 14, Issue 3.
- Buch, Roger et al. (2004): Rapportskrivning. Råd og vink til studerende, 2. oplag, Syddansk Universitetsforlag.
- Bøndergaard, Jette (2001): "*Habermas – om at ville noget i verden*" pp. 121-132 i Reinholm, Niels; Pedersen, Hans Skadkær (2001): Pædagogiske grundfortællinger, KvaN.
- Canadian Journal of Communication (1991): "*Fourth Generation Evaluation*", Vol.16, No.2.
- Carpenter, Mick (2004): "*Book Reviews. Fourth Generation Evaluation*", *Concepts and Transformation*, 9:3, 2004, pp 305-313, John Benjamins Publishing Company.
- Christensen, Jørgen Grønnegård et al. (2006): Politik og forvaltning, 2. udgave, Academica.

- Christensen, Søren; Jensen, Poul-Erik Daugaard (1986): Kontrol i det stille. Om magt og deltagelse, Samfundslitteratur.
- Christiansen, Peter Munk et al.(2006):Demokrati, magt og politik i Danmark, Gyldendal.
- Dahler-Larsen, Peter (2001A): Den Rituelle Refleksion – om evaluering i organisationer, Odense Universitetsforlag.
- Dahler-Larsen, Peter (2001B): *"From Programme Theory to Constructivism: On Tragic, Magic and Competing Programmes"*, Evaluation, Vol.7, No.3, 2001, pp.331-349.
- Dahler-Larsen, Peter (2002A): Evaluering kortlagt. En undersøgelse af amters og kommuners evalueringspraksis på det sociale område, Systime.
- Dahler-Larsen, Peter (2002B): At fremstille kvalitative data, Syddansk Universitetsforlag.
- Dahler-Larsen, Peter (2003): Selvevalueringens hvide sejl, Syddansk Universitet.
- Dahler-Larsen, Peter (2006A): Egne forelæsningsnotater, masterstuderende Syddansk Universitet.
- Dahler-Larsen, Peter (2006B): Evalueringsskulptur. Et begreb bliver til, Syddansk Universitetsforlag.
- Dahler-Larsen, Peter; Krogstrup, Hanne Kathrine (red.)(2001): Tendenser i evaluering, Syddansk Universitet.
- Dahler-Larsen, Peter; Krogstrup, Hanne Kathrine (2003): Nye Veje i Evaluering. Håndbog i tre evalueringsmodeller, Systime.
- Day, Patricia; Klein, Rudolf (1987): Accountabilities. Five Public Services, Tavistock Publications, London and New York.
- Donaldson, Stewart I.; Lipsey, Mark W. (2006): *"Roles for Theory in Contemporary Evaluation Practice: Developing Practical Knowledge"*, i Shaw, Ian F. et al.(2006): Handbook of Evaluation. Policies, Programs and Practices, SAGE Publications, London pp. 56-75.
- Dreyfus, Hubert; Dreyfus, Stuart (1991): Intuitiv Ekspertise. Den bristende drøm om tænkende maskiner, Nysyn, Munksgaard.
- Esping-Andersen, G.D. et al. (2002):Why we need a New Welfare state. Oxford University Press.
- Fink, Hans (2000): John Dewey, De store tænkere, Rosinante.
- Fishman, Daniel B. (1992): *"Postmodernism comes to program evaluation. A critical Review of Guba and Lincoln`s Fourth Generation Evaluation"*, Evaluation and Program Planning, Vol.15, pp.263-270. Pergamon Press Ltd.
- Flyvbjerg, Bent (2001): Making Social Science Matter. Why social inquiry fails and how it can succeed again, Cambridge University Press.
- Garde, Jens (1981): *"Saglige krav"*, pp.131-201 i Forvaltningsret. Almindelige emner, 1981, Institut for Offentlig Ret, Aarhus Universitet, Juristforbundets Forlag.
- Gergen K. J. (1991): *"The social constructionist movement in modern psychology"*. American Psychologist, 40, pp.266-275.
- Giddens, Anthony (1993): Sociology, Second Edition, Polity Press, UK.

- Greene, J. (1996): "*Qualitative Evaluation and Scientific Citizenship: Reflections and Refractions*", *Evaluation* 2 (3), pp. 277-289.
- Guba, Egon G.; Lincoln, Yvonna S. (1981): Effective evaluation. San Francisco, Jossey-Bass.
- Guba, Egon G.; Lincoln, Yvonna S. (1989): Fourth Generation Evaluation, SAGA Publications, London.
- Guba, Egon G. (ed.)(1990): The Paradigm Dialog, SAGE Publications, London.
- Guba, Egon G. (1992): "*Relativism*", *Curriculum Inquiry*, 22, pp.17-23.
- Guba, Egon G.; Lincoln, Yvonna S. (1994): "*Competing Paradigms in Qualitative Research*", pp. 105-117 i Denzin, Norman K.; Lincoln, Yvonna S. (1994): Handbook of Qualitative Methods, Second Edition, SAGE Publication.
- Ham, Christopher; Hill, Michael (1993): The Policy Process in the Modern Capitalist State, Second Edition, Harvester Wheatsheaf, New York.
- Hansen, Hanne Foss (2000): "*Evaluering i Danmark*", *Nordisk Administrativt Tidsskrift*, nr. 2, 2000, 81. årgang, pp.156-177.
- Hansen, Hanne Foss (2003): Evaluering i staten. Kontrol, læring eller forandring, Samfundslitteratur.
- Hansen, Hanne Foss (2005): "*Kvalitet i evalueringspraksis*", *Nordisk Administrativt Tidsskrift*, Nr. 1/2005, 86. årgang, pp.27-43.
- Haralambos, Michael (1980): Sociology. Themes and Perspectives, University Tutorial Press, Great Britain.
- Hartnack, Justus; Sløk, Johannes (red.)(2000) De store tænkere. Kant, Rosinante.
- Heap, James L. (1995): "*Constructionism in the Rhetoric and Practice of Fourth Generation Evaluation*", *Evaluation and Program Planning*, Vol.18, No.1, pp. 51-61.
- Held, David (1993): Models of Democracy, Polity Press.
- Hellevik, Ottar (2003): Forskningsmetode i sociologi og statsvitenskap, Universitetsforlaget, Oslo.
- Hoff, Jens (1993): "*Medborgerskab, brugerrolle og magt*", pp.75-106 i Andersen, Johannes et al.(1993): Medborgerskab – Demokrati og politisk deltagelse, Systime.
- Huebner, Angela J.; Betts, Sherry C. (1999): "*Examining Fourth Generation Evaluation. Application to Positive Youth Development*", *Evaluation*, Vol.5. pp.340-358, SAGE Publication.
- Illeris, Knud (1999): Læring – aktuel læringsteori i spændingsfeltet mellem Piaget, Freud og Marx, Roskilde Universitetsforlag.
- Jacobsen, Bo et al.(1999): Videnskabsteori, Gyldendal.
- Jensen, Claus Haagen (1981): "*Forvaltningsrettens kilder*", pp.89-130 i Forvaltningsret. Almindelige emner, 1981, Institut for Offentlig Ret, Aarhus Universitet, Juristforbundets Forlag.
- Kelstrup, Morten (1995): "*Om det danske demokrati og den europæiske integration*", Copenhagen Research Project on European Integration, Arbejdspapirer 3/1995, Institut for statskundskab, København.

- Knudsen, Christian (1990): "*Empirisk-analytisk videnskabsteori. Del II: Fra naiv til sofistikeret falsifikationisme*", pp.99-118 i Andersen, Heine (red.)(1990), *Videnskabsteori og metodelære*, bind 1, 3. udgave, Samfundslitteratur.
- Koch, Tina (1994): "*Beyond measurement: fourth-generation evaluation in nursing*", *Journal of Advanced Nursing*, 1994, 20, pp.1148-1155.
- Koch, Tina (2000): "*Having a say: negotiation in fourth-generation evaluation*", *Journal of Advanced Nursing* 2000, 31. (1) pp.117-125.
- Krogstrup, Hanne Kathrine (2003): *Evalueringsmodeller. Evaluering på det sociale område*, Systime. Den findes også på nettet: <http://www.social.dk/netpublikationer/2003/pdf/evalueringm.pdf>
- Kushner, S. (1996): "*The Limits of Constructivism in Evaluation*", *Evaluation* 2(2), pp.189-200.
- Kvale, Steinar (1997) *InterView. En introduktion til det kvalitative forskningsinterview*, Hans Reitzels Forlag.
- Labinowicz, Ed (1980): *The Piaget Primer. Thinking, Learning, Teaching*, Addison-Wesley Publishing Company.
- Lai, Morris K.(1991): "*Field-based Concerns About Fourth-Generation Evaluation Theory*", Paper presented at the annual meeting of the American Educational Research Association, Chicago, EDRS, <http://www.eric.ed.gov/>
- Larsen, Peter Gorm (2006): "*Evaluering af leve-bomiljø på Pleje- og aktivitetscenter Højvangen, afsnit Lynghøjen, Tjele Kommune*". Findes på adressen www.policy.dk
- Larsen, Peter Gorm (2007A): "*Anbefalinger i et evalueringsteoretisk perspektiv med eksempler fra Danmarks Evalueringstinstituts praksis*", årsopgave i evalueringsteori, masterstudie i evaluering, Syddansk Universitet. Findes på adressen: <http://www.policy.dk/>
- Larsen, Peter Gorm (2007B): "*Kulturkamp på plejehjemmet*", Kronik i Nyhedsmagasinet Danske Kommuner, nr.33. 2007. Findes på adressen: <http://www.danskekommuner.dk/default.asp?id=86407>
- Laughlin, R.; Broadbent, J. (1996): "*Redesigning Fourth Generation Evaluation: An Evaluation Model for Public-Sector Reforms in the UK?*" *Evaluation* 2 (4), pp.431-451.
- Lauritzen, Laurits (1972): "*Anarkismen og syndikalismen*" pp.231-265, i *Politiske ideologier. Fra Platon til Mao*, Politikens Forlag 1972.
- Lay, Margaret; Papadopoulos, Irena (2007): "*An Exploration of Fourth Generation Evaluation in Practice*", pp.495-504, *Evaluation*, 2007, Vol. 13 No. 4.
- Lincoln, Yvonna S.;Guba, Egon G. (1985): *Naturalistic inquiry*, Beverly Hills, CA:SAGE.
- Lincoln, Y.S. (1990): "*The making of constructivist: A remembrance of transformations past*", in Guba, E. (Ed.), *The paradigm dialog* (pp67-87), Newbury Park, SAGE Publications.
- Lindgren, Sven-Åke (1996): "*Michel Foucault*" pp. 314-329, i Andersen, Heine; Kaspersen, Lars Bo (1996): *Klassisk og moderne samfundsteori*, Hans Reitzels Forlag.

- Luhmann, Niklas (2000): Sociale systemer. Grundrids til en almen teori, Hans Reitzels Forlag.
- Lübcke, Poul (red.)(1982A): Vor tids filosofi. Videnskab og sprog, Politikens Forlag.
- Lübcke, Poul (red.)(1982B): Vor tids filosofi. Engagement og forståelse, Politikens Forlag.
- Lübcke, Poul (red.)(1983): Politikens filosofi leksikon, Gyldendals Bogklubber.
- Mark, Melvin M. et al. (2006): *"Introduction. The Evaluation of Policies, Programs And Practices"* i Shaw, Ian F. et al. (2006): Handbook of Evaluation. Policies, Programs and Practices, SAGE Publications, London pp. 1-30.
- Moll, Luis C. (ed.): Vygotsky and education. Instructional implications and applications of sociohistorical psychology, Cambridge University Press.
- Mortensen, Nils (1996): *"Amerikansk pragmatisme"* pp. 120-132 i Andersen, Heine; Kaspersen, Lars Bo (1996): Klassisk og moderne samfundsteori, Hans Reitzels Forlag.
- Næss, Arne (1980): Vitenskapsfilosofi, 3. udgave, Universitetsforlaget, Oslo.
- Olsen, Henning (2002): Kvalitative kvaler, Akademisk Forlag.
- O`Neil, Tom (1995): *"Implementation Frailties of Guba and Lincoln´s Fourth Generation Evaluation"*, Studies in Educational Evaluation, Vol.21, pp.5-21.
- Patton, Michael, Quinn (1990): *"The Evaluator`s Responsibility for Utilization"* i Alkin, Marvin C. (1990): Debates on Evaluation, SAGE Publications pp. 185-202.
- Patton, Michael Quinn (1997): Utilization-Focused Evaluation: The New Century Text, SAGE Publication.
- Patton, Michael Quinn (2002): Qualitative Research & Evaluation methods, SAGE.
- Pawson, Ray (1996): *"Three Steeps to Constructivist Heaven"*, Evaluation, 1996, Vol.2 No.2., pp.213-219, SAGE Publications, London.
- Pawson, Ray & Tilley, Nick (1997): Realistic Evaluation, London, SAGA Publications.
- Piaget, Jean (2000): *"Ligevægtsbegrebets rolle i psykologien"*, pp.26-36 i Illeris, Knud (red.)(2000): Tekster om læring, Roskilde Universitetsforlag.
- Qvortrup, Lars (1998): Det hyperkomplekse samfund. 14 fortællinger om informationssamfundet, Gyldendal.
- Qvortrup, Lars (2004): Det vidende samfund. Mysteriet om viden, læring og dannelse, Unge Pædagoger.
- Rasmussen, Erik (1981): Lighedsbegreber, Berlingske Leksikon Bibliotek. Berlingske Forlag, København.
- Rasmussen, Jens (1997): Socialisering og læring i der refleksive moderne, Unge Pædagoger.
- Rasmussen, Jens (2004): Undervisning i det refleksivt moderne. Politik, profession, pædagogik, Hans Reitzels Forlag.
- Rieper, Olaf (2004): Håndbog i evaluering. Metoder til at dokumentere og vurdere proces og effekt af offentlige indsatser, AKF Forlaget.

- Rossi, Peter H.; Lipsey, Mark W.; Freeman, Howard E. (2003): Evaluation. A Systematic Approach, Seventh Edition, SAGE Publication, London.
- Schön, Donald A.(2001): Den reflekterende praktiker. Hvordan professionelle tænker, når de arbejder, Klim.
- Schutz, Alfred (2005): Alfred Schutz. Hverdagslivets sociologi, Hans Reitzels Forlag.
- Schwandt, Thomas A. (2001): Dictionary of Qualitative Inquiry, Second Edition, SAGE Publications, London.
- Schwandt, Thomas A. (2005): "The centrality of practice to evaluation" i American Journal of Evaluation Vol.26, no.1 pp. 95-105.
- Schwandt, Thomas A.; Burgon, Holli (2006): "Evaluation and the study of lived experience" i Shaw, Ian F. et al. (2006): Handbook of Evaluation. Policies, Programs and Practices, SAGE Publications, London pp.98-117.
- Scriven, Michael (1991A): Evaluation Thesaurus, Fourth Edition, SAGE.
- Scriven, Michael (1991B): "Beyond Formative and Summative Evaluation" i Milbrey McLaughlin; D.C. Philips (3.ds.) Evaluation and Education: At Quarter Century, Chicago, University of Chicago Press.
- Shadish, R. William et al.(1991): Foundations of Program Evaluation. Theories of Practise, SAGE Publications, London.
- Shadish, R. William (1998): "Evaluation Theory is Who We Are", American Journal of Evaluation, Winter 1998, vol.19, Issue 1, pp.1-19.
- Shaw, Ian F. et al.(2006): Handbook of Evaluation. Policies, Programs and Practices, SAGE Publications, London.
- Stake, Robert E. (2004): Responsive Evaluation, Thousand Oaks, SAGE.
- Statskundskab (2004): Leksikon i Statskundskab, Akademisk Forlag.
- Sørensen, Marianne (2005): "Politisk ansvar i de danske kommuner", Speciale ved Institut for Statskundskab, Aarhus Universitet, pp.3-30, findes i materialesamling 1 til Politologi og offentlig administration, MIE-uddannelsen, efterår 2006.
- Sørensen, Max (1973): Statsforfatningsret, 2. udgave ved Peter Germer, Jurist- og Økonomforbundets forlag.
- Theilgaard, Jesper (2006): Det danske vejr, Gyldendal.
- Thomsen, Jens Peter Frølund (2000): Magt og indflydelse, Magtudredningen, Institut for statskundskab, Aarhus Universitet.
- Vedung, Evert (1998): Utvärdering i politik och förvaltning, Andra upplagan, Studentlitteratur.
- Weiss, Carol H. (1977): "Research for Policy 's Sake:The Enlightenment Function of Social Research", Policy Analysis, Vol.3, pp.531-545.
- Weiss, Carol (1990A): "If Program Decisions Hinged Only on Information. A Response to Patton" i Alkin, Marvin C. (1990): Debates on Evaluation, SAGE Publications, pp.208-222.
- Weiss, Carol (1990B): "Evaluation for decisions: Is anybody there? Does anybody care?" i Alkin, Marvin C. (1990): Debates on Evaluation, SAGE Publications, pp.171-184.
- Wenneberg, Søren Barlebo (2002): Socialkonstruktivisme – Positioner, problemer og perspektiver, Samfundslitteratur.

Winter, Søren (1994): Implementering og effektivitet, Systime Academic.
Østerud, Øyvind (1995): Statsvitenskap. Innføring i politisk analyse,
Universitetsforlaget, Oslo.