

# Neljännän sukupolven evaluaation jälkeinen arviointitutkimus?

Petri Virtanen

## ABSTRACT

The aim of article is to focus on three themes in current debate on evaluation research. First, the nature of evaluation "school of thoughts" that have been in existence since 1980s and various evaluation standards are discussed. Second, the article highlights new competence demands for evaluators at the beginning of the new century, and finally this article takes a critical look of the shift that is taking place in evaluations carried in public sector domain today. It is argued, that there has not been but quasi schools of thoughts of evaluation in the domain of public sector during the last decades. The only substantial mode of thought, on which all the quasi schools of thought has been based on, has related to causal reasoning and causality-based impact evaluations. Also the existing evaluation standards have been utilised without reasonable criticism. Also it is emphasized, that there are totally new competence demands for the modern evaluators. Old-fashioned methods and ideas do not necessarily meet the knowledge demands of the modern public sector organisations and policy institutions. As a whole, the nature of evaluation has shifted towards consultation and knowledge creation, interpretation and management.

*Key words: Schools of thought in evaluation, evaluation standards, causal reasoning, evaluation as a form of knowledge management*

## MAAILMA, EUROOPPA JA SUOMI ARVIOIVAT

Arviointitutkimus on nykyisin sekä symbolisessa että toiminnallisessa mielessä kiinteä osa julkishallinnon arkipäivää kaikkialla maailmassa. Julkishallinnon arvioinnin toimintamallista on viimeistään 2000-luvun alussa tullut globaali ilmiö (esim. Furubo & Sandahl 2002; Wollmann 2003a; Pollitt 2003) ja se on levinnyt "perinteisestä" käytötarkoituksestaan - hallinnon reformien arvon todentamisesta - kaikenlaiseen hallinnolliseen toimintaan<sup>1</sup>.

Suomen kehitys ei poikkea muun maailman kehityksestä. Täälläkin arvioidaan kaikkialla julkishallinnossa. Silti käsitteellisessä mielessä arviointitutkimuksen merkitys on kaikkea muuta kuin selvä. Arviointi, seuranta, tarkastus ja valvonta ovat edelleen toisiinsa limittyviä ja rajoiltaan epäselviä käsitteitä ja toimintoja. Mainituista toiminnallisista päällekkäisyyksistä huolimatta arviointitutkimuksen rooli tulee edelleen muuttumaan. Kysymys on siitä, miten ja millä reunaehdoilla arviointitutkimuksesta on tulossa hallinnon tietotuotannon osa ja politiikan (ensisijaisesti policies, ei politics) tekemisen mahdollistaja.

Arviointitutkimuksen kentällä toimivat lienevät kuulleet Egon Guban ja Yvonna Lincolnin teoksesta Fourth Generation Evaluation, joka julkaistiin vuonna 1989 ja sillä oli ilmestyessään ja osittain vielä edelleenkin on läpiluoataava merkitys arviointitutkimuksen kentällä. Teoksesta puhuttiin aikanaan paljon, sen tiimoilta järjestettiin seminaareja, sitä ylistettiin - ja vastustettiin. Teoksen merkityksestä kertoo se, että se oli aikanaan jonkun aikaa kustantajansa (Sage Publications) myydyin tiedekirja. 2000-luvun ensimmäisten vuosien näkökulmasta arvioiden Guban ja Lincolnin

nin teosta on kuitenkin enemmän tai vähemmän kohdannut Karl Marxin kohtalo: heistä edelleen puhutaan paljon, mutta sitäkin harvemmin ovat lukeneet heidän tekstejään.

Vaikka Guban ja Lincolnin teos todellakin herätti huomiota julkaisuhetkellään ja vielä kauan sen jälkeenkin, ei teoksen asiasisältö ollut mitenkään uutta, innovatiivista, vallankumouksellisuudesta puhumattakaan. Pelkistäen kyse oli siitä, että teoksessa esitettiin synteesi 1950-luvun jälkipuoliskolta virinneestä ja erityisesti 1970- ja 1980-luvulla voimissaan olleesta arviointitutkimuksen metodologisista kiistasta, jossa vastapuolina erityisesti Yhdysvalloissa olivat naturalistisen, konstruktivistisen tutkimusmetodologian kannattajat ja positivistisen, määrällistä informaatiota ainoana oikeana tiedon muotona pitävät tutkijat. Guba ja Lincoln pitivät tavoiteisiin sidottua ohjelma-arvioinnin lähestymistapaa paitsi managerialistisena, pahimmassa tapauksessa valtaa pitävien vallan pönkittämisinä, mutta myös ontologisesti ja epistemologisesti mahdottomana. Nimenomaan tämä kritiikki on heidän teoksensa keskipisteessä ja kiteytyi heidän teoksessaan voimakkuudella, joka hakee vertaistaan hyökkäyksen jyrkkyydessä suhteessa siihen asti vallalla olleeseen positivistishenkiseen ohjelma-arvioinnin lähestymistapaan.

Guban ja Lincolnin hyökkäys oli siinäkin mielessä kohtuuton, että on vaikea löytää vakavasti otettavia tutkijoita, jotka pitävät määrällistä aineistoa ainoana oikeana aineistotyyppinä. Vaikka Guba ja Lincoln eivät olleet kovinkaan omaperäisiä sisällöllisessä mielessä, osasivat he antaa amerikkalaiseen tapaan omalle tulkinnalleen houkuttelevaa vetoavuusarvoa sisältävän otsikon ja leiman ja loivat tulkinnastaan vuosia alan toimijoiden ja harrastajien piirissä eläneen konseptin. Juuri tämän brändin muodostamistaidon osalta Guba ja Lincoln ovat uniikkeja ja muistuttavat siinä mielessä paljolti pelkästään itseään, koska heitä ennen tämäntyyppinen tuotteistaminen oli ollut, ainakin suurelta osin, tuntematonta tai ainakin peiteltyä.

Merkillepantavaa siis on, että Guba ja Lincoln loivat magnum opuksellaan sittemmin 1990-luvulla ja erityisesti sen jälkipuoliskolla vakiintuneen perinteen: arviointitutkimuksen kentällä monet itsensä koulukunnan arvoiseksi tuntevat tutkijat julkaisivat mitä erilaisimmista lähtökohdista ja motiiveista liikkeelle lähtien monentasoisia evaluation-loppuisia teoksia. Esi-

merkinä voidaan mainita vaikkapa Huey-Tsyh Cheniin henkilöitynyt teoriaperusteinen koulukunta, David M. Fettermaniin henkilöitynyt arvioinnin kohteiden valtaistamiseen tähtäävä koulukunta, erityisesti Ray Pawsoniin (ja osittain Nick Tilley'hyn) henkilöitynyt realistisen arvioinnin koulukunta (ks. esim. Chen 2005; Fetterman ym. 1996; Fetterman 2001 ja 2004; Pawson & Tilley 1997; Pawson 2005a ja 2005b), ja monet muut<sup>2</sup>.

On ollut hämmentävää huomata, ettei tätä kirjallisuuden lajityyppiä ole juurikaan kritisoitu näihin päiviin asti, ei kansainvälisillä areenoilla eikä etenäkään Suomessa. Kovinkaan paljoa ei ensinnäkään ole pohdittu sitä, missä määrin tämä lajityyppi ja sen liepeiltä löytyvä monopolivinen kommentaarikirjallisuus on jäänyt epämääräiseksi sekoitukseksi tieteellistä kausaalipäättelyä, tutkimustyön metodologiaa, arvioinnin kohteiden valtaistamista, yhteiskunnallisen hyvän edistämistä ja 1950- ja 1960-luvulta alkaen vaikuttaneiden ja arviointitutkimuksen kentälle positoituvien yhdysvaltalaisen yhteiskuntatieteilijöiden, kasvatustieteilijöiden ja hallintotieteellisesti suuntautuneiden organisaatiotutkijoiden (erityisesti Cronbach, Guba, Lincoln, Patton, Scriven, Stake, Stufflebeam, Weiss, ja monet muut) ajatusten toistamista ja suoranaista kopiointia. Toinen hämmentävä piirre on ollut se, missä määrin tähän lajityyppiin on sisältynyt eritasoista hurmahenkistä hengennostatusta ja "1990-luvun arviointigurujen" omaa oletettua suuruutta.

Kaikkein hämmentävintä on ollut todeta, että Suomessa tähän keskusteluun ei ole osallistuttu käytännössä lainkaan. Täällä on lähinnä tyydytty kopioimaan kansainvälisiltä areenoilta välittyneitä lajityypin konsepteja alisosiaalistettuina malleina. Hyvänä esimerkkinä tästä käy edellä mainitut valtaistamiseen tähtäävä fettermanilainen ajattelutapa 1990-luvun lopulla ja vuosituhaten vaihteen jälkeen itsensä pienellä ja vakiintumattomalla suomalaisella arviointitutkimuskentällä läpi lyönyt realistinen arviointitapa, joita on sovellettu täällä sellaisella hurmahenkisyydellä, joka todennäköisesti häkellyttäisi itse lähestymistavan alkuperäiskehittäjätkin.

Tämän artikkelin tavoitteena ei ole Guban ja Lincolnin (1989) teoksen keskeisten ideoiden esittely eikä 1990-luvun arviointitutkimuksen koulukuntien metodologisten ohjelmien kriittinen analyysi. Tätä teemaa on tarkoitus jatkaa edellä sanotun mukaisesti toisaalla (Ahonen & Virtanen

2005). Sen sijaan tämän artikkelin tarkoituksena on pohtia arviointitutkimuksen olemusta kolmen jäljempänä mainitun teeman näkökulmasta. Mainitut teemat kytkeytyvät monin tavoin toisiinsa ja tässä on tarkoituksenmukaista ryhmitellä ne kolmeen luokkaan, joka toimii tässä myös artikkelin päälukuryhmittelyn pohjana: (i) arviointitutkimuksen koulukuntien luonne ja siihen liittyvä metodologinen "paluu", (ii) arviointitutkimuksen lähestymistapojen soveltamisen mahdollisuudet ja sen kytkennät arviointitutkimuksen tekijältä vaadittaviin osaamisvaatimuksiin, ja (iii) arviointitutkimuksen merkityksen painopistesiihtymän julkishallinnon kehittämistoiminnassa. Tiivistäen tässä artikkelissa väitetään, että

- Arviointitutkimuksen kentän metodologinen lohkoutuminen 1990-luvulla ei ole merkinnyt aitojen, itsenäisten ja metodologisesti kiinnostavien koulukuntien syntymistä, vaan kyse on ollut eräänlaisesta bourdieulaisesta erottautumisen politiikasta, tavallaan brändien muodostamisesta omine vetoavuusarvoineen.
- Arviointitutkimuksen kentän portinvartijusmekanismien tunnistaminen on edellytys esimerkiksi 1990-luvulla laajalle levinneiden arviointistandardien luonteen ymmärtämiseksi.
- Arviointitutkimuksen tekijöiltä vaaditaan 2000-luvun alussa aivan uudentyypisiä osaamisvalmiuksia, mikä liittyy ennen muuta arviointitutkimuksen luonteen muutokseen.
- Arviointitutkimuksen merkitys julkishallinnon politiikan suunnittelussa ja toimeenpanossa radikaalisti muuttunut viimeksi kuluneen viidentoista vuoden aikana. Arviointitutkimuksen metodologiset kiistat ovat siirtyneet taka-alalle ja arvioinnista on tullut kokonaisvaltaista julkisen politiikan tietotuotantoa.

#### ARVIINTITUTKIMUKSEN LOHKOUTUNUT KENTTÄ JA METODOLOGINEN "PALUU"

##### *Koulukuntien muodostumisesta*

Barbara Czarniawska (2003) julkaisi hiljattain puhuttelevan artikkelin tieteenalojen määrittelyn vaikeudesta. Nämä ironiset havainnot sopivat hyvin myös arviointitutkimuksen kenttään: se ei ole yhdenmukainen toiminnan alue tai toi-

mintatapa, selkeydestä puhumattakaan. Arviointitutkimukseen yleisesti liitettävien käsitteiden, lähestymistapojen, aineistonkeruustrategioiden ja sovellettavien tutkimusmenetelmien kirjo ei helpota tämän tutkimustyyppin sisällön selkeätä hahmottamista.

Käytännön kokemus arviointityötä tehneelle kuitenkin on, että on jotain mitä arviointitutkimus on ja jotain mitä se ei ole. Ehkä olisi syytä pysähtyä hetkeksi miettimään sitä, millainen arviointitutkimuksen kenttä loppujen lopuksi on, siis metodologisessa mielessä, ja miksi kentän ylipäänsä pitäisi olla metodologisessa mielessä lokeroitunut?

Lähtökohtaisesti voidaan ajatella, että arviointitutkimuksen kentän metodologinen ja lähestymistavallinen lokerointi olisi haitallista lähinnä siksi, että huomion kiinnittäminen tutkimusmetodologian kokonaisuuteen tai arvioitavan kohteen kokonaisuuteen häipyä jollain tavalla taka-alalle tai tulee ainakin monessa suhteessa toisarvoiseksi. Tämä ilmiö on universaali ja näkyy hyvin tieteessä yleisemminkin. Tiede pirstoutuu yhä kapeampiin erikoisaloihin. Maailmasta on tältä osin tullut tieteentekijöiden koulutuksen osalta eräänlainen millimetriosuamisen maailma; tarkastelun kohteena olevan millimetrin sisältä löytyy aina jokin vielä pienempi tarkasteluyksikkö, jolle erikoistua ja jota tutkia mitä erilaisimmin metodein. Tuntuu siltä, ettei pyrkimys yleiskatsauksellisuuteen, raja-aitojen ylittämiseen, yhteyksien löytämiseen kaiken käsillä olevan erilaisuuden keskellä ole jäänyt kenenkään tehtäväksi. Arviointiteoria ja arviointitutkimusta koskeva teoria on tavallaan kulkenut jo jonkin aikaa eri raiteilla käytännön arviointityöstä, millä ei kuitenkaan ole ollut suuren suurta merkitystä käytännön arviointityössä.

Arviointitutkimuksen kenttä on tutkimusmetodologisessa mielessä ollut 1970-luvulta alkaen aina vuosituhatlukuun loppuun asti hajautuva kenttä. Guba ja Lincoln kytkevät arviointitutkimuksen synnyin tieteen kehitykseen ja puhuvat 1970- ja 1980-lukujen osalta jo neljännen sukupolven evaluaatiosta. Ottamatta tässä kantaa tällaisten modernissa historian tutkimuksessa todettuun kronologisten periodisointien vaikeuteen voidaan heihin viitata todeta, että arviointitutkimus on kehittynyt nykyiseen muotoonsa yhteiskunta-, talous- ja kasvatustieteen piiristä liikkeelle lähteneenä soveltavana tutkimuksen alueena, joka sittemmin on haarautunut yhä nyansoidum-

pien arviointitutkimuksen osakenttien logiikan mukaisesti ja uusien esiin nousseiden poliittis-hallinnollisten kysyntätekijöiden perusteella (tutkimusaiheiden emergenssin mukaisesti tapahtunut kehitys).

Ilkka Niiniluodon (2003) käyttämää terminologiaa mukaillen arviointitutkimuksesta on tullut tieteenala taidollisen tieteistymisen myötä, siis eräänlainen suunnittelutiede esimerkiksi lääkintä-, insinööri-, maatalous- ja liiketaloustieteiden rinnalle ja tästä on ollut seurauksena erinäiset uskottavuus- ja muut pulmallisuudet. Vanhat akateemiset oppialat yliopistoissa varmasti katsovat näitä suunnittelutiedetuloikkaita ja arviointitutkimustakin joukon jatkeena syvästi epäillen, sillä niiden on uskottu perustuvan pelkkiin opportunistisiin ja keinotekoisin ajatusrakennelmiin ilman omia aitoja teorioita ja metodeita. Oireellista on, ettei Suomessa ole yhtään, päinvastoin kuin monissa muissa maissa, arviointitutkimukseen ensisijaisesti keskittyvää professuuria, joskin se ja sen liitännäisteemat kuuluvat läheisesti moneen professuurin opetusvelvollisuuksiin ja tieteenalamäärityksiin.

#### *Standardien esiinmarssi*

Arviointitutkimuksen kentän lohkoutuminen on liittynyt läheisesti arviointitoiminnan standardien kehittämiseen. Yhdysvalloissa laadittiin ensimmäiset arviointitoiminnan standardit 1970-luvulla standardien laatimisaallon levitessä kaikkialle maailmaan viimeistään 1990-luvun kuluessa. 2000-luvun alussa ollaan tilanteessa, jossa arviointitoiminnan standardeja on olemassa kaikissa maailmankolkissa, useimmiten vuonna 1994 vahvistettujen American Evaluation Associationin ohjelma-arvioinnin standardien kopioina tai muunnelmina tosin. Esimerkiksi syyskuussa 2004 Berliinissä pidetyn European Evaluation Society'n konferenssissa keskusteltiin paljon erilaisten standardien rakenteesta ja soveltamisen mielekkyydestä. Nytemmin, 2000-luvun alussa, on kansainvälisillä foorumeilla alkamassa kriittinen keskustelu arviointitoiminnan eettisten periaatteiden soveltamisesta ja ennen muuta niiden soveltamisen eettisen partikularismin ongelmista (tästä keskustelusta esim. Virtanen 2004; Virtanen & Laitinen 2004).

Olemassa olevia arviointistandardeja kriittisesti tarkasteltaessa voidaan semanttisista ja syn-

taktisista muotoiluista helpolla päätellä niiden tekemiseen liittyneiden henkilöiden kädenjälki. Standardeihin liittyy paljon symbolisia ja taloudellisia intressejä. Arvioinnin standardien ja laatusuosituksen kehittäjillä on mahdollisuus säännellä sitä, millaisia toimijoita arviointitutkimuksen kentälle pääsee ja millaisten oppien mukaan arviointitutkimusta kulloinkin pitäisi tehdä. Käytännössä tämä tapahtuu siten, että joissakin tapauksissa arviointien tilaajat edellyttävät joidenkin tiettyjen arviointistandardien noudattamista, joillakin arviointiyhdistyksillä on oikeus pitää arviointistandardien soveltamiseen tähtäävää koulutusta, ja niin edelleen.

Tietenkin voidaan poleemisesti kysyä, miksi arviointistandardit ylipäänsä ovat olemassa? Kysymyksen ääneen esittäminen ei tarkoita arviointityöhön liittyvien eettisten periaatteiden vähätelyä, koska arviointityön voi todellakin liittyä monenlaista vastuuttomuutta. Se voi esimerkiksi liittyä menetelmien soveltamiseen, johtopäätösten konstruointiin tai kehittämissuosituksen manipulointiin. Patologiseen valehteluun ei käytännössä törmää arviointitutkimuksia lukiessa, mutta periaatteessa sekin lienee mahdollista. Kokonaan toinen kysymys on, esiintyykö arviointitoiminnan keskuudessa muistiaukkojen korjaamista bona fide - siis eräänlaista konfabulointia - sellaisissa tapauksissa, joissa asioista tai ilmiöistä ei tiedetä tarpeeksi, tai lainkaan.

Arviointistandardeja koskeva kritiikki on kohdistunut niiden metodologiseen riippuvuuteen tietyistä koulukunnista ja siihen, ettei ehdottomien käskyjen perusteella voi kattaa monimuotoisia päätöksentekotilanteita. Ehdottomat käskyt ovat lähtökohtaisesti jos eivät mielettömiä niin ainakin hyvin vaikeasti noudatettavissa, koska niillä ei voida kattaa kaikkea tässä tapauksessa tarkastelun kohteena olevan arviointiprosessin eri vaiheita. Edellä sanotun perusteella ehdottomat käskyt, kuten (1) Tee X tai ehdolliset käskyt, kuten (2) Tilanteessa B, tee X, eivät perustu siihen, että ne ottaisivat huomioon erilaisten ilmiöiden syy- ja vaikutussuhteita ilmaisevia säännönmukaisuuksia. Tämän vuoksi tarvitaan tehtävätietoon liittyviä lauseita, kuten (3) X:n tekeminen tilanteessa B saa aikaan vaikutuksen A, joskin nämäkään eivät pysty vastaamaan edellä mainittuun arviointiprosessin partikularismin haasteeseen kuin rajallisesti. Edellä mainituista voidaan vielä johtaa ehdollisia toimintasuosituksia, jotka voivat olla teknisen normin muotoon kirjoitettuja lauseita,

kuten (4) Jos haluat A:n ja uskot olevasi tilanteessa B, sinun pitää tehdä X. Tekniseen standardinormiin pyrkiminen on siis hyvin haastavaa ja 2000-luvun alun tilanteen valossa voidaan sanoa, että suurilta osin yritykset tähän tavoitteeseen pääsemiseksi ovat parhaimmillaankin jättäneet hyvän yrityksen asteelle.

#### *Metodologinen "paluu"*

Arviointitutkimuksen kentän 1980- ja 1990-luvuilla tapahtuneen lohkoutumisen taustalla on kuitenkin koko ajan ollut jotain hyvin pysyvää metodologisessa mielessä kaikkien näiden vuosien aikana. Oikeastaan on olemassa selkeä polku, joka on kantanut arviointitutkimuksen sekalaisessa ja vaikeasti hahmottuvassa kentässä ja sen kautta käytännön arviointityön kentillä, koostuen vaikuttavuusarvioinnin erilaisista muunnelmista. Sen juuret juontavat taas vuosisatojen päähän, humelaiseen kausaalipäätelyyn.

David Humen lähtökohta on hyvin käytännöllinen ja arviointitutkimuksen näkökulmasta vastaväitteitä herättämätön. Syyn ja vaikutuksen välistä suhdetta ei nimittäin voi todentaa apriorisesti, vaan nimenomaan kokemusperäisesti. Ilman kokemusperäistä tulkintaa, havainnointia ja mittausta emme voi hänen mukaansa tehdä päätelmiä kausaalisesta suhteesta syyn ja vaikutuksen välillä. Edellä sanottu herättää myös kysymyksiä arviointitutkimuksen näkökulmasta asiaa tarkasteltuna. Kysymys voidaan nimittäin asettaa myös arviointikriteerien näkökulmasta: minkä tiedon varassa voimme päätellä - tarkemmin sanottuna, tehdä evaluatiivisen tulkinnan - joidenkin asioiden syy- ja vaikutusyhteydestä? Onko niin, että arvioinnin tekijöiltä vaaditaan kokemusperäistä tietoa arvioinnin kohteena olevasta toiminnasta, ikään kuin preskriptiivisessä mielessä? Käytännön kokemus puoltaisi tällaista ehtoa, koska muutenhan voitaisiin ajatella, että arviointi on mekaanista tutkimusmenetelmien soveltamista, ortodoksisen eli oikeaoppiseksi hyväksytyyn arviointitiedon tuottamista käytettäessä Pierre Bourdieun tunnetuksi tekemää ilmaisua.

Humen ajattelutapa syyn ja vaikutuksen välisestä suhteesta ei ollut mitenkään ongelmaton ja hän itse oli hyvin tietoinen ajattelutapansa rajallisuudesta. Syyn ja vaikutuksen välinen suhde nimittäin on hyvin harvoin suoraviivainen, tai yli-

päänsä helposti ymmärrettävissä oleva. Maailma on kompleksinen, samoin sen sisältämät lainalaisuudet ja säännönmukaisuudet, eikä nämä tee helpoksi uteliaalle mielelle mainittujen syy- ja vaikutussuhteiden paikantamisen, tulkinnan ja selittämisen. Arviointitutkimukseen kuuluu joka tapauksessa kausaalipäätelmien tekeminen. Tätä lähtökohtaa voi puolustaa ainakin kahdesta näkökulmasta. Inhimillinen toiminta - tässä tapauksessa julkinen politiikka ja sen osana hallintokoneisto - ei voi perustua muuhun kuin syiden ja vaikutusten tuntemiseen. Päinvastainen tilanne on mieleton. Toisaalta voidaan ajatella, että kaikkialla siellä missä on tämänlaista inhimillistä toimintaa, pitää olla kausaalista selittämistä. Tämä tarkoittaa kaikessa yksinkertaisuudessaan ymmärrystä syiden ja vaikutuksen välisestä yhteydestä. Mikäli tämä näkökohta unohtetaan, voimme heittää hyvästit tarkoituksenmukaiselle arviointitutkimukselle, jonka tavoitteena on yhteiskunnallisen hyvän edistäminen<sup>3</sup>.

Perti Tötön (2003) mainiota ilmaisua mukailien arviointitutkimuksessa tarvitaan kausaalisuhteen paljastamiseksi teoreettisia arvauksia ja oletuksia siitä, mikä X voi Y:n aiheuttaa, sekä empiiristä näyttöä, että X todella aiheuttaa Y:n. Vaatimus kausaalipäätelystä on kuitenkin ollut herkkä paikka monille arviointitutkimuksen kentällä toimivalle. Voi vaan kysyä, miksi näin on ollut? Vastaus tähän kenties liittyy metodologisiin kysymyksiin, onhan näet niin, että kausaalipäätelmien tekemiseen vaaditaan kahdenlaista tietoa: testattujen hypoteesien myötä kasautunutta teoreettista tietoa sekä puhtaimmassa ja joiden mielestä uskottavimmassa muodossa tai kokeellista, korrelatiivista tai vähintäänkin kvasikokeellista empiiristä tietoa joiden ilmiöiden syiden ja seurausten välisestä suhteesta.

Ilmeisesti juuri kokeellisten ja kvasikokeellisten asetelmien vuoksi kausaalijattelua on karsastettu arviointitutkimuksessa. Voitaisiinko siis ajatella, että kausaalisuhteen toteaminen olisi hyväksyttävämpää, jos iänikuiset koulukuntariidat kokeellisten ja kvasikokeellisten ja konstruktivisen koulukunnan välillä jätettäisi taka-alalle?

Voitaisiinko löytää toinen tie ja mahdollisuus, jossa kausaalisuhteen ymmärtämisen tavoittelu olisi arvioinnin tekijöiden johtotähti, mutta vaille kokeellisen, kvasikokeellisen ja ex post facto -asetelman "historiallisia" rasitteita?<sup>4</sup> Kyse ei ehkä olekaan niin isosta asiasta kuin voisi luulla. Vaikka avoimissa järjestelmissä kausaaliset

mekanismit ovat piilossa erilaisten tapahtumien ja muiden ilmiöiden takana, ei niiden esiin kaivaminen aina edellytä kokeellista asetelmaa. Ne voidaan kaivaa esille esimerkiksi kriittisen realismin hengessä osin kokeellisen intervention, osin abstrahoidulla, osin näitä kahta ääripäätä toisiinsa sekoittaen? (tästä myös Mohr 1995). Yksi mahdollinen etenemisreitti kausaalijattelun suuntaan voisi olla soveltaa kussakin arviointitoimeksiannosta yksinkertaista kontrafaktuaalista päättelyä, joka onnistuu aina arvioitavasta kohteesta ja arviointitoimeksiannosta riippumatta. Minkä tahansa vaikkapa projekti- tai ohjelma-arvioinnin yhteydessä voi kysyä, mitä olisi tapahtunut jos tätä projektia tai ohjelmaa ei olisi lainkaan ollut olemassa riippumatta siitä, perustuuko tutkimusasetelma ja aineistonkeruustrategia määrällisiin tai laadullisiin menetelmiin.

David Hume on hyvä referenssi arviointitutkimuksen tekijöille kausaalipäättelyä koskevien ajatusten lisäksi myös toisesta syystä. Häntä on nimittäin pidetty äärimmäisyyksiin asti menevänä skeptikkona. Skeptisyys korostuu erityisesti myös niin sanotussa tavoitteista vapaassa arviointimallissa, jossa ei sitouduta tarkastelemaan sitä, mitä arvioinnin kohteena oleva ohjelma tai projekti on onnistunut saavuttamaan sille asetetut tavoitteet. Tavoitteista vapaa arviointimalli lähtee liikkeelle periaatteellisesta ajatuksesta (joka tietysti rajautuu käytännön arviointikysymyksiksi arviointihankkeen edetessä), että mitä kaikkea mahdollista ohjelman tai projektin toteutuksen aikana on tapahtunut ottamatta kantaa, ovatko vaikutukset olleet aiottuja tai ei-aiottuja, positiivisia tai negatiivisia, ja niin edelleen.

#### ALISOSIAALISTETUISTA LÄHESTYMISTAVALLISISTA MALLEISTA TARKOITUKSEN- MUKAISIIIN LÄHESTYMISTAPOIHIN

*Arviointitiedon tuottamisen logiikasta: yksi mahdollinen teoreettinen lähestymistapa*

Arvioinnin koulukunta- ja standardikritiikin yhteydessä ei saisi unohtaa sitä, miten arviointitutkimuksia tehdään. Miten arvioinnin kodifioidut lähestymistavat käännetään käytännössä arviointiasetelmiksi ja empiiristen tutkimusongelmien ratkaisumalleiksi? Tässä jaksossa pohditaan sitä, miten arvioinnin lähestymistavallisia ja metodologisia valintoja on sovellettu käytännössä.

Pahimmillaan kyse on alisosiaalistettujen ulkoa kopioitujen mallien kritiikittömästä, mielikuvituksestomasta ja mekaanisesta toistamisesta, parhaimmillaan syvällisen evaluatiivisen mielikuvituksen hyödyntämisestä niin, että kuhunkin arviointiongelmaan kehitetään tarkoituksenmukainen arviointiasetelma.

Tässä jaksossa kritiikin kohteena on erilaisten arviointitutkimusten lähestymistapojen ja menetelmien mekaaninen soveltaminen, joka on ollut seurausta ulkoaopittujen mallien kritiikittömästä tulkinnasta. Arviointitutkimuksen tekemisen alisosiaalistetut mallit perustuvat formaaliin logiikkaan ja ne edustavat suoraviivaista, lineaarista tutkimustyön ideaalia. Kun suunnitteluvaiheessa on kontrolloitu kaikki potentiaaliset epävarmuustekijät, mikään ei voi mennä vikaan ja tutkimus voidaan toteuttaa suunnitellusti. Alisosiaalistetut muutosteorioiden voidaan esittää vaiheistettuina jaksoina, jotka voidaan todentaa ulkopuolisesti ja edistymisen kvantifoiden. Muutos on korostuneen lineaarinen toimitus, jonka vaiheet erottuvat selkeästi toisistaan.

Seuraavassa oleva tarkastelu perustuu oletukseen, että arviointitutkimus on soveltavaa tutkimusta, jolla on aina olemassa jokin tilaaja. Tässä tarkastellaan esimerkin omaisesti sitä, millainen maailma avautuu eteemme tarkasteltaessa arvonlisän tuottamista arviointipalvelujen tuottajan näkökulmasta. Kysymys on rajapinnasta asiakkaan ja lisäarvoa tuottavan palvelujen toimittajan välillä. Tässä rajapinnassa vallitsee oma logiikkansa kun asiaa katsotaan palvelujen tuottajan näkökulmasta - sellaisesta, mikä ottaa huomioon palvelujen tuottajan toimintatavat asiakassuhteen hallintaa silmällä pitäen. Jeffrey F. Rayport ja Bernard J. Jaworski (2004) sanovat, että asiakassuhteen hallinnassa on vähintään neljä tasoa ja tällä havainnoilla on merkitystä myös arviointitoiminnan palvelujentuottajien toimintaa analysoitaessa.

Arviointitutkimusta tekevän ja arviointitutkimuksen tilaajan välillä pitää olla ensinnäkin (1) fyysinen foorumi, siis tila jolla kohdata. Palvelun tuottajan näkökulmasta tulee siis olla jokin paikka, jossa asiakkaita voi ylipäänsä kohdata. Arviointitoiminnan kannalta havainto on mielenkiintoinen, koska kyseessä ei ole konkreettisen tuotteen myynti, vaan nimenomaan asiantuntijapalvelun tuottaminen. Tästä näkökulmasta tarkasteluna voidaan tietysti esittää vastaväite, että yhteiskuntahan on nykyään tietoyhteiskunta, jossa



tietoverkot mahdollistavat ajatustenvaihdon ja sähköisen tiedonvälityksen. Näin tietysti on, käytännössä kuitenkin arviointitoimeksiantojen onnistunut toteuttaminen todellakin edellyttää fyysistä kohtaamista tilaajan ja toimittajan välillä. Tällä varmistutaan muun muassa yhteisestä käsityksestä suhteesta siihen, mitä ollaan tekemässä<sup>5</sup>.

Asiakasrajapinnan toinen taso on luonteeltaan (2) kognitiivinen. Se tarkoittaa, että palvelun toimittaja onnistuu pääsemään arviointipalvelun tilaajan eli asiakkaan kanssa samalle "aaltopituudelle", mikä tarkoittaa käytännössä sitä, että arviointipalvelun tuottaja asettuu ikään kuin arvioinnin tilaajan asemaan ja todella pyrkii hahmottamaan maailmaa arvioinnin tilaajan näkökulmasta. Arvioinnin toteutuksessa tässä liikutaan puolueellisen ja riippumattoman arvioinnin vaikeasti hahmotuvassa välimaastossa, koska tässäkin tapauksessa arvioinnin tekijän pitäisi pystyä toimimaan niin puolueettomasti kuin mahdollisesti. Kognitiivisella tasolla toimiminen ei siis saa merkitä riippumattomuudesta luopumista.

Asiakasrajapinnan kolmas pinta on (3) emotionaalinen ja asenteisiin liittyvä. Arviointitutkimuksen tekijän (so. palvelun toimittajan) on oltava kohtelias, olla ikään kuin oikealla tavalla "emotionaalisesti kalibroitu" suhteessa asiakkaan tilanteeseen ja tarpeisiin. Tähän sisältyy vaatimus ikään kuin mahdollisimman suuren arvonlisän tuottamisesta nimenomaan siinä suhteessa, miten asiakas palvelutoimeksiantoa arvioi ja suhteellistaa. Toimitetusta palvelusta tulee olla jotain hyötyä asiakkaalle ottamatta kantaa siihen, ovatko arvioinnin tulokset arvioinnin kohteen suhteen myönteisiä, kielteisiä tai jotain näiden väliltä. Tästä syystä johtuen arvioinnin tekijöiden tulisi kiinnittää erityistä huomiota siihen, millaisiin johdopäätöksiin ja kehittämisehdotuksiin he päätyvät ja miten nämä muotoillaan. Varsinkin kehittämisehdotusten pitäisi olla hyvin konkreettisia ja realistisia. Arvioinnin tekijöiden pitäisi pystyä artikuloimaan [i] mitä [ii] jonkun pitäisi tehdä [iii] jossain aikataulussa, [iv] mitä tämä tarkoittaa tai vaatii käytännössä esimerkiksi resurssien suhteen ja [v] mitä mahdollisia vaihtoehtoisia toimintamalleja arviointipalvelun tilaajalla voisi olla käytössään arvioinnin jälkeisiä toimenpiteitä suunnitelluissaan.

Asiakasrajapinnan neljäs oleellinen taso liittyy (4) saavutettavuuteen tai sananmukaisesti Rayportin ja Jaworskin termein käännettynä kytket-

tävyyteen (connectedness). Arviointitutkimuksen tekijän pitää ymmärtää arvioinnin tilaajan resurssiyhteydet viitaten näillä eritasoisin asioihin, muun muassa siihen, millaisia palveluntarvekokonaisuuksia asiakkailla on. Tällä voitaisiin arviointitoiminnan kannalta tarkoittaa käytännön arviointityössä nykyisin yleisesti esiin nousevaa kysymystä arvioinnin roolista. Usein kuulee nimittäin sanottavan, että ulkoinen arviointi on muuttunut tai muuttumassa yhä enemmän konsulttivaksi työotteeksi. Tähän teemaan palataan tässä artikkelissa jäljempänä.

Rayportin ja Jaworskin ajatuksia miettiessä huomataankin, että näinhän juuri omia etujaan ja tietysti myös asiakkaidensa tarpeita maksimoivan palvelun tuottajan kannattaakin toimia. Arviointitiedon tuottamisen jälkeen arviointien tilaajat tarvitsevat usein laaja-alaista asiantuntijatukea tuotetun tiedon prosessoimiseksi - ikään kuin käännettäessä se hallinnon toimintaympäristöön sopivaksi ja päätöksenteon mahdollistamiseksi - ja olisi absurdia ajatella, että arviointipalvelujen tilaajan tarpeet loppuisivat sillä hetkellä kun arviointiraportti valmistuu painokoneen uumenista. Näin ajatellen huomataan, että samalla kun arviointitiedon tuottaja maksimoi omaa etuaan ja vahvistaa omaa markkina-asemaansa jo olemassa olevan asiakkaan silmissä, vahvistaa se samalla asiakkaan kokemusta kokonaisvaltaisesta palvelukokemuksesta.

Olemme mutkikkaan yhtälön edessä. Jos uskomme mitä Rayport ja Jaworski sanovat, niin arviointitoiminnan kentällä jalansijan saaminen - tämä havainto siis koskee ennen muuta arviointitutkimuksen tekijöitä - edellyttää "kytkentää" eli arviointiasiantuntijatieiden tuotteistamista arviointitiedon hyödyntäjän käyttöhyödykkeeksi. On siis arviointitiedon tuottajan edun mukaista kasata symbolista pääomaa ikään kuin arviointitiedon hyödyntäjän "mieliksi". Ajatus siis pahimmillaan kuuluu: se mikä symbolisen (ja osin myös taloudellisen) pääoman kasaamisessa kenties arviointien tekijän näkökulmasta voitetaan, samalla arvioinnin tekijä häviää osan uskottavuudestaan ja leimautuu ortodoksisen tiedon tuottajaksi.

#### *Neljännän sukupolven jälkeen tarvittavat arviointitaitot*

Edellistä pohdintaa seuraava kysymys on, millaisia taitoja arvioinnin tekijältä vaaditaan jos

tähtäämme siihen, ettemme sortuisi alisosiaalitetuttujen arviointimetodologian ja -tekniikoiden soveltajiksi? Voidaan ajatella, että kompetenssi-vaatimukset ovat viidenlaisia. Arviointitutkimuksen tekijöillä pitää olla monesta osatekijästä koostuvaa valtaa kykyä (power to) saada aikaiseksi jotain.

*Hyvyys ihmisenä.* Arviointitutkimuksen tekijältä edellytetään aitoa rehellisyyttä itseään, arvioitavaa kohdetta ja arvioinnin tilaajaa kohden. Arviointisijan on oltava hyvä ihminen siinä mielessä, ettei totuutta saa vääristellä, kaikista tehdyistä aineistohavainnoista pitää reilusti kertoa, hyvät ja huonot asiat pitää pystyä kertomaan avoimesti ja asian vaatimassa tasapainossa. Vaikka tämä kuulostaa varmaan hirveän naiivilta, vaatii aito refleksiivisyys kurinalaisuutta ja tähän kurinalaisuuteen kuuluu myös rehellisyys ja bona fide -tyyppinen kirjoittaminen.

*Verkostoituminen kentän avaintoimijoiden kanssa.* Esimerkiksi Pierre Bourdieun (1994; 2004a ja 2004b) kenttäteorian valossa on ilmeistä, että arviointitoiminnan kentän toimijat toimivat keskenään verkostoituen. Näin ajatellen tuntuisi loogiselta, että myös yksittäiset henkilö- ja instituutiotoimijat verkostoituisivat keskenään, tai että ainakin tämä olisi heidän etujensa mukaista. Arvioinnin tilaajien kannattaa verkostoitua keskenään ja sama koskee myös arvioinnin tekijöitä<sup>6</sup>.

*Tarkoituksenmukainen arvioitavan kohteen ymmärtäminen.* Vaatimuksesta arvioitavan kohteen ymmärtämiseen sisältyy paljon tarunhoitoa. Joidenkin mielestä arviointia ei voi tehdä, ellei arvioinnin tekijä ole ollut konkreettisesti mukana tekemässä jotain sittemmin arvioitavaa toimintaa. Ohjelmia ei ikään kuin voisi arvioida, jos et ole ollut ohjelmajohtajana tai projektipäällikkönä, ja mieluummin juuri nyt tarkastelun ja arvioinnin kohteeksi tulevassa ohjelmassa tai projektissa. Tähän sisältyy merkittävä näkökulman vinoutumisen harha, mitä ei aina tunnisteta. Joskus voit nimittäin olla niin lähellä arvioitavaa kohdetta ettei enää näe sitä. Toisaalta jotkut arviointien tilaajat edellyttävät arvioinnin tekijöiltä täydellistä riippumattomuutta ja joissakin tapauksissa tällainen vaatimus menee mielettömyydeksi. Mohrin (1995) ajatuksiin vedoten voisi sanoa, että arvioinnissa on olennaista kahden haasteen ratkaisu: oikeiden kysymysten esittäminen ja niihin oikeiden vastausten saaminen. Ja jotta oikeita kysymyksiä voi esittää, pitää arvioitavana olevan kohteen toiminnallinen logiikka tuntea riittävällä

tarkkuudella.

Mikä sitten ratkaisuksi? Riittäisikö se, että arvioinnin tekijöillä on riittävä ymmärrys arvioitavan kohteen - instituutio, projekti, ohjelma, hanke, policy, tai vastaava - toimintalogiikasta. Eli, jos arvioinnin kohteena on vaikkapa jokin ministeriön käynnistämä ja toimeenpanema ohjelma, arvioinnin tekijällä pitää olla ymmärrys miten valtionhallinnon tilivirastojen ohjelmia suunnitellaan, budjetoidaan, ohjataan, niiden tavoitteita mitataan, ja niin edelleen. Tämä kannanotto tulee huomattavan lähelle sitä, mikä on teoriaperusteisen arviointitutkimuksen keskeinen lähtökohta esimerkiksi ohjelma-arvioinnissa (Chen 2005). Näin ajatellen arvioinnin tekijän tehtävänä on ensimmäisessä vaiheessa konstruoitava arvioitava ohjelma loogisena mallina muodostuen erilaisista toiminnallisista aktiviteeteista ja näiden vuorovaikutussuhteista. Arvioinnin tekijän on siis ymmärrettävä se "maasto", minne hän putoaa ottaessaan vastaan arviointitoimeksiannon ja nopeasti muodostettava yleiskäsitys arvioitavan arvioinnin kohteena olevan toiminnan yleispiirteistä (toiminnan tavoitteet, nykyvaihe, keskeiset kehittämissaasteet, ja niin edelleen). Tämä ei suinkaan tarkoita sitä, että arvioinnin tekijän pitäisi etukäteen tietää nämä asia, koska sellaisessa (pahimmassa) tapauksessa arvioinnin tekijän vääristyneet intuitiiviset hypoteesit kääntyvät nopeasti johtopäätöksiksi.

Arvioinnin tekijän on vaan otettava nopeasti tilanne "haltuun" ja aina tämä ei ole helppoa. Jokatapauksessa arvioinnin tekijällä tulee olla valmiuksia kausaaliseen päättelyyn ja tämän tapaisen ajattelumallin soveltamiseen. Arvioinnin tekijän pitää pystyä selittämään asioita sen vuoksi, että reaali maailman havainnot eivät automaattisesti merkitse ilmiöiden ja tapahtuneen syiden ymmärtämistä ("taivas on sininen ja vaikka tämän voi todentaa katseella emme vielä tiedä miksi taivas on sininen"). Selittäminen on - Peter Liptonin (2004) termein - sitä, että tietoon (tietoisuuteen kasvuun) lisätään jotain, jotta ymmärtäisimme mistä on kysymys. Tätä "jotain lisäämistä" voisi tässä yhteydessä kutsua evaluatiiviseksi tulkinaksi.

*Reflektioiva tutkimusmenetelmällinen osaaminen.* Kaikkien arvioinnin tekijöiden ei tarvitse osata kaikkia tutkimusmenetelmiä, mutta jokaisen heistä tulee tuntea erilaisten tutkimusmenetelmien soveltuvuus ja soveltuvuuden rajat suhteessa kulloiseenkin tutkimusongelmaan.



Arvioinnin tekeminen on älyllistä toimintaa, joka lähtee liikkeelle tutkimusongelmasta, johtaa siitä ongelman ratkaisukeinot ja toteuttaa salapolii-sitehtävä eli ongelman ratkaisu. Tätä taustaa vasten on hyvin mielenkiintoista, että silti edelleen tieteenalojen sisälle syntyy nykyisin koulukuntia tai tutkimustraditoita, jotka ovat kehkeytyneet yhden aineiston keräämisen tai analysointitekniikan ympärille. Vastaavasti voidaan sanoa, että arviointitutkimuksen kentällekin on syntynyt koulukuntatyyppejä fraktioita jonkun tietyn erillisteen varaan.

Suomestakin on kohtalaisen helppo löytää - suppean arviointitutkimuskenttämme sisältä - jonkin tietyn tutkimusmenetelmän tai tutkimuksellisen lähestymistään monomaanikkoja, jotka jäykästi pitäytyvät omilla lähtökuopissaan. Esimerkkinä voi vaikkapa mainita vertaileva analyysi hallintotieteen alalla ja realistisen arviointiparadigman kritiikin soveltaminen lähinnä sosiaali-, terveys- ja työllisyyspolitiikan ohjelmissa ja projekteissa.

Huey-Tsyh Chenin (2005) termiä käytettäessä arviointitutkimuksen menetelmän valinta perustuu situationaaliseen harkintaan: on valittava se, mikä toimii parhaiten jossain tietyssä tilanteessa, eikä ole olemassa mitään universaalia, aina toimivaa tutkimusmenetelmää. Arvioinnin tekijän tulee olla kriittinen käyttämiensä tutkimusmenetelmien suhteen ja annettava tältäkin osiin aseet "vastustajien" käsiin heidän arvioidessaan arviointitoimeksiannon metodologisia lähtökohtia.

*Kolmiportainen konsensuaalinen tulkinta tulosten tuottamisessa.* Arviointitutkimusta oppii parhaiten sitä yhdessä tekemällä muiden kanssa. Ehkä kaikkein tehokkain tapa korjata virheet ja niihin liittyvät pelot, on kyky nauraa omille virheilleen. Näin ollen voidaan olettaa, että arviointitutkimuksessa havaintojen tuottaminen ei riitä, eikä siinä analyysivaiheessa sovellettu reflektiivinen tutkimusmenetelmällinen osaaminen. Arvioiva, evaluaatiivinen tulkinta syntyy kun havainnoista johdetaan johtopäätöksiä ja näistä edelleen toimenpide-ehdotuksia. Tässä yhteydessä toimii konsensuaalisuus parhaimmalla mahdollisella tavalla. Arviointitutkimuksen tulosten koeteltavuutta lisää se, arvioinnin johtopäätöksiä (1) arvioidaan kriittisesti lähellä olevan tutkijatiimin tai -yhteisön sisällä<sup>7</sup>, sen (2) jälkeen yhdessä arvioinnin tilaajan ja (3) lopuksi yhdessä, jos suinkin mahdollista, arvioinnin kohteen edustajien kanssa. Tällainen kolmiportainen konsensuaali-

suus ei tarkoita arvioinnin tekijöiden mielipiteiden vesittämistä, vaan nimenomaan johtopäätösten ja toimenpide-ehdotusten oikeaan osumisen varmistamista.

## ARVIINTITUTKIMUS SUOMESSA 2000-LUVULLA: REFORMIEN UUDISTUSTEN TODENTAMISESTA HALLINNON TIETOTUO- TANTOON JA TIEDONHALLINTAAN

### *Arviointi vakiintuu*

Edellä sanotun perusteella näyttää siis aidosti siltä, että samalla kun arviointitutkimus edelleen hakeaa paikkaansa muiden sitä lähelle sijoittuvien toimintamuotojen rinnalla, se muuntuu sisällöllisesti muun muassa siihen liitettävien motiivien osalta. Julkishallinnon reformeja ja arviointitoiminnan kehitystä on pidetty toisiinsa läheisesti linkittyneinä teemoina jo pitkään. Esimerkiksi Wollmann (2003b) sanoo, että mainitut asiat ovat olleet jo 30 vuotta kuin siamilaiset kaksoset. Mainittu Wollmann - ja monet muut hänen lisäkseen (esim. Pollitt & Bouckaert 2003; Derlien & Rist 2002; Thoenig 2003) - katsoo, että julkishallinnon arviointitoiminta on kehittynyt nimenomaan kytköksissä 1980- ja 1990-luvuilla läpimurron saavuttaneen uuden julkisjohtamisen toimintafilosofian rinnalla.

Vaikka suomalainen julkishallinnon arviointitoimintaa on ollut olemassa määrittelytavasta riippuen jo kymmeniä vuosia, 1990-luvulla tilanne muuttui. 1990-luvulla valtiovarainministeriön hallinnon kehittämissosastolle muodostui merkittävä rooli arviointitoiminnan koordinoijana ja uusien arviointitoimintaa koskevien vaikutteiden välittäminen suomalaiseen hallintoon nimenomaan juuri OECD-maista ja enenevässä määrin myös Euroopan komission suunnasta. Suomessa valtiovarainministeriö tilasi 1990-luvun kuluessa ja erityisesti sen jälkipuoliskolla alan asiantuntijoilta - muun muassa Christopher Pollittilta, Geert Bouckaertilta ja heidän tutkijatiimeiltään - analyyseja keskushallinnon rakenteellisiin selvityksiin liittyen. Näiden selvitysten tarkoitusperä liittyi tavallaan juuri näihin hallinnon kehittämistoimenpiteiden dokumentointiin "ulkopuolisin" silmin. Vastaavia analyyseja hallinto teetti myös suomalaisilla asiantuntijoilla. Merkittävän roolin tässä suhteessa sai Helsingin yliopiston professori Markku Temmes, jonka nimi voidaan liittää

1990-luvulla varsin moniin ministeriöiden tilaamiin arviointitutkimuksiin tai arviointitutkimuksia muistuttaviin selvityksiin. (Tästä kehityksestä ks. Ahonen ym. 2002).

Varsinaisten politiikkalohkojen, ohjelmien (usein myös projektien) ja instituutioiden arviointityössä käytettiin Suomessa hyväksi vielä 1990-luvulla kohtalaisen suppeaa kotimaista arviointiosaajien joukkoa, jonka viiteryhmä löytyi tavallisesti yliopistoista, tutkimuslaitoksista ja ehkä jossain määrin myös yksityisistä konsulttitoimistoista. Jo 1990-luvulla - ja etenkin sen jälkipuoliskolla - oli nähtävissä arvioinnin tekijöiden erikoistumista tietyille politiikkasektoreille ja 2000-luvulla tilanne muuttui 1990-luvun tilanteesta oleellisesti. 2000-luvulla julkishallinnon arviointitoiminnan kenttä on pirstoutunut toimijoiden osaamisen, intressialueiden ja erilaisten kilpailu- ja kamppailuasetelmien myötä erilaisiin ryhmittymiin, joiden edustajat tulevat niin yliopistollisista, julkisten tutkimuslaitosten ja yksityisen sektorin konsulttitoimistojen yhteisöistä.

Seuraavassa tarkastellaan alustavasti julkishallinnon arviointitoiminnan muutosta neljän erityisen vastakkainasettelun näkökulmasta. Tarkastelu perustuu jossain määrin Wollmannin (2003b) alunperin esittämään typologiaan, jota on kuitenkin muokattu tässä yhteydessä Suomen oloihin sopivaksi.

#### *Perinteinen arviointitoiminta ja metaevaluaatio*

Perinteinen arviointitoiminta on kohdistunut politiikkalohkoihin, ohjelmiin, projekteihin, erilaisiin hankkeisiin ja julkisiin instituutioihin (esimerkiksi valtion virastot ja laitokset). Tässä arviointitavassa keskeisenä lähtökohtana on ollut tarkastelun kohteena olevalle toiminnalle asetettavat tavoitteet. Toimintaa siis tarkastellaan siis suhteessa asetettuihin tavoitteisiin. Metaevaluaatio taas on lähtökohtaisesti hyvin erityyppistä kuin edellä mainittu perinteinen arviointitapa. Siinä voi olla kysymys ainakin kahdesta asiasta eli metaevaluaation käsite voidaan tulkita ainakin kahdella tavalla. Sillä voidaan ensinnäkin viitata arvioinnin laadun arviointiin, mikä käytännössä voidaan suorittaa vaikkapa jollain valitulla arviointitoiminnan standardilla. Tämänäköisestä metaevaluaatiota on Suomessa toteutettu jonkun verran, muun muassa rakennerrahastoarviointien, valtion virastojen ja laitosten arviointien ja lääninhal-

litusten peruspalvelujen arvioinnin yhteydessä. Toinen tapa mieltää metaevaluaation käsite on se, että sillä tarkoitetaan vaikkapa jonkun ohjelmakokonaisuuden arvioimista siten, että ohjelmaan kuuluvien yksittäisten hankkeiden (ja näitä hankeaineistoja koskevia) aineistoja ja tietoja käytetään ikään kuin ohjelmatasoisessa tarkastelussa "toisen käden lähteenä" sen jälkeen kun niitä on alun perin käytetty mainituissa hankearvioinneissa. Suomessa tämänäköisestä metaevaluaatiota on sovellettu erityisesti rakennerrahasto-ohjelmien arvioinneissa, joissa toimenpidekokonaisuuksien arviointien tekijät periaatteessa "luovuttavat" aineistonsa ohjelmatasoisen arvioinnin tekijän aineistomassaksi.

Tulevaisuudessa eri tavoin toteutettujen metaevaluaatioiden määrä kasvaa jo siitä syystä, että perinteistä arviointitoimintaa on jo olemassa niin paljon. Kyse on siis tavallaan tiedon suodattamisesta varsinaista päätöksentekoa mahdollistamiseksi.

#### *Sisäinen ja ulkoinen arviointi*

Varsin usein vielä nykyisin näkee keskustelua itse tai ulkopuolisen tahon toteuttaman arvioinnin hyvydestä tai huonoudesta. Käytännössä tämä jako on hyvin kategorinen, ja vailla käytännön merkitystä. Nykyään valtionhallinnossa toteutetaan arviointitoimintaa sekä sisäisenä että ulkoisena toimintana eli osa arvioinneista tehdään itse ja osa teetetään ulkopuolisilla asiantuntijoilla. Vielä nykyisinkin vallalla oleva selvitysmiesinstituutio edustaa tavallaan yhtä merkkipaalia ulkoisen arvioinnin toteutusmallissa. Eli tällaisissa tapauksissa hallinto tilaa ulkopuolisen käsityksen jonkun asian tilasta ja toivottavasta kehityssuunnasta. On tietysti helppo uskoa, että itse tehtyyn arviointiin voi lähtökohtaisesti liittyä erilaisia "kotiinpäin vetämisen" epäilyjä eli pelkoja eräänlaisten "pohjoiskorealaisten" arviointien tekemisestä. Tällainen pelko liittyy siis ajatukseen, että itse tehty arviointi olisi jollain tavalla epäluotettavampaa kuin ulkoisen (riippumattoman) tekijän tekemä arviointi<sup>8</sup>.

Arviointien tekijöitä voi tässä yhteydessä kehoittaa aitoon kriittisyyteen ja ja refleksiiviseen epäilyyn. On nimittäin niin, että ulkoistenkin - näennäisesti riippumattomien - arvioinnin tekijöiden joukossa on instituutioita ja toimijoita, jotka tuottavat oikeaoppista tietoa arvioinnin tilaajille

ja näin menetellessään voi vaan arvailla heidän motiivejaan, jotka tietysti voivat olla hyvin moninaisia. Kyseessä on siis tilanne, jossa ulkoinen arvioinnin tekijä haluaa esimerkiksi miellyttää arvioinnin tilaajaa. Tuloksena on näennäisesti korrekti raportti, mutta jossa ei ole kuitenkin aidosti haluttu kyseenalaistaa arvioinnin kohteena olleen ohjelman, hankkeen tai projektin järjestyttävyyttä. Kyse on tilanteesta, jossa tiedon tuottaja toimittaa "rauhottavaa" tietoa vallan tai hallitsevan luokan käyttöön ja samalla - kyynisesti nähtynä - he maksimoivat seuraavan arviointitoimeksiannon saamisen<sup>9</sup>.

Tulevaisuudessa sisäisen ja ulkoisen arvioinnin välinen kuilu kapenee entisestään, koska arviointitekniikat ja -menetelmät kehittyvät koko ajan. Olennaista tässä muutoksessa on se -luonnollisesti varmistuttuamme arvioinnin validiteettista ja reliabiliteetista - miten hyödynnettävää arviointitutkimuksilla tuotettu tieto loppujen lopuksi on. Tämä näkökulma ei ole mitenkään uusi tai uniikki arviointitutkimuksen metodologisella kentällä, muun muassa Michael Q. Patton (1997; 2001) on asiasta puhunut jo vuosia.

#### *Etukäteisarviointi, jatkuva arviointi ja jälkikäteis-arviointi*

Muun muassa kansainvälisten rahoittajien (Maailmanpankki, Yhdistyneet Kansakunnat) ja instituutioiden (Euroopan komissio) arviointipoliitikassa vakiintunut tapa ymmärtää arviointi kolmiulotteisena toimintana heijastui voimakkaasti Suomeen 1990-luvulla. Arviointia alettiin suorittaa - niin sanotusta common sense -mielikuvasta poiketen - ennen arvioitavan kohteen toimeenpanoa ja sen aikana sen lisäksi, että arviointi voidaan tehdä kun arvioinnin kohteena oleva toimenpide on päättynyt. Mainitusta kehityksestä huolimatta 1990-luvulla arvioitiin julkishallinnon arviointitoiminnan erityishaasteeksi Suomessa arviointitiedon heikkoa kytkeä päätöksentekoon ja vähäistä etukäteisarviointipainotusta. (Harrinvirta ym. 1998). Sittemmin, erityisesti Euroopan unionin jäsenyyden ja rakennerrahastoarviointien myötä suomalainen hallintokulttuuri sai vaikutteita eurooppalaisista arviointitoiminnan perinteistä, joissa korostui mainittu etukäteisarviointien merkityksen korostaminen. Etukäteisarviointien tekeminen ja tilaaminen on tämän myötä lisääntynyt myös varsinaisten rakennerrahasto-

ohjelmien arviointien ulkopuolella. Yhtenä merkkipaaluna voidaan tässä suhteessa myös pitää lainsäädäntövalmisteluun kytketty vaatimus tiettyjen oletettujen vaikutusten kartoituksesta.

Näyttää siltä, että lähivuosina etukäteisarvioinnin merkitys tulee vahvistumaan perinteisesti hyvin esillä olleiden jatkuvan ja jälkikäteisarvioinnin rinnalla. Julkishallinnon tietotarpeiden suuntautuessa yhä kasvavassa määrin tulevaisuuteen päin tuntuisi myös luonnolliselta, että erilaisille ennakoitiedon tuottamiselle ja tulevaisuuden tutkimuksen menetelmille erikoistuneille tiedontuottajille olisi yhä enemmän kysyntää.

#### *Arviointi tuomarina olemisena ja oppimisen mahdollistajana*

Kuten edellä on tuotu esille, ovat arviointityylit ja -lähestymistavat ovat hyvin moninaisia, arvioinnin metodeista ja aineistonkeruustrategioista puhumattakaan. Arvioinnin motiivit voidaan periaatteellisella tasolla palauttaa kahteen puhtaaksiviljeltyyn malliin, yhtäältä tuomarina olemisena (arvioinnin rooli ensisijaisesti tilintekovastuun osoittajana) ja toisaalta oppimisen mahdollistajana, ikään kuin toiminnan "korjaajana". Tämä malli pohjautuu pitkälti niihin ajatuksiin, joita aikanaan esitti Michael Q. Patton arviointitiedon hyödyntämisen tärkeyttä painottavassa, arviointitoimijoiden joukossa kirjoittajansa viimeistään maailmanmaineeseen nostaneessa teoksessaan Utilization-focused Evaluation (1997). Patton erotti omassa typologiassa omaksi luokakseen yleisesti tiedon tuottamiseen liittyvät motiivit, joiden ilmenemismuotona on kaiketi ennen muuta akateeminen julkaisutuotanto sen sijaan, että kyseeseen tulisivat ensisijaisesti arvioinnin tilaajan toimesta julkaistut niteet.

Seuraavaksi heräävä kysymys on, millainen kanta näihin motiiveihin olisi sitten otettava? Jos motiivien suhteen on pakko jokin kanta ottaa, olisi se ehkä kannanotto pluralismin puolesta eli kaikkia mainittuja motiiveja tarvitaan ja ne kaikki ovat - arviointitehtävästä riippuen - tarkoituksenmukaisia. Arviointi voi olla tilintekovastuun osoittamista, mutta myös parantamiskohteiden osoittamista, organisatorisen oppimisen mahdollistamista, kulloisista painotuksista riippuen. Vielä 1990-luvulla puhuttiin paljon arviointitiedon hyödyntämisen tähdellisyydestä poliittis-hallinnollisessa päätöksenteossa, eikä vähäisimmässä

määrin edellä mainitun Pattonin vaikutuksesta. Nyt, 2000-luvulla, voidaan aidommin puhua arviointitiedon tuottamisesta osana julkishallinnon tiedon ja tietämyksen hallintaa (knowledge management). Oletettavaa on, että julkishallinnon arviointitoiminnan kulttuuri vahvistuu tulevaisuudessa ja sen merkitys, luonne ja toimintatavat selkiintyvät sitä läheisesti sivuavien toiminnan muotojen suhteen. Kokonaan oma lukunsa on arviointiteknologian kehittyminen. Missä määrin voidaan jatkossa puhua aidosta tutkimus- tai arviointiavusteisesta julkisesta politiikasta on iso yhteiskuntapolitiittinen kysymys. Muutos vaatii osaamisen kasvattamista niin arviointitiedon tuottajien kuin hyödyntäjienkin parissa.

### JULKISHALLINNON PÄÄTÖKSENTEKOA TUKEVA ARVIINTITUTKIMUS - TAKANA LUPAAVA MENNEISYYS?

Julkisen sektorin ja hallinnon rakenteiden ja päätöksentekojärjestelmien tulevaisuus ratkaisee pitkälti myös sen, millaisia vaatimuksia ja millaista kysyntää arviointitoimintaan tulevaisuudessa kohdistuu. Oletettavaa on, että ainakin laatuvaatimukset kasvavat, volyymimeilessä arviointitoiminta kasvaa edelleen ja arviointi kytketty entistä enemmän konsultointiin ja toimintatapojen kehittämiseen. Monentasoinen hallinto ei jatkosakaan tule toimeen ilman palautemekanismeja ja toiminnan kehittämisen kyseenalaistamisen välineitä, jollaiseksi arviointi ilman muuta on luokiteltava.

Edellä luonnosteltu kehitys johtaa siihen, että arviointitutkimuksen rooli muuttuu entisestään lähivuosina. Arviointiin liittyi aikanaan paljon odotuksia ja varmasti osaan näistä haasteista arviointitutkimus on pystynyt vastaamaan. 1980-luvulle juurensa juontaneet ja 1990-luvulla voimansa päivät kokeneet arviointitutkimuksen suunnannäyttäjät eivät ehkä sittenkään, eivätkä ainakaan heitä replikoivat kvasikoulukunnat, onnistuneet ratkaisemaan arviointitutkimuksen asemaa ja roolia koskevaa kysymystä julkishallinnon rakenteiden peruspilarina. 1990-luvulla kaikkialle levinneet pahimmassa tapauksessa manipulatiiviset arviointistandardit eivät myös onnistuneet pelastamaan tilannetta kuin ainoastaan osittain. Vaikka ne toki auttoivat kiinnittämään huomiota arviointiprosessien laatuun,

sisältyi niihin monenlaisia ongelmia, joista eettisiin valintoihin liittyvä partikularismin pulma ei ole vähimmästä päästä.

On luultavaa, että arviointitutkimuksen rooli ja asema julkishallinnon osana selkiytyy sen tiedonmuodostusta ja tiedonkäyttöä koskevien valintojen myötä. Arviointitutkimuksesta on hyvää vauhtia tulossa organisatorisen tiedon tuottamista ja tiedonhallintaa. Tästä aukeaa monia kiinnostavia kysymyksiä ja tutkimuksellisia haasteita. Olisi oleellista tarkastella sitä, millaista organisaation tieto on, millainen on organisaation tiedonkäsittelyprosessi, mihin toimintaan tiedon käyttäminen liittyy ja ennen muuta siihen, miten uutta tietoa luodaan. Organisaatiolle tarpeellinen tieto syntyy aina vuorovaikutustilanteissa. Tästä syystä huomion kiinnittäminen yksilöiden toimintaan on merkityksellistä. Ei ole olemassa mitään tietoa sellaisenaan, vaan tietoa konstruoidaan jatkuvasti erilaisissa vuorovaikutustilanteissa (Tsoukas & Vladimirou 2005; Tsoukas & Papoulias 2005).

Arviointitutkimus tuottaa analysoitua tietoa, joka ei ole vielä tietämystä. Tietämys on tavallaan organisatorisen tiedon korkein muoto. Se sisältää organisaation jäsenten tekemän tulkinnan raakadatasta ja siitä edelleen johdetusta informaatiosta. Muuntuakseen tietämykseksi - tavallaan merkitykselliseksi informaatioksi - tieto tulee tulkita siinä organisatorisessa kontekstissa, jota kyseinen raakatieto tai tiedonjyvä koskee. Tätä voisi havainnollistaa vaikkapa kartalla. Pelkkä kartta - fyysinen tai sähköinen artefakti - osoittaa maantieteellisiä yhteyksiä ja etäisyyksiä. Pelkkä kartan käsissä pitäminen ei vielä auta kartanlukijaa, koska kartta ei osaa lukea itseään. Kartanlukijalla pitää olla kyky sijoittaa itsensä kartalle, sen lisäksi kartanlukijalla pitää tavallisesti olla käsitys siitä, mihin hän on menossa, ja kolmanneksi kartanlukijalla pitää olla hahmotuskyky siitä, miten nykyisen paikan ja tavoiteltavan paikan välillä kannattaa edetä (mikä reitti on tarkoituksenmukaisin, esimerkiksi kävellen linnuntietä lyhyimmältä vaikuttava reitti ei aina oleärkevin, koska siinä voi joutua kiipeämään kohtuuttoman paljon). Aivan samaan tapaan organisaatiolla ei ole käyttöä tiedosta sinänsä jos sitä ei osata liittää kontekstiinsa ja jos tiedon rajallisuutta ei tunneta. Edellisestä on lyhyt harppaus ajatukseen organisaatiosta itsesäätelevinä järjestelminä. Tällöin olennainen kysymys on miten tämä itsesäätelevä järjestelmä toimii, miten yhteistä tietoisuutta

ja tietämyspohjaa rakennetaan ja ennen muuta, mille oletuksille tämä tietämyspohja kulloinkin rakentuu?

## KIRJALLISUUS

- Ahonen, Pertti, Virtanen, Petri 2005. Arvonlisää julkishallinnon arvioinnista? Suomalainen arviointitutkimuksen kenttä 2000-luvun alussa. Julkaisematon käsikirjoitus. Ilmestyy vuoden 2005 aikana.
- Ahonen, Pertti, Virtanen, Petri, Uusikylä, Petri 2002. Evaluation in Finland. Teoksessa Furubo, Jan-Eric ym. (toim.) *Evaluation Atlas*. New Brunswick: Transaction Publishers, s. 47-61.
- Bourdieu, Pierre 1994. *Sur la théorie de l'action*. Paris: Seuil.
- Bourdieu, Pierre 2004a *Esquisse pour une auto-analyse*. Paris, Éd. Raisons d'Agir.
- Bourdieu, Pierre 2004b. *Science of Science and Reflexivity*. Cambridge: Polity Press.
- Chen, Huey-Tsyh 2005. *Practical Program Evaluation. Assessing and Improving Planning, Implementation and Effectiveness*. Thousand Oaks ym.: Sage.
- Czamiawska, Barbara 2003. *This Way to Paradise: On Creole Researchers, Hybrid Disciplines, and Pidgin Writing*. *Organization* 10 (3) ss. 430-434.
- Derlien, Hans-Ulrich, Rist, Ray C. 2002. *Policy Evaluation in International Comparison*. Teoksessa Furubo ym. (toim.) *Evaluation Atlas*. New Brunswick: Transaction Publishers, s. 439-455.
- Fetterman, David M. 2004. *Empowerment Evaluation Principles in Practice*. New York: Guilford Publications.
- Fetterman, David M. 2001. *Foundations of Empowerment Evaluation*. Thousand Oaks, CA: Sage.
- Fetterman, David M., Kaftarian, Shakeh J., Wandersman, Abraham (toim.) 1996. *Empowerment Evaluation: Knowledge and Tools for Self-assessment and Accountability*. Newbury Park, CA: Sage.
- Furubo, Jan-Eric, Sandahl, Rolf (toim.) 2002. *Evaluation Atlas*. New Brunswick: Transaction Publishers.
- Guba, Egon G., Lincoln, Yvonna, S. 1989. *Fourth Generation Evaluation*. Newbury Park: Sage.
- Harrinvirta, Markku, Uusikylä, Petri, Virtanen, Petri 1998. *Arvioinnin tila valtionhallinnossa. Valtionvarainministeriön tutkimukset ja selvitykset 7/1998*. Helsinki.
- Lipton, Peter 2004. *Inference to the Best Explanation*. London & New York: Routledge.
- Mohr, Lawrence B. 1995. *Impact Analysis for Program Evaluation*. Thousand Oaks ym.: Sage.
- Niiniluoto, Ilkka 2003. *Totuuden rakastaminen*. Keuruu: Otava.
- Patton, Michael Q. 2001. *Use as a Criterion of Quality in Evaluation*. Teoksessa Benson, A.P. ym. (toim.) *Visions of Quality: How Evaluators Define, Understand and Represent Program Quality*. JAI/Elsevier Science. Oxford, s. 155-180.
- Patton, Michael Q. 1997. *Utilization-focused Evaluation*. 1. p. 1986. Newbury Park ym.: Sage.
- Pawson, Ray 2005a. *Evidence Based Policy: A Realist Perspective*. Ilmestyy vuonna 2005.
- Pawson, Ray 2005b. *Measure for Measures. Manifesto for Empirical Sociology*. London: Taylor & Francis Ltd.
- Pawson, Ray, Tilley, Nick 1997. *Realistic Evaluation*. London: Sage.
- Porter, Les J., Tanner, Steve J. 2004. *Assessing Business Excellence. A Guide to Business Excellence and Self-assessment*. Amsterdam ym.: Elsevier Butterworth Heineman.
- Pollitt, Christopher 2003. *The Essential Public Manager*. Philadelphia: Open University Press.
- Pollitt, Christopher, Bouckaert, Geert 2003. *Evaluating Public Management Reforms: An International Perspective*. Teoksessa Wollman, Hellmut (toim.) *Evaluation in Public-Sector Reform: Concepts, and Practice in International Perspective*. Cheltenham & Northampton: Edward Elgar, s. 12-35.
- Rayport, Jeffrey F., Jaworski, Bernard, J. 2004. "Best Face Forward." *Harvard Business Review*, December 2004, s. 47-58.
- Stake, Robert E. 2004. *Standards-Based & Responsive Evaluation*. Thousand Oaks ym.: Sage.
- Thoenig, Jean-Claude 2003. *Learning From Evaluation Practice: The Case of Public-Sector Reforms*. Teoksessa Wollman, Hellmut (toim.) *Evaluation in Public-Sector Reform: Concepts, and Practice in International Perspective*. Cheltenham & Northampton: Edward Elgar, s. 209-230.
- Tsoukas, Haridimos, Papoulias, Demetrios P. 2005. *Understanding Social Reforms: A Conceptual Analysis*. Teoksessa Haridimos Tsoukas, *Complex Knowledge: Studies in Organizational Epistemology*. Oxford: Oxford University Press, s. 165-180.
- Tsoukas, Haridimos, Vladimirov, Efi (2005). *What is Organizational Knowledge?* Teoksessa Haridimos Tsoukas, *Complex Knowledge: Studies in Organizational Epistemology*. Oxford: Oxford University Press, s. 117-140.
- Töttö, Pertti 2003. *Syvälistä ja pinnallista. Teoria, empiria ja kausalisuus sosiaalitutkimuksessa*. Tampere: Vastapaino.
- Virtanen, Petri 1998. *Refleksiivinen evaluaatio ja tiedon hyödyntäminen*. *Hallinnon tutkimus* 4 (17), s. 262-271.
- Virtanen, Petri 2004. *The Rise and Fall of Evaluation Standards*. *Hallinnon tutkimus* 2 (23), s. 16-29.
- Virtanen, Petri, Laitinen, Ilpo 2004. *Beyond Evaluation Standards? European Journal of Spatial Development* (<http://www.nordregio.se/EJSD>). ISSN 1650-9544-Refereed Articles Octo 2004 No. 13.
- Wollmann, Hellmut (toim.) 2003a. *Evaluation in Public-Sector Reform: Concepts, and Practice in International Perspective*. Cheltenham & Northampton: Edward Elgar.



Wollmann, Hellmut 2003b. Evaluation in Public-Sector Reform: Towards a 'Third Wave' of Evaluation? Teoksessa Wollman, Hellmut (toim.): Evaluation in Public-Sector Reform: Concepts, and Practice in International Perspective. Cheltenham & Northampton: Edward Elgar, s. 1-11.

## VIITTEET

<sup>1</sup> Tässä artikkelissa käytetään systemaattisesti arviointitutkimuksen käsitettä tietäen, ettei se ole kokonaan ongelmatonta. Artikkelin teemojen kehittäminen perustuu ideoihin, jotka on sisällytetty artikkelin kirjoittajan työstämiin jaksoihin laajemmassa käsikirjoituksessa (Ahonen & Virtanen 2005), jota tämän artikkelin kirjoittaja on tehnyt yhdessä Pertti Ahosen kanssa loppusyksystä 2004 alkaen. Kyseinen käsikirjoitus julkaistaan monografiana vuoden 2005 aikana. Kentän käsite ei tämän artikkelin käyttötarkoituksessa eikä edellä mainitussa käsikirjoituksessa ole mitenkään ylimalkainen, vaan sillä on hyvinkin tarkasti mietitty käyttötarkoitus. Kentästä puhuttaessa viitataan tässä yhteydessä vuonna 2002 edesmenneen maailmankuulun ranskalaisen yhteiskuntatieteilijän, Collège de Francen sosiologian professori Pierre Bourdieun kenttäteorian käsitteeseen ja siihen läheisesti linkittyviin konsepteihin, kuten esimerkiksi pääoma ja habitus. Tekijä haluaa kiittää Mikko Wennbergiä ja Ville Valovirtaa artikkelin kriittisestä kommentoinnista, Hallinnon Tutkimuksen anonyymiä referaattia myötäsukaisesta suhtautumisesta ja Pertti Ahosta ajatuksellisista virikkeistä viime vuosien aikana.

<sup>2</sup> Kuriositeettina voi mainita, että arviointitutkimuksen yksi arvostetuimmista ja siteeratuimmista vaikuttajista - Robert E. Stake - julkaisi vuonna 2004 teoksen Standards-Based & Responsive Evaluation (Stake 2004), joka on oikeastaan hyvä esimerkiksi 1990-luvulla alkaneesta alan yleisesitysten tuottamisen perinteestä. Siinä sanotaan olemassa olevasta oikeastaan kaikki, mutta ei juuri mitään uutta.

<sup>3</sup> Tätä näkökohtaa ei saa sekoittaa siihen, mistä Guba ja Lincoln (1989) puhuvat arvostellessaan manageriaalista ulkokohtaista ja erityisesti niin sanottua kolmannen sukupolven evaluaatiota. Gubaa ja Lincolnia ei hyvällä tahdollakaan voi tulkita niin, että he edustavat jotain sellaista evaluaatiotulkintaa, mihin ei liittyisi kausaliiteetin ajatus. Vaikka he ovat haluttomia tavoitteisiin sidottua arviointia kohtaan, eivätkä suostu määrittelemään edes sitä, mitä he arvioinnilla tarkoittavat, kytkevät he silti tarkastelunsa johonkin inhimilliseen toimintaan ja sen vaikutuksiin, joko aiottuihin tai ei-aiottuihin.

<sup>4</sup> Näistä kysymyksistä kiinnostuneiden kannattaa tutustua esimerkiksi Mohrin (1995) seikkaperäisen esityksen avulla. Samaisesta teoksesta löytyy havainnollinen käsiteanalyysi fyysisen, faktuaalisen ja kontrafaktuaalisen kausaliiteetin luonteesta.

<sup>5</sup> Ei siis esimerkiksi ole aivan puhdas sattuma, että

varsin usein - erityisesti Euroopan unionin varoin rahoituksissa arvioinneissa miltei poikkeuksetta - arvioinnin tilaajat edellyttävät arviointipalvelun tuottajilta varsinaisen alustavan arviointisuunnitelman hyväksymisen jälkeen niin sanotun täsmennetyt työsuunnitelman, joka perustuu arvioinnin tilaajan ja toimittajan väliseen keskusteluun ja yhdessä muodostettuun tulokintaan arviointitehtävästä.

<sup>6</sup> Näin arvioinnin tilaajat voivat saada tietoa kentän markkinatilanteesta, eri yksilö- ja instioutuutiotoimijoiden kyvykkyydestä ja hyvistä ja huonoista suorituksista, vertailevaa evidenssiä eri palveluntuottajien hinnoittelusta, ja niin edelleen. Arviointitoimeksiannon voittaminen kilpailutilanteessa riippuu ratkaisevasti siitä, millainen on arviointitiimin habitus. Joissakin tapauksissa tällä on selvästi suurempi merkitys kuin arviointipalvelun hinnalla, koska tilaajat laskevat usein paljon tiimin uskottavuuden varaan. Myös arviointitutkimusten tekijöiden keskuudessa toisilta oppiminen on oleellinen osa arviointityön tekemistä. Tässä tällä viitataan siihen, että arviointitiimin sisällä "vartuneempien" tekijöiden kokemuksista kannattaa ottaa opiksi (jos voidaan vakuuttua siitä, "nuoremmat" pystyvät suodattaamaan "vartuneempien" pahimmat ennakkoluulot ja kaava-maisuudet pois heidän puheestaan).

<sup>7</sup> Arvioinnit tehdään nykyään poikkeuksetta Suomessa ministeriöiden toimeksiannosta niin, että niissä on monta yksittäistä toimijaa/tutkijaa. Tämä mahdollistaa tämän "yhdessä virheille nauramisen".

<sup>8</sup> Sivumennen sanoen: tällaisia epäilyjä ei juurikaan liity laatupalkintokriteeristöjen mukaisesti tehtyihin itsearviointeihin. Esimerkiksi Suomen laatupalkintokilpailussa arvioinnin suorittajia kouluttaessa korostetaan sitä, etteivät laatupalkintokilpailun hakemuksensa jättäneet organisaatiot ole valehdelleet hakemuksessaan eli toiminnankuvauksessaan. Hakemuksessa esitettävät asiat ovat lähtökohtaisesti siis tosia. Vastaavasti kokemukset laatupalkintojen soveltamisesta - varsinkin harjaantuneiden soveltajien keskuudessa - arviointi suoritetaan hyvin kriittisesti ja on arvioitu, että arvioinnissa ollaan jopa ylikiittäviä omille puutteille ja toimintatapaongelmille. Itsearviointimallien suorittamisesta; tästä ks. Porter & Tanner 2004.

<sup>9</sup> Ei tarvitse olla kovinkaan pitkään mukana julkishallinnon arviointitoiminnan kentällä kun huomaa, miten eri arviointitoimijat - yksilöt ja instituutiot - toimivat maksimoidakseen asemaansa arvioinnin kentällä ja kiillottavat omaa habitustaan. Ei liene kohtuutonta sanoa eikä todennäköisesti ylitä kovin suurta uutiskynnystä väitettäessä, että joillakin arvioinnin tilaajilla on hallinnossa omat "hovihankkijansa", jotka ovat vahvistaneet asemansa tuottamalla tietyn tyyppisiä, arvioinnin tilaajia miellyttäviä tuloksia. Arvioinnin tekijöiden kannalta tilanne ei ole aivan ongelmaton, koska Suomi on pieni maa ja Eurooppakin kohtalaisen pieni maanos, jossa "paha kello kuuluu pitkälle". Näennäinen voitto saat- taakin olla huono saavutus, jos sen hintana on maineen menetys "pitkässä juoksussa". Tämä pohdinta on tietysti kovien väitteiden esittämistä, säännellään-



hän nykyisin julkishallinnon palvelutuotantoa erilaisilla kilpailusäännöksillä. Kilpailuttaminen säädettiin Suomessa julkishallinnossa pakolliseksi vuonna 1994 lailla julkisista hankinnoista. Kilpailuttaminen liittyy paitsi julkisiin palveluihin, myös kaikkiin julkisiin hankintoihin. Suomen julkisten hankintojen mittasuhteesta saa jonkun käsityksen kun todetaan, että sen kokonaisarvo on vuosittain noin 20 miljardia euroa, mikä on noin 15 % bruttokansantuotteesta. Arviointitoiminnan osuus tästä on tietysti aivan marginaalinen, mutta kuitenkin merkittävä, koska yksi yhden ministeriön tilaama kaksivuotinen ohjelma-arviointi maksaa tilaajalleen tavallisesti vähintään 100 000 - 120 000 euroa. Tästä problematiikasta ks. Virtanen 1998.