



Arvioinnin teemanumero 2006

Suomen arviointiyhdistys

www.finnishevaluationsociety.net

Toimittanut Liisa Horelli



Vaikuttavuusarviointi mantrana ja kamppailulajina

Liisa Horelli

Suomen Arviointiseuran vuosikokouksen teemana oli vuonna 2005 vaikuttavuuden arviointi. Perinteeksi muodostuneen tavan mukaan vuosikokouksen aihe jalostetaan Hallinnon Tutkimus-lehdessä ilmestyväksi arvioinnin teemanumeroksi. Erityisnumeron toimittajana jouduin tutustumaan aiheeseen vähän laajemminkin. Hämmennyin vaikuttavuuden arvioinnin massiivisesta levinneisyydestä. Tässä numerossa kirjoittava Marketta Rajavaara (2006) googlasi käsiteparilla "vaikuttavuuden arviointi" ja sai yhteensä 33 900 osuaa. Hänen mukaansa vaikuttavuuden käsitettä käytetään nykyisin kaikkien hyvinvointivaltion toimintojen, palvelujen ja etuuksien ohjauksessa. Vaikuttavuuden arvioinnin kohteena ovat mitä moninaisimmat toiminnot sosiaalityöstä taidemuseo- ja tutkimustoimintaan sekä kielipetukseen ja poliisien päällystökoulutukseen saakka. Vaikuttavuus on nostettu tuottavuuden ja innovaatiokyvyn rinnalla taloudellisen ja yhteiskunnallisen menestymisen kriteeriksi. Vaikuttavuudesta näyttääkin tulleen eräänlainen pakkopaita, jopa mantra. Toisaalta ei kuitenkaan olla varmoja, mitä vaikuttavuus ja sen arviointi merkitsevät. Vaikuttavuuden määritelmiä ja oikeita arviointikäytäntöjä etsitään kuin Graalin maljaa. Myös korkeakoulut ja yliopistot ovat velvoitetut sisäisen arviointinsa lisäksi pohtimaan omaa alaansa laajempia yhteiskunnallisia vaikutuksia (vrt. Korkeakoulujen kolmas tehtävä). Tämän teemanumeron tavoitteena on purkaa auki vaikuttavuuden käsitettä ja vaikuttavuusarvioinnin käytäntöjä sekä pohtia hallinnan ohjauksen ja arvioinnin uusia avauksia.

ARVIINTIYHTEISKUNNAN RAKENTUMI-NEN

Vaikuttavuuden arviointibuumi ei ole syntynyt tyhjiössä. Evert Vedungin (2003) paljon siteeratut arvioinnin aallot kuvaavat hyvin, miten arviointi on vähitellen läpäissyt aikakautemme kulttuurin ja politiikan. Arvioinnin ensimmäistä aaltoa 1960-1970-luvuilla leimasi rationaalinen ja objektiivinen näkökulma yhteiskunnalliseen päätöksentekoon ja suunnitteluun. Poliitikka lähenee tiedettä ja politiikasta haluttiin riisua ns. lyhytnäköinen (puolue)poliittinen pelivara mahdollisimman vähäiseksi. (Ks. Roininen 2006.)

Usko positivistisen tieteen mahdollisuuksiin päätöksenteon ja suunnittelun apuna eli vahvana. Tieteellinen arviointi nähtiin yhtenä osana keskitettyä hallinnointia, mutta toisaalta myös reaktiona sen äärimuotoja vastaan. Päätöksenteon avuksi valjastettiin erilaisia toimenpiteitä, joista Vedung käyttää yhteisnimitystä "politiikka-analyysi" (policy analysis). Siihen kuuluvat menetelmät, kuten järjestelmä- ja kustannus-hyötyanalyysit, loivat perustaa myöhemmälle arviointitutkimuksen metodologiselle kehitykselle. Aallon kärjistynyt malli, julkisten interventioiden insinöörimalli rakentui keskitetyn päätöksenteon ja hierarkkisen täytäntöönpanon perustalle.

1980-luvulla alkanutta toista arviointiaaltoa Vedung kutsuu uudeksi julkisjohtamisen (New Public Management) malliksi. Sen taustalla on uusliberalistinen, markkinalähtöinen aatesuunta ja tieteellisen liikkeenjohdon kautta koko julkisen sektorin toimintaan levinneet ohjaukseen. Radikaalin rationalismin suuri kertomus vaimeni siihen kohdistuneen kritiikin kasvaessa ja epä-

luulo keskitettyä päätöksentekoa ja suunnittelua kohtaan vahvistui. Keskitetyn hallinnon todettiin erkanevan liian kauaksi päätösten täytäntöönpanosta. Poliitikan vahva tieteellistämisyrittäminen väheni ja usko johtamiseen, välilliseen ohjaukseen sekä kuluttaja- ja kansalaislähtöisyyteen kasvoi.

Välillinen ohjaus toi julkiselle sektorille mukanaan mm. tulosyksiköt, tulostimit, suoritusperusteisen palkkauksen, laatujärjestelmät sekä toimintojen ulkoistamisen ja kilpailuttamisen. Aiempaa keskitettyyn hallintoihin liittyvää suoraa ohjausta ja valvontaa puretaan sitä mukaa, kun uudet välilliset informaatio- ja arviointiohjauskeinot valtaavat alaa.

Arviointiohjausta on nykyisin ryhdytty pitämään uudenlaisena julkista hallintoa läpäisevänä ohjausideologiana. Sen katsotaan soveltuvan nykyiseen yhteiskuntaan aiempaa, keskitetysti toteutettua normiohjausta paremmin. Normiohjaus perustui sääddöksiin ja normitettuun resursien jakoon. Arviointiohjaus perustuu sitä vastoin toiminnan laadun ja tarkoituksenmukaisuuden seurantaan sekä tulosten ja vaikutusten arviointiin. (Roininen 2006.)

Koska arvioinnista on muodostunut merkittävä yhteiskunnallinen ilmiö, jolla on syviä ja laajoja vaikutuksia läpi yhteiskunnan ja useiden ihmisten elämään, voidaan Vedungin arviointiaaltojen kuvausta pitää myös selostuksena arviointiyhteiskunnan rakentumisesta (Simola & Rinne 2006). Arviointiyhteiskunta on Rantalan & Sulkusen (2006) mukaan myös läpiprojektoitunut. Uusiksi valtaajien organisaatiomuodoiksi näytävät muodostuneet projektit ja niihin liittyvät arviointiyhteiskunnan ominaisuudet. Niitä ovat ohjelma- ja suunnittelu, tulosjohtaminen, byrokraattisia hierarkioita korvaavat verkostot sekä yhteiskunnallisten vastuurakenteiden hajaantuminen. Siirtyminen keskitetyn vallan ja vastuun normatiivisesta, hyvän elämän ja tasa-arvon politiikasta projektiyhteiskunnan itsevastuullisuuteen on merkinnyt uutta kirjoittamatonta yhteiskuntasopimusta. Sen ytimessä on arviointi ja erityisesti vaikuttavuuden arviointi.

ARVIOINTIYHTEISKUNNAN PAULOISSA

Koulutussosiologit Hannu Simola ja Risto Rinne kyselevät kriittisessä artikkelissaan, miksi arviointiin liittyviä yhteiskunnallisia vaikutuksia on

tutkittu niin vähän, vaikka osa vaikutuksista on silmiinpistävästi haitallisia. Arvioinnista itsestään on tullut osa politiikkaa. He soveltavat analyysissään Peter Dahler-Larsenin (2003) tekemää eritelyä arvioinnin ja politiikan välisistä suhteista. Arvioinnin toimintapolitiikka (evaluation policy) viittaa keskusteluun siitä, mihin arvioinnilla pyritään ja miten se toteutetaan. Arvioinnin politiikka (evaluation politics) liittyy puolestaan poliittiseen valtapeliin siitä, kuka arvioinnista hyötyy sekä kuka saa asettaa arvioinnin kriteerit ja rajat. Arvioinnin poliittisuudessa (the political in evaluation) on kyse siitä, miten arviointi liittyy inhimillisen toiminnan sääntelyyn, merkitysten rakentumiseen, käsitykseemme itsestämme yhteiskunnallisina toimijoina ja identiteettimme rakentumiseen. Toisin kuin edelliset, viimeksi mainittu on vain vähäisessä määrin noussut merkittäväksi osaksi tutkimusta, varsinkin Suomessa.

Kirjoittajien mukaan yhteiskunnalliset sivuvaikutukset eivät näytä olevan kenenkään intressissä. Arvioinnin tilaajat keskittyvät hallinnoimaan arviointitekniikkaa, kun taas arviointiammattilaiset korostavat arvioinnin menetelmällisen osaamisen tärkeyttä. Mutta mitä nuo sivuvaikutukset sitten ovat? Niitä ovat muun muassa lisääntyneet kustannukset ja byrokratia, luottamuksen rapautuminen, monet siviiliuhrit erityisesti korkeakoulumaailmassa sekä uuden "auditoitavan" subjektin sosiaalinen rakentuminen. Viimeistään tässä vaiheessa saa Simolan ja Rinteen foudcaultlainen analyysi jatkuvan hallinnon, valvonnan ja arvioinnin katkerista hedelmistä lukijan kysymään ahdistuneena: onko arviointiyhteiskunnasta ulospääsyä? Artikkelin kirjoittajat tarjoavat erääksi pelastuskeinoksi governmentality-viitekehystä käsin tehtyä tutkimusta arvioinnista ja sen vaikutuksista. Governmentality merkitsee Kantolan (2002) mukaan hallinnan tapaa. Tuossa viitekehyksessä on mahdollista kysyä, miten meitä hallitaan ja miten me hallitsemme itseämme. Silloin voivat arviointiin liittyvät alakysymykset olla: millaisia järkevyyden lajeja (episteme), näkyvyyden muotoja (theasis), toiminnan tapoja (techne) ja toimijoiden identiteettejä (ethos) laadunarviointi tuottaa?

JÄRKEILYTTYLIEN KIRJOJA

Marketta Rajavaara esittää artikkelissaan ratkaisun kysymykseen, miksi eksponentiaalisesta

levämisestään huolimatta vaikuttavuudelle ja vaikuttavuuden arvioinnille ei ole yksiselitteistä määritelmää. Hänen mielestään arviointi voidaan jäsentää kahdella tavalla. Toisaalta se voi olla itsenäinen tietokulttuuri tai -käytäntö (evaluation as epistemic practice). Toisaalta arviointi voidaan sisällyttää monenlaisiin tietokulttuureihin julkisessa toiminnassa tai laajemmin yhteiskunnassa (evaluation in epistemic practices). Viimeksi mainittu tuo näkyviin sen, että vaikuttavuuden arvioinnin haasteet ymmärretään toisistaan poikkeavin tavoin eri tietokulttuureissa. Vaikuttavuuden tietokulttuurit ovat todellakin erilaisia paikallisesti, eri tieteenaloilla sekä erilaisissa hallinto- ja asiantuntijaympäristöissä. Tietokulttuurit voivat tukea toisiaan, mutta myös kamppailla keskenään. Lisäksi kunkin tietokulttuurin sisällä vallitsee monenlaisia vaikutustiedon haluja, käytäntöjä sekä käsitteitä pätevästä tiedosta.

Erilaisissa tietokulttuureissa myös järkeillään vaikuttavuutta vaihtelevasti. Rajavaara tarkastelee Ian Hackingin (2002) kehittämän analyttisen työkalun, järkeilytylin (style of reasoning), avulla vaikuttavuuden arvioinnin käytäntöjä. Hackingin mukaan jokaisen järkeilytylin sisällä kehittyvät omat todistamisen ja osoittamisen kriteerit. Ne määrittävät totuusehdot tiedon alalle, jolla niitä sovelletaan. Siten järkeilytylin on enemmän kuin menetelmien kokoelma, jonka avulla hankitaan uusia tosiasioita. Järkeilytylit eivät sinällään ole objektiivisia, vaan ne määrittävät ehdot sille, mikä katsotaan objektiiviseksi. Kun jokainen järkeilytylin luo omat totuuskriteerinsä, niistä tulee itsensä oikeuttavia.

Vaikuttavuuden arviointia tehdään monenlaisen järkeilytylien perustalta. Rajavaara esittelee viisi järkeilytyyliä, nimittäin tavoitteita ja tarpeita mittaavan, kokeellisen ja tieteellisiin vaikuttavuusnäyttöihin perustuvan, auditoivan, mekanismeihin keskittyvän sekä vuorovaikutteisen järkeilytylin. Teemanumeron kannalta on mielenkiintoista, että eri järkeilytylien parissa nimetään tai määritellään vaikuttavuus eri tavoin. Vaikuttavuus voi siten liittyä tavoitteiden saavuttamiseen, tarpeiden tyydyttämiseen, koeryhmän tilanteen muutoksiin, toimintamekanismeihin, yhteistoiminnallisuuteen ja vuorovaikutukseen, systemaattisiin näyttökatsauksiin tai tilivelvollisuuden tietojärjestelmiin.

Mikään edellä kuvatuista ei välttämättä ole ainut oikea, vaan järkeilytylin ja sen piirissä nähty vaikuttavuus voivat vaihdella kontekstista riip-

puen. Artikkelin lopussa kirjoittaja varoittaa ”vaikuttavuussodista”, siis tilanteista, joissa tiedon pätevyys ehdot osoittautuvat keskeiseksi kiistakapulaksi. Sellainen on mahdollista hyvinvointivaltion tietokulttuureissa, jos vaikuttavuustiedolla hallinnointia vielä tehostetaan.

KAKSI ESIMERKKIÄ JÄRKEILYTYYLEISTÄ

Arviointipäällikkö Petteri Paasio esittelee viitekehysten arvioinnista yleensä sekä sen erityisalasta, vaikuttavuudesta. Hän tarkastelee arviointia yleisellä tasolla keskeisenä inhimillisen toiminnan muotona, jossa tehdään arvottavia päätelmiä tiedon ja tosiasioiden perusteella. Arvioinnin ydinkysymys on silloin arvojen merkityksen ymmärtäminen. Erityisviitekehysten tärkein arvo on vaikuttavuuden tunnistaminen toiminnassa. Paasio osoittautuu innokkaaksi mekanismeihin keskittyvän järkeilytylin kannattajaksi. Hän pitää Chenin (2005) vaikuttavuusteoriaa sekä Pawsenin ja Tillyn (1997) kehittämää realistista arviointia lupaavina vaikuttavuuden arvioinnin välineinä. Realistisessa arvioinnissa selvitetään, mitkä vaikutusmekanismit sisältyvät interventioon, miten ne toimivat sekä millaisissa olosuhteissa vaikutukset ilmenevät. Mitä osuvammin intervention taustaoletukset vastaavat yhteiskunnallisia olosuhteita ja kohderyhmän tarpeita, sitä suuremmalla syyllä voidaan odottaa myönteisiä vaikuttavuustuloksia.

Marketta Rajavaara nostaa artikkelissaan esiin realistisen arvioinnin teoriaperusteisen (theory-based) ja Chenin teorialähtöisen (theory-driven) vaikuttavuusarvioinnin eron. Chenin kehittämä arviointiteori on tulevaisuuteen suuntautunut. Siinä painotetaan normatiivisia muutosteorioita. Se sopii erityisen hyvin yhdyskuntasuunnittelun ja osallistumisen arviointiin. Sitä vastoin teoriaperustaisesta, realistisesta arvioinnista puuttuu muutosteorian tarkastelu. Siinä korostetaan pikemminkin toteutuneiden ohjelmien ja politiikkojen mekanismien selvittämistä. Paasion artikkelin sisältämät kaaviokuvat selventävät näitä, harvemmin esiintuotuja eroja.

Juha Koivisto ja Riitta Haverinen esittelevät puolestaan esimerkin tieteellisiin vaikuttavuusnäyttöihin perustuvasta järkeilytylistä. He myös kuvaavat, miten tästä näkökulmasta tehdään vaikuttavuuden arviointi. Näyttöperustaisessa vaikuttavuusjärkeilyssä tehdään systemaattisia

katsauksia vaikuttavuustiedosta erilaisilta toiminnan alueilta. Kyseessä on 1990-luvulta alkanut "evidenssiliiike", joka pyrkii levittämään vaikuttavuustietoon perustuvia toimintakäytäntöjä terveydenhuoltoon, sosiaalialalle, mielenterveytyöhön, kasvatukseen, terapioihin ja kuntoutukseen (Rajavaara 2006). Näyttöperusteisen arviointityön tueksi on perustettu arviointi-instituutioita muun muassa Englannissa, Tanskassa ja Ruotsissa. Artikkelin päätavoitteena on jäsentää ja selkeyttää tutkimuskatsausten mahdollisuuksia vaikuttavuuden arvioinnin välineenä tilanteessa, jossa vaikuttavuuden arvioinnin paradigmoja on useita. Kirjoittajat tarkastelevat erityisesti sitä, missä määrin kokeellista tutkimusta systematisoivat katsaukset tuottavat luotettavaa ja siirrettävää vaikuttavuustietoa sosiaalialalla. Huolellisesti kuvattuina tapaustutkimuksina ovat perheinterventioiden vaikuttavuus sekä ehkäisevien kotikäyntien vaikutukset iäkkäiden toimintakykyyn.

Systemaattisten tutkimuskatsausten tekeminen on runsaasti työtä ja resursseja vaativaa toimintaa. Valitettavasti katsaukset, jotka systematisoivat satunnaistetun koeasetelman perusteella tehtyjä tutkimuksia, näyttävät tuottavan varsin triviaalia tietoa toimenpiteiden vaikuttavuudesta, ainakin sosiaalialalla. Artikkelin kirjoittajat päätyvätkin suosittelemaan moninäkökulmaista otetta katsausten teossa. Tuolloin tulisi keskittyä sekä interventiomalleihin, prosessiin, kontekstiin että tuloksiin. Moninäkökulmaisessa työssä tulisi koota ja systematisoida erilaisten lähestymistapojen, koeasetelmien, realistisen ja mahdollisesti uusien arviointityylien, kuten relationaalisen arvioinnin avulla tuotettua tietoa (Koivisto 2006).

TILAUS UUSILLE AVAUKSILLE

Siirtyminen viranomaisvetoisesta epävarmuuden minimoimisesta hajautettuun ja monimutkaiseen yhteiskuntarakenteeseen on johtanut epäselvyyden lisääntymiseen (Sotara 1996). Epäselvyyttä on myös lisännyt projektiyhteiskunnan toiminnan "itseaiheutettu" monimutkaisuus (Rantala & Sulkunen 2006; Kaakinen 2005). Ongelmat ovat muuttuneet sellaisiksi, ettei niitä kyetä ratkaisemaan osa- tai sektorikohtaisesti, vaan niiden ratkomiseksi on otettava huomioon yhä monimutkaisempia kokonaisuuksia. Lineaarinen, jatkuvaan edistysuskoon liittyvä kehitysnäkemyk, jolla on paljon yhteistä ensimmäisen

arviointiaallon aikana vallinneen radikaalin rationalismin kanssa, on tullut tiensä päähän. Se on korvautumassa evolutionaarilla ja kompleksisuusteoreettisilla (Mitleton-Kelly 2003) lähestymistavoilla.

Dominoivien järjestytylien varassa tapahtuva arviointiohjaus ei näytä toimivan. Lisäksi vaikuttavuusarviointien ylikorostaminen vie huomion pois uusien ohjaustapojen ja niihin liittyvien arviointien kehittämisestä. Muun muassa kompleksisuusteoriaan perustuvat pehmeät strategiat, osallistuvan yhteishallinnan erilaiset muodot ja aikasuunnittelu (Horelli & Wallin tulossa) edellyttävät tuekseen uusia verkostomaisia ja relationaalisia (Koivisto 2006) arviointitapoja. Arvioinnin tietokulttuurien ja järjestytylien välinen kamppailu siis jatkuu ja uudet avaukset tarvitsevat tilaa. Euroopan arviointiseuran tähän asti suurimmassa konferenssissa Lontoossa annetaankin kaikkien kukkien kukkia. Lähes 400 esityksessä, yli 40 maasta, käsitellään eri lähtökohdista "arviointiyhteiskunnan kriittisiä liittymiä" ja etsitään samalla arvioinnin tulevaisuuksia (EES 2006).

Lopuksi haluan kiittää lämpimästi teemanumeron kirjoittajia mielenkiintoisesta yhteistyöstä ja keskusteluista, joita kävimme välillä hyvinkin kiihkeästi. Kiitokset myös referee-lausuntojen antajille, jotka innostivat artikkeleiden tekijät yhä mittavampiin ponnistuksiin.

LÄHTEET

- Chen, Huey-Tsyh: Practical Program Evaluation. Sage Publications, Thousand Oaks, CA 2005.
- Dahler-Larsen, Peter: (2003). Det politiske i evaluering. Studies in Educational Policy and Educational Philosophy, E-Tidskrift, 2 (2003): 1, s. 1-12.
- European Evaluation Society: Evaluation in Society. Critical Connections. Joint International Conference, London, 4-6- October 2006. <http://www.profbriefings.co.uk/EISCC2006> .
- Hacking, Ian: Historical Ontology, Harvard University Press, Cambridge, Massachusetts, London, England 2002.
- Horelli, Liisa & Wallin, Sirkku (Toim.): Arjen ajan hallintaa - kokemuksia suomalaisesta aikasuunnittelusta. Tiede, Helsinki, tulossa.
- Kaakinen, Juha: Arvioinnin paikka. Arvioitsija intressien ristipaineessa. Hallinnon Tutkimus, Arvioinnin teemanumero, 23 (2005): 2, s. 94-99
- Kantola, Anu: Markkinakuri ja managerivalta. Poliittinen hallinta Suomen 1990-luvun talouskriisissä. Viestinnän laitos/Loki kirjat, Helsinki, 2002.
- Koivisto, Juha: Sosiaalialan näyttökeskustelu: miten

- sosiaalisten interventioiden vaikuttavuus osoitetaan? *Janus* 14(2006) 1, s. 53-60.
- Koivisto, Juha & Haverinen, Riitta: Systemaattiset tutkimuskatsaukset vaikuttavuuden arvioinnin välineenä sosiaalialalla. *Hallinnon Tutkimus, Arvioinnin teemanumero*, (2006): 3, s. 108-125.
- Mittleton-Kelly, Eve: Complexity Research - Approaches and Methods: The LSE Complexity Group Integrated Methodology. Teoksessa A. Keskinen, M. Aaltonen and E. (Toim.) *Organisational Complexity*, s.55-74. Finland Futures Research Centre, Turku School of economics and Business Administration, Turku, 2003.
- Paasio, Petteri: Yleinen ja erityinen viitekehys arvioinnista. *Hallinnon Tutkimus, Arvioinnin teemanumero*, (2006): 3, s. 92.107.
- Pawson, Ray & Tilley, Nick: *Realistic Evaluation*. Sage Publications, London 1997.
- Rajavaara, Marketta: Vaikuttavuuden tietokulttuurit. *Hallinnon Tutkimus, Arvioinnin teemanumero*, (2006): 3, s. 81-91.
- Rantala, Kati & Sulkunen, Pekka (toim.): *Projektiyhteiskunnan kääntöpuolia*. Gaudeamus Kirja, Tampere-Paino, Tampere, 2006.
- Roininen, Janne: Näkökulmia arviointiyhteiskuntaan ja sen ohjauskeinoihin. TKK/YTK (Julkaisematon käsikirjoitus), Espoo, 2006
- Simola, Hannu & Rinne, Risto: Koulutuksen laadun arvioinnin yhteiskunnallisten vaikutusten tutkimisesta. *Hallinnon Tutkimus, Arvioinnin teemanumero*, (2006): 3, s. 66-80.
- Sotarauta, Markku: Kohti epäselvyyden hallintaa. Pehmeä strategia 2000 -luvun alun suunnittelun lähtökohdista. *Acta Futura Fennica* No 6. Tulevaisuuden tutkimuksen seura. Gummerus Kirjapaino Oy, Jyväskylä, 1996.
- Vedung, Evert: Arviointialto ja sen liikkeelle panevat voimat. STAKES, FinSoc työpapereita, Helsinki, 2/2003.



Koulutuksen laadunarvioinnin yhteiskunnallisten vaikutusten tutkimisesta

Hannu Simola ja Risto Rinne

ABSTRACT

Researching political effects of quality evaluation in education

The subject of the article is the social and political effects of evaluation, especially those of quality assessment in higher education. From the perspective of the sociology and politics of education, we will ask why studies on extensive and multifarious effects connected to evaluation activities are so scarce and exceptional. Relevant issues to be studied then are, for example, the political character of evaluation, the rise of evaluation and quality professions and the shared interests of key groups in evaluation. It has been claimed that evaluation itself has become a steady part of governance and politics. The concentration on rational goals, sophisticated means and scientifically legitimated techniques of evaluation, seems to make it easy to forget many embarrassing problems that come from the social, cultural and political context of evaluation. This discourse, shared by an epistemic community we call 'The Quality and Evaluation Industrial Complex', could be characterised as 'Wishful Rationalism'. In the final part of the article, we will make a proposal to analyse evaluation effects from the framework of governmentality research. When evaluation is seen as a mode of governance,

we might start asking, for example, what are the rationalities, techniques, identities and visibilities of governance that will be constructed but also destroyed through evaluation and quality assessment procedures. We claim that this kind of approach, might open some analytical and critical understanding of essential societal and political problems, such as how quality evaluation effects will be connected to changes in economical and political structures, as well as to power relations in our society.

Lienee jo yleisesti hyväksyttyä, että sellaiset käsitteet kuin *quality revolution* (Newton 2002), *evaluation industry* (Leeuw 2002) ja *audit explosion* (Power 2003) kuvaavat osuvasti ja liioittelematta uuden laatuparadigman ja laadunarvioinnin asemaa nyky-yhteiskunnassa. Alan pohjoismaisen veteraani Evert Vedung (2003) pitääkin arviointia aikamme 'gigatrendinä', joka läpäisee harvinaisen kokonaisvaltaisesti koko aikakautemme kulttuurin ja politiikan. Arvioinnin kehittäjät kuuluttavatkin jo arviointiaallon 'viidennen sukupolven' esiinmarssia (Guba & Lincoln, 1989; Virtanen, 2001), ja uusi arviointi- ja laatuosaajien' (Uusikylä 1999, 17) ammattikunta on kiistatta muodostumassa. Samalla arvioinnista itsenäisenä tieteenalana omine julkaisu- ja keskustelufoorumineen puhutaan yhä vakavammassa mielessä (Scriven 1991; Teddlie & Reynolds 2000).

Laadussa ja sen arvioinnissa on siis kyse merkittävästä yhteiskunnallisesta ilmiöstä, jolla voi olettaa olevan syvällekyviä ja laajoja vaikutuksia mitä moninaisimmilla alueilla useimpien ihmisten elämässä jälkiteollisissa yhteiskunnissa. Tähänastinen arviointiin ja sen vaikutuksiin kohdistunut tutkimus on kuitenkin pääosin keskittynyt tarkastelemaan ilmiötä arvioitsijoiden ja arvioinnin tilaajien näkökulmasta. Vaikka esimerkiksi vaikuttavuustutkimuksissa varsin usein muistetaan mainita tarkoittamattomista seuraamuksista tai ei-aiotuista ja sivuvaikutuksista, on fokus lähes poikkeuksetta arvioinnille asetettujen tavoitteiden toteutumisen tai toteutumattomuuden tarkastelussa (ks. esim. Rajavaara 2006; Leeu 2002). Voinee sanoa, että kyse arviointiin liittyvässä tutkimuksessa on ollut pikemminkin siitä, miten tutkimus palvelee arviointia, sen tilaajia ja suorittajia kuin siitä, mitä tutkimus voisi sanoa arvioinnista yhteiskunnallisena ilmiönä. Tutkimuksen ensisijainen tähtäin on ollut tuottaa 'policy-relevanttia' tietoa tavoitteiden saavuttamisesta; tutkimus on ollut pikemminkin *for evaluation* ja *on evaluation*.

Peter Dahler-Larsenin (2003; ks. myös Segerholm 2003) tekemä erittely arvioinnin ja politiikan suhteista on tässä mielessä valaiseva. Arvioinnin toimintapolitiikalla (*evaluation policy*) tarkoitetaan keskustelua siitä, mihin arvioinnilla pyritään ja miten se toteutetaan. Arvioinnin politiikka (*evaluation politics*) puolestaan liittyy poliittiseen valtapeliin siitä, kuka mitään arvioinnista hyötyy ja kuka saa asettaa arvioinnin kriteerit ja rajat. Arvioinnin poliittisuudessa (*the political in evaluation*) on kyse laajemmista ja syvemmistä yhteiskunnallisista ilmiöistä: siitä, miten arviointi liittyy inhimillisen toiminnan sääntelyyn, merkitysten rakentumiseen, käsityksemme itsestämme yhteiskunnallisina toimijoina ja identiteettimme rakentumiseen. Arvioinnin toimintapolitiikasta on tutkimusta tehty eniten ja arvioinnin politiikasta jonkin verran. Niistä on käyty myös julkista keskustelua - niin kansainvälisesti kuin Suomessa. Arvioinnin yhteiskunnallisten vaikutusten ja seuraamusten tutkimus ei sen sijaan näytä nousseen missään merkittäväksi osaksi tutkimusta¹.

Mistä puhumme kun puhumme arvioinnista, erityisesti laadunarvioinnista? Arvioinnin käsitteellinen kenttä on mahdollisimman sekava. Tampereen yliopiston finanssihallinnon ja julkisyhteisöjen laskentatoimen professorin Pertti Ahosen (1998, 257) mukaan arvioinnin käsit-

teen ala on kasvamistaan kasvanut ja samalla käsitteen merkityssisältö on hämärtynyt. Arviointi on ahmaissut muiden käsitteiden alaa, muun muassa seurantaa, suunnittelua, laadun kehittämistä ja varmistamista sekä myös tarkastusta. Ahosen (1996, 154) määritelmän mukaan "arviointiksi on syytä määritellä se, ja vain se, jota tällä sanalla on tarkoitettu ja nyt ja vastaisuudessa tarkoitetaan niissä käyttöyhteyksissä, jossa sana on esiintynyt ja esiintyy." Laadunarvioinnin voi sanoa kuuluvan vakiintuneessa jaottelussa pääosin tilivelvollisuusarviointiin (Rajavaara 206, 74), mutta käsitteet voivat olla tärkeämpiä sen vuoksi, mitä ne tekevät kuin siksi, mitä ne merkitsevät (ks. emt., 14; Rose 1999b, 9). Arvioinnin käsitteistön moniselitteisyys, epämääräisyys ja amebamaisuus ei ole oikeastaan ongelma, vaan olennainen osa sitä itseään (ks. Power 1999).

Tässä artikkelissa tarkastelemme arvioinnin ja eriytyneemmin laadunarvioinnin yhteiskunnallisia vaikutuksia korkeakoulutuksessa². Kohteena on siis arvioinnin jatkuvasti laajentuvan spektrin yhden 'lajityypin' vaikutukset erällä julkishallinnon alueella. Pyrimme koulutussosiologisesta ja -poliittisesta näkökulmasta pohtimaan ensinnäkin sitä, miksi arviointiin liittyviä laajempia yhteiskunnallisia vaikutuksia - tai edes arvioinnin 'sivuvaikutuksia' - on tutkittu niin vähän. Tässä tarkastelemme erityisesti arvioinnin tilaajien ja tuottajien intressejä: kyse on arvioinnin poliittisesta luonteesta, arviointiprofession noususta, mutta myös ja erityisesti arvioinnin avainryhmien intressejä palvelevasta puhetavasta. Toiseksi hahmottelemme arvioinnin yhteiskunnallisiin vaikutuksiin liittyviä tutkimustehtäviä, jotka ensiarvoisesti tärkeydestään huolimatta ovat jääneet tai on jätetty tähänastisen tutkimuksen katvealueelle. Perinteisen arviointitutkimuksen näkökulman lisäksi nostamme esiin hallinnan tavan tutkimuksesta nousevan mahdollisuuden jäsentää tutkimustehtäviä.

LAADUNARVIOINTI JA POLITIIKKA

Arvioinnissa, mutta aivan erityisesti laadunarvioinnissa tai 'laadunvarmistuksessa' ja sen 'auditoinnissa' sekä tämän pohjalta tapahtuvissa 'akreditoinneissa', on siis kyse politiikasta, erityisesti hallintapolitiikasta. On ongelmallista - tai paremminkin peräti naiivia ja harhaanjohtavaa - puhua laadusta ja arvioinnista ilmiönä 'sinänsä'

tai 'itsenään', abstraktioina, irrallaan historiallis-yhteiskunnallisesta ja erityisesti poliittisesta kontekstistaan, ikään kuin olisi kysymys neutraaleista teknis-rationaalisista kehittämishankkeista. Arvioinnit perustuvat aina arvoihin, niitä mittaviin kriteereihin ja arvojen järjestykseen.

Kun tänään puhutaan laadusta tai arvioinnista, on kyse aivan eri asioista kuin pari vuosikymmentä sitten. Viimeistään 1980-luvun loppupuolelta lähtien laatu ja arviointi on kiinnitettävä sellaisiin syvästi poliittisiin käsitteisiin kuin esimerkiksi arvioiva valtio, auditointiyhteiskunta, uusi julkisjohtaminen, kilpailuvaltio ja koulutus pääoman korkeimpia linnakkeita ja kiihtyvää osaamiskilpailua ajatellen akateeminen kapitalismi.³ Laatu ja arviointi sijoittuvat nykyisissä hallintajärjestelmissä aivan toiseen paikkaan ja niillä on aivan toisenlainen tehtävä kuin aikaisemmin. On ajateltavissa, että 'Vanhan liiton' maailmaa hallittiin byrokratian, professionalismin ja tasa-arvon eetoksen kautta. Siinä laadulla ja arvioinnilla oli vähäinen rooli.⁴ 'Uuden liiton' maailmaa toimii puolestaan markkinakurin, managerivallan ja erinomaisuuden eetoksen voimin. Siinä laatu ja arviointi ovat uuden hallinnan tavan ytimessä.⁵ Niistä on tullut hallinnon tukijärjestelmä ja kiinteä osa organisatorisen valvonnan ja hallinnan rutinoitua regimiä⁶. Arvioinnista on itsestään tullut olennainen osa politiikkaa ja hallintaa.

Uudesta julkisjohtamisesta (NPM) ja uudesta managerialismista käytävät keskustelut ovat hallinneet julkisen hallinnon uudistamisesta käytävää keskustelua maailmalla jo ainakin parin viime vuosikymmenen ajan. Ne nousivat hallintoreformistiseen keskusteluun viimeistäänkin silloin, kun OECD ja sen jäsenvaltioiden hallitukset törmäsivät laskeviin talouslukuihin ja julkisten palvelujen kasvaviin vaatimuksiin. Uuden julkisjohtamisen avulla nostettiin tulevaisuuden avaimiksi visio julkisen sektorin johtajista yrittäjinä, jotka veisivät kansakunnat uuteen yksityisen sektorin mallintamaan parempaan maailmaan. Siinä eivät muuttuisi ainoastaan hallinnan vanhat käytännöt, vaan niiden tilalle rakentuisivat myös kokonaan uudenlainen liike-elämän arvomaailma ja kulttuuri. Julkisen valinnan teorian mukaisesti uskottiin, että kaikkea inhimillistä toimintaa motivoi ensisijassa oma etu ja että markkinakilpailu julkisellakin sektorilla tuottaisi keskitettyä byrokrattista hallintaa paremmin tulosta tehdyille investoinneille.

Tosiasiassa suuri osa NPM-liikettä on ollut

pikemminkin managerialistista ideologiaa kuin harkittua ja kontrolloitua uudistusten todellisia vaikutuksia tarkastelevaa strategiaa. Lisääntyvässä kovassa kritiikissä on myös suoraan kysytty, mitä uutta lopultakaan uusi hallintatapa tuo, kun käyttäytymismittausta kehiteltiin jo 1910-luvulla, suoritusbudjetointia 1950-luvulla, tulosjohtamista 1960-luvulla ja yksityistäminen oli julkisenkin hallinnon työkalupakissa jo 1500-luvulla. Kriitikot tahtovat samalla muistuttaa, että julkinen yrittäjä ja 'uusmanagerialismi' pikemminkin uhkaavat sellaisia demokraattisia ja perustuslaillisia arvoja kuin reiluus, oikeus, edustuksellisuus ja osallistuminen. (Meek 2003, 8-10).

Etenkin yliopisto- ja korkeakoulutuksen kentälle on ulottunut 'laatuvalvankumoukseksi' luonnehdittu toimintapolitiikka. Sen perusajatuksena on pitää tutkimusta ja koulutusta markkinatavarana, joka ei olennaisesti eroa muista markkinoilla kaupan olevista palveluista ja hyödykkeistä. Tämä ajattelu kiteytyy Kansainvälisen kauppajärjestön (WTO) jäsenten keskinäisessä GATS-sopimuksessa. Ydinkäsitteiksi ovat muotoutuneet tilivelvollisuus (*accountability*), arviointi (*evaluation*) ja laadunvarmistus (*quality assurance*). Nämä käsitteet tulevat suoraan talouselämästä ja tavaratuotannosta, joissa niitä on myös enemmän tai vähemmän menestyksellisesti sovellettu. Kyse on siis ensisijaisesti korkeakoulutukseen kytkettyyn inhimilliseen pääomaan kiinnittyvästä globaalista talous- ja markkinapolitiikasta.

Suomalaisen korkeakoulutuksen laadunarvioinnin yksi paljon puhuva hallinnollis-poliittinen perusteksti on Opetusministeriön muistio *Korkeakoulutuksen laadunvarmistus*⁷. Sen esitysten mukaan Suomen yliopistoissa ollaan kauttaaltaan rakentamassa paraikaa uusia laadunvarmistuksen mekanismeja. Muistiossa ei rivien väleissä eikä oikeastaan riveillääkään pyritä kieltämään laadunarvioinnin poliittista olemusta. Muistion mukaan paine korkeakoulujen laadunvarmistukseen tulee pääosin yliopistojen ulkopuolelta ja liittyy visioon, mielikuvaan, utopiaan eurooppalaisista korkeakoulumarkkinoista, joiden vaatimuksiin kansallisten yliopistojen on auditoinnillaan vastattava. Takana on Euroopan Unionin konsepti - ei siis mikään, joka on, vaan jokin sellainen, joka on tuleva ja poliittisen tarpeen mukainen.

ARVIOINTI VASTAUKSENA HALLINNOITAVUUDEN ONGELMAAN

Nikolas Rose (1999b) korostaa sitä, että vasta hallinnoitavuuden mahdollisuus antaa hallituksille vallan hallita. Tämä on kautta historian edellyttänyt hallitsijoita ja hallituksia luomaan ja pystyttämään lukemattomia hallitsemisen tekniikoita, strategioita, laskelmia ja arvioita, joiden avulla tuhansia hajautuneita paikkoja ja asioita voidaan luokitella, laskea, järjestää, arvioida ja hallita. Näin on voitu saattaa voimaan perustuslaillinen, taloudellinen, organisatorinen, juridinen ja ideologinen valta, jolla sitten on voitu arvioida, säädellä ja hallita taloudellista elämää ja instituutioiden toimintaa, väestön terveyttä ja tapoja tai vaikkapa rahvaan säädyllisyyttä ja siveellisyyttä. (mts. 17). Kärjekkäimmillään tätä samaa kuvaa Michel Foucault (1980) esimerkiksi kirjassaan *Tarkkailla ja rangaista* Benthamin Panopticonin käsitteen avulla. Kyseessä on jatkuvan hallinnon, valvonnan ja arvioinnin läpäisemä arkkitehtoninen rakenne, jossa kaikkia yksittäisiä kansalaisia ja heidän toimintaansa, käyttäytymistään ja aikaansaannoksiaan voidaan alati tarkkailla ja arvioida. (Rinne 2001, 106).

Alkuperältään taloudellisen laskennallisuuden ja mitattavuuden periaatteet siirrettiin myös sellaisille alueille, joita olivat aiemmin säädelleet vanhat virkavaltaiset säädökset sekä professionaaliset normit. Aiemmin voittoa tavoittelemattomista organisaatioista, kuten esimerkiksi yliopistoista ja sairaaloista, alkoi muokkautua pikkuyrityksiä, joiden tuloksia alettiin arvioida ja mitata erinäköisin indikaattorein. Rose (1999b, 152) kutsuu julkisen sektorin uutta tulosvastuullisuuden ja arvioinnin hallintateknologiaa 'etähallinnaksi' (*governing at distance*). Uudella hallinnan teknologialla "aiemmin abstraktit tilat tehtiin materiaalisiksi organisaation tilan fysikaalisella uudelleentuotteistamisella", mikä sitten sai muotonsa uusissa nimeämisissä, uusissa budjettimomenteissa, uusissa arviointi-indikaattoreissa, uusissa menestyskäyrissä ja koko uudessa julkisessa manageriaalisessa kulttuurissa.

Kukin uusi arvioinnin ja mitattavuuden alle tuleva tila kutsuu asujaimistoaan arvioimaan ja mittaamaan itseään, kääntämään toimintonsa mitattaviksi ja taloudellisiksi termeiksi maksimoimaan tehokkuuden ja tulot, leikkaamaan tuhlauksen, uudelleenmuokkaamaan kustannustehottomat toimet. Mielivalta tulee näin kesy-

tetyksi, liberalisoiduksi sekä tunnistetuksi neutraalina ja objektiivisena laskentana ja arviointina (Rose 1999b, 152-154; Rose & Miller 1992). Rosen mukaan tätä kautta on siirrytty 'tarkastajayhteiskuntaan' (*audit society*). (Vrt. Rinne 2001, 107).

Samaa tarkastaja- tai auditointiyhteiskunnan ideaa on kehitellyt Michael Power (1994; 1999; 2003). Hänen mukaansa tarkastukset ovat alleviivatuksi tulleet korvaamaan sen luottamuksen, jota hallitukset aikanaan tunsivat professioiden ja asiantuntijoiden viisautta ja osaamista kohtaan. Tämä murtuma näkyy Powerin mukaan yhtäläillä niin kouluissa, sairaaloissa, yliopistoissa kuin laajemminkin yrityksissä. Power katsoo, että arvioinnissa on tavallaan kyse 'kontrollin kontrollista' ja 'todentamisen rituaaleista'.

Vedung (2004) näkee Max Weberiin vedoten arvioinnin olevan uusi vaihe länsimaisen yhteiskunnan etenevässä modernisoinnissa ja rationalisoinnissa. Se on osa tieteellisen ja rationaalisen laskennallisuuden esiinmarssia. Arvioinnin pitkä perinne voidaan Dahler-Larsenin (2000) mukaan jaotella kolmeen vaiheeseen. Ensimmäistä vaihetta voidaan kutsua rationaalis-objektiiviseksi vaiheeksi. Tuossa vaiheessa korostettiin arvioinnin tieteellistä objektiivisuutta ja tehtiin tarkkaa työnjakoa yhtäältä arvioinnista intresseeratun poliittisen päätöksenteon ja toisaalta varsinaista arviointia suorittavan positivistisen tieteen välillä. Tuloksena oli usein "pölyisiä arviointiraportteja tiukkoine tieteellisine mittauksineen ja poliittisesti irrelevantteine löydöksineen..." (mts. 72). Vedung (2004) näkee tämän vaiheen lähinnä ns. insinöörimallin vaiheena. Toista vaihetta voidaan kutsua subjektiivis-argumentoivaksi vaiheeksi. Tässä vaiheessa korostettiin arvioinnin subjektiivista ja samalla myös poliittista luonnetta. Nähtiin, että kaikkea arviointia ympäröivät useat kilpailevat intressit. Tuloksena oli paljon ristiriitoja ja pyrkimys välttää arviointien väärinkäyttöä ja vääränlaisia arvioitsijoita. Kolmas vaihe on kuitenkin nykyisen arviointikompleksin ymmärtämisen kannalta oleellisin. Tässä vaiheessa arvioinnit integroitiin osiksi rutinoituja organisoidun ohjauksen ja kontrollin regiimejä. Arvioinneista tuli osa 'normaaleja' hallinnollisia prosesseja.

Kolmas vaihe kytkeytyi uuden julkisjohtamisen läpimurtoon. NPM luonnollistui tietynlaiseen arviointiin perustuvan tieteellisen liikkeenjohdon kautta koko julkisen sektorin toiminnan malliksi. Käyttämisen, laadun, tehokkuuden ja asia-

kastytyväisyyden arviointi arkipäiväistyi rutiiniksi osana uutta tavoite- ja tulosjohtamisen luonnollista kulttuuria. Tämä on vaihe, jossa samalla perustettiin yhä uusia arvioinnin julkisia instituutioita, kuten Suomessa mm. Korkeakoulujen arviointineuvosto ja Koulutuksen arviointineuvosto. Samanaikaisesti arviointiseurojen ekspansio lähti käyntiin.

Englantilaisen koulutuspolitiikan tutkijan Stephen J. Ballin (2001) mukaan koulutuksen laadunarviointi ja koko arviointipuhe kytkeytyvät osaksi ylikansallista ja globaalia uutta kulttuuria. Se samoin kuin puhe ”oppivasta yhteiskunnasta” tai ”tietoon perustuvasta taloudesta” ovat osa uutta konsensusta ja sen rakentamista. Puhe palvelee ja symboloi hänen mukaansa talouspolitiikan määräävää asemaa ja kasvavaa kolonialisaatiota koko koulutus- ja kasvatuspolitiikassa. Se ovat osa uutta politiikkakehystä ja ”politiikkaepidemiaa”, joka valtaa alaa kansallisvaltioiden rajojen yli. Tämän politiikkaepidemian ainutlaatuisuus on siinä, että ”se ei muuta ainoastaan sitä, mitä ihmiset kouluttajina, oppineina tai tutkijoina tekevät, vaan se muuttaa sen, mitä ihmiset ovat.” (Ball 2003, 215). Kyse on pitkälti samasta asiasta, mistä Richard Sennett (1998) puhuu ’luonteen korroosiona’ ja Power (1994) ’epäluotattamuksen taantumana’.

Ballin (2001; 2004) mukaan uuteen epidemiaan kytkeytyvät keskeisesti politiikan teknologian uudet elementit, kokonainen uusi paradigma. Näitä elementtejä ovat markkinamuotoisuus, managerialismi ja suorituskeskeisyys. Uusi markkinamuotoisuus tuottaa koulutuksen kentälle uuden moraalisen ympäristön ja laajemmin tuota kenttää läpäisevän uuden sivilisaation muodon. Hänen mukaansa se synnyttää ’itsekkyyden kulttuuria’, jossa henkilökohtaiset pyrkimykset ja motiivit sekä oman edun tavoittelu ovat universaaleja arvoja tärkeämpiä. Hajautetuilla markkinoilla pragmatismi ja itsekkyyden nähdään luonnollisina ihmisten välisten suhteiden jäsentäjinä. Managerialismin kautta julkisista koulutuspalveluista poistetaan perinteiset professionaalisen kiinnittymisen muodot ja tilalle tuodaan uudenlainen liikkeenjohdollinen yrityskulttuuri. Suorituskeskeisyydessä taas työntekijöiden ja organisaation suoritukset pannaan näyttille ja tuodaan jatkuvan arvioinnin ja vertailun alle. Jotta voitaisiin tulla arvioiduksi, organisaation täytyy aktiivisesti muuntaa itsensä arvioidavan hyödykkeen muotoon, pyrkiä patologisesti luomaan pelejä, joilla

vastataan indikaattorikulttuurin vaatimuksiin (Ball 2003, 224-225). Yksilöiden ja organisaatioiden ’näytökset’ alkavat toimia tuottavuuden ja suorituskyvyn mittareina ja laadun näytteillepanon hetkinä. (Ball 2001, 30-36; 2003)

ARVIOINTIPROFESSION INTRESSIT

Arviointielinkeinon ja -teollisuuden⁸ kenttä on voimakkaasti ylikuumentunut ja professionalistien intressien hallitsema. Tämä johtunee pääosin politiikan lupaamista lähes rajattomista liiketoimintamahdollisuuksista. Ahosen (2001, 103) mukaan arvioinnin kenttää luonnehtiikin vakava jakautuneisuus. Ahonen kirjoittaa:

”Eri tahot pyrkivät paaluttamaan evaluaation kenttää, jotta kullekin niistä muodostuisi monopolietuja antava läänitys ajatellen sitä, mitä evaluaatioon on legitiimiä lukea. Olipa kyse evaluaatioon sovelletuista lähestymistavoista ja metodeista tai niistä tarkoituksista, joissa evaluaatiota harjoitetaan, paalutuksia nousee paalutusten jälkeen ja monopoli- ja reviiritaisteluita puhkeaa, raivoaa mutta lopulta laantuu.”

Kyse on tietenkin nousevan profession - arvioinnin, hallinnan ja suunnittelun ammattilaisten, eräänlaisen laadunhallinnan ammattikunnan - pyrkimyksistä ajaa etujaan. Heihin kuten muihinkin professioihin sopii mainiosti yhdysvaltalaisen sosiologin Randall Collinsin (1990) luonnehdinta professioista, ’idealisoituista ammattikunnista’, jotka ilmoittautuvat yksinoikeudella ratkomaan ’itseluotuja ongelmia’ (*self-created problems*). Näitä ei olisi sellaisenaan olemassa ilman heidän olemassaoloaan: ihmiset kyllä riitelisivät keskenään, mutta vasta juristien ammattikunnan, oikeuslaitoksen ja oikeustieteen synty teki riidoista juridisia; ihmiset kyllä oppisivat ja olisivat oppimatta kaikenlaista, mutta vasta opettajien ammattikunnan, koulun ja kasvatustieteen synty teki ongelmasta pedagogisen.

Myös koulutuksen kentän hallinnasta käytävä reviiritaistelu on olemukseltaan ja vaikutuksiltaan syvästi poliittista. *Quality in Higher Education* -lehden päätoimittaja Lee Harvey (2004), laatu- ja arviointimies *primus inter pares*, yhdisti vakuuttavasti poliittisuusteesin ja professionalismiteesin runsas vuosi sitten ilmestyneessä artikkelissaan *Maailmojen sota: kuka voittaa taistelun laatuheruudesta?* Harvey pohtii artikke-

lissaan ehdotetun ylikansallisen World Quality Register'in pystyttämisen motiiveja ja vaikutuksia käyttäen sellaisia luonnehdintoja kuin laadunvarmennusprofession intressit, korkeakoulutuksen viejämäiden intressit ja laatuimperialismi. Hän pelkää, että kansainvälisessä katsannossa laadunvarmistuksessa yhtyvät arviointiteollisuuden professionalistinen projekti ja korkeakoulutuksen vientikaupan vapautusliike. Tämä merkitsisi hänen mukaansa sitä, että yliopistokoulutuksen ydin, opiskelijoiden oppiminen jää auttamattomasti managerialistisen politiikan, uusliberalistisen talouden ja professionalististen intressien jalkoihin.

Yksi selvimpiä seuraamuksia uuden julkishallinnon (NPM) voimistumisesta yliopistoissa näyttää olevan vahva hyökkäys professioiden ja etenkin akateemisten professioiden kimppuun. Ne ovatkin tämän seurauksena menettäneet prestiisiään ja sosiaalista asemaansa. Voidaan kenties puhua suorastaan akateemisten professioiden asteittaisesta proletarisoitumisesta ja autonomisen aseman vahvasta heikkenemisestä. Hallitsevaa asemaa heidän rinnallaan ja ylitseen on ottanut 'uusi manageriaalinen luokka', joka uudella julkishallinnan politiikallaan on kerännyt valtaa itselleen. (Amaral, Fulton & Larsen 2003, 284-285).

Niinpä kun luemme EAQAn (European Association for Quality Assurance in Higher Education) laatukäsikirjaa *Standards and Guidelines for Quality Assurance in the European Higher Education Area*⁹ tai sen suomalaista vastinetta, KKAN auditointikäsikirjaa *Korkeakoulujen laadunvarmistusjärjestelmien auditointi*¹⁰, luemme tekstiä, jonka yksi olennainen funktio on tehdä itsensä tarpeelliseksi eli oikeuttaa ja vahvistaa ammattikunnan asemaa ja prestiisiä. Sama koskee myös paikallisia erittelyjä ja toimenpideohjelmia, esimerkiksi HY:n laitosjohtajat keväällä 2006 muutosvastarintaan lietsonutta asiakirjaa *Laadunvarmistuksen nykytila Helsingin yliopistossa*¹¹. Kun viimeksi mainitussa dokumentissa todetaan, että laadunvarmistus kattaa "yliopiston koko toiminnan", se on "jatkovaa" ja että kaikesta ("toiminnoista, tavoitteista, vastuuta-hoista ja -henkilöistä, dokumentoinnista, laadunvarmistusmekanismista ja laadun ilmaisimista") on oltava dokumentit, on kyse myös siitä, että tietty ammattikunta paaluttaa monopoliaan.

LAADUNARVIOINTI JA TAVOITERATIONAALINEN HARHA KOULUTUKSESSA

Jos haluamme ymmärtää arvioinnin kentän sokeita pisteitä, saattaa olla hyödyllistä tehdä ekskursio koulureformien historiaan. Kaikista yhteiskunnallisista instituutioista juuri koulutuksessa voidaan sanoa arviointia toteutetun kaikkein pisimpään. Oppilaiden oppisuoritusten arvioinnin perinteet yltävät yhtä pitkälle kuin institutionalisoitunut koulutuskin. Juuri koulusaavutusten tutkiminen ja testaaminen muodostavat arvioinnin ensimmäisen sukupolven ydintoiminnan (Callahan 1963; Valtonen 1997). Toisaalta arvioinnin uudenlainen tuonti koko koulutuksen kentälle¹² on nähtävissä koulutuksen hallinnan reformina. Sen avulla pyritään saamaan aikaan tavoitteiden suuntaista muutosta kouluinstituutiosta ja lisäämään toiminnan vaikuttavuutta tuloksissa.

Yhdysvaltalaiset koulutushistorioitsijat David Tyack ja Larry Cuban (1995) toteavat, että yksi koulureformien tutkimuksen lopputulemista on se, että tärkeämpää olisi tutkia sitä, miten koulut muuttavat reformeja kuin sitä, miten reformit muuttavat kouluja. Tämä johtuu osittain siitä, että kentän toimijoilla, opettajilla 'katutason byrokratteina' on paljon suurempi vaikutus reformien lopulliseen toteutumiseen kuin yleensä uskotaan. Toisaalta taustalla on koulun institutionaalisen kontekstista nouseva tekijä, jota Tyack ja Cuban (emt.) kutsuvat "koulunpidon kieliopiksi". Samaa säätelevää tekijää on kutsuttu myös mm. koulun kehystekijöiksi (esim. Lundgren 1972) tai koulun käytäntöjen - ajan, tilan ja rituaalien järjestelmäksi (Rinne et al 1984). Nämä tekijät pitkälti muodostavat reunaehdot reformien läpiviennin mahdollisuuksille. Olennaista on, että reformin vaikutuksista ei voi puhua ottamatta huomioon näitä institutionaalisia kehystekijöitä. Niinpä lukemattomia kertoja on voitu todeta, että uudistukset kyllä muuttavat kouluja, mutta harvoin niin kuin on ajateltu: "reformit eivät pelkästään määrittelellä johda tarkoitettuihin tuloksiin" (Pressman & Wildavsky 1974; vrt Cerych & Sabatier 1986).

Miten on sitten ymmärrettävissä, että näin olennaiset tekijät ovat jääneet niin vähäiselle huomiolle reformien tutkimuksessa? Koulureformien kohdalla yksi selitysyrittäminen on, että reformien vääjäämättömän epäonnistumisen ja uusien reformien jatkuvan tarpeen taustalla toimii erityinen reformidiskurssi, jota voidaan kutsua toiveajatte-

lun rationalismiksi (Simola 2000, 2002).

Koulutusala sitkeästi hallinneessa ajattelussa, nk. *Tyler Rationalessa*¹³ kaiken uudistamisen vipupisteeksi esitettiin tavoitteiden asettamista. Näistä mahdollisimman hyvin määritellyistä tavoitteista voitaisiin sitten johtaa niin parhaat mahdolliset menettelytavat (opetusjärjestelyt) kuin rationaaliset arviointimenettelytkin. Luke-mattomat versiot komiteamietinnöissä, opetus-suunnitelmissa ja oppikirjoissa toistavat kaavaa TAVOITTEET --> OPETUSJÄRJESTELYT --> oppiminen --> ARVIOINTI. On merkille panta-vaa, että *Tyler Rationale* jättää sekä toimijat että toiminnan kontekstin suunnittelulle olennaisten seikkojen ulkopuolelle.

Tällainen toiveajattelun rationalismi johtaa hel-posti toimijuuden ja toiminnan historiallisesti rakentuneen sosiaalisen, kulttuurisen ja institu-tionaalisen kontekstin sivuuttamiseen. Huomio kapenee rationaalsiin tavoitteisiin, keinoihin ja arviointiin, kun taas toimijat ja toiminnan kon-teksti nähdään annettuna ja itsestään selvänä, sellaisena, joka on luonnollista, pysyvää ja refor-min sekä toimijan vaikutusmahdollisuuksien ulko-puolella. Syntyy reformien jatkuva kehä, jossa (oivallisista) tavoitteista ja (mainioista) opetus-järjestelyistä huolimatta (epärealistinen) reformi vääjäämättä epäonnistuu ja tarvitaan taas uusi. Syyllinen epäonnistumiseen löytyy helposti arvo-valtahierarkian alimmalta tasolta (opettajat), ja eliitti voi aloittaa uuden reformin suunnittelun.

Toiveajattelun rationalismin sitkeyttä reformi-diskurssissa voidaan selittää sillä, että se sopii oivallisesti reformin keskeisten toimijoiden intresseihin. Reformien 'tilaajille' (useimmiten (koulu)viranomaiset ja -poliitikot) on tärkeätä, että organisaatiolle (koululle) asetetaan kor-kealentoisia ja näyttäviä tavoitteita. Reformien 'tuottajia' (useimmiten (koulutus)suunnittelijat ja (pedagogiset) asiantuntijat) taas kiinnostaa oman menetelmiin ja tekniikkoihin perustuvan asian-tuntija-asemansa säilyttäminen ja vahvistaminen. Kontekstin reaalisen ja raadollisen todellisuuden vaikutusten huomioon ottaminen jää näin lat-tiatason toimijoiden huoleksi. Kaikkia reformin osapuolia yhdistää kuitenkin se, että reformissa ja sen arvioinnissa luodaan kuva instituutiosta (koulusta) kehittyvänä ja edistyvänä, tehtävänsä täyttävänä ja julkisen luottamuksen ansaitsevana instituutiona. (ks. esim. Weick 1976; Popkewitz 1991).

On merkityksellistä, että vallitseva tulosjoh-

tamisen ideologia vahvistaa *Tyler Rationalen* mukaista ajattelutapaa. Tulosjohtamisen ideolo-gia nostaa yhtäältä tavoitteet ja niiden täsmäl-lisen määrittelyn uuden koulutuksen hallinnan keskiöön (Ks. esim. Hirvi 1996; Lundgren 2003). Jotta tulosneuvotteluja voitaisiin käydä, on mää-riteltävä tavoitteet niin, että niiden toteutumista voidaan arvioida. Toisaalta se myös erottaa tavoitteiden asettamisen ja niiden toteuttamisen. 1990-luvun alussa opetushallituksen tuolloinen pääjohtaja hahmotteli tulosjohtamisen ytimen koulumaailmassa selkeästi:

Aito tulosohjaus koulumaailmassa edellyttää kahta osapuolta: tavoitteita asettavaa ja resurssit osoit-tavaa ohjausyksikköä sekä tuotteet ja palvelut tuottavaa tasoa eli oppilaitoksia. (Hirvi 1991a)

Niissä [Peruskoulun opetussuunnitelman perus-teissa] olisi ilmaistuna keskeiset oppimisen ja kasvamisen tavoitteet, joista käsin johdetaan oppi-ainekohtaiset tavoitteet, määritellään pakolliset, vapaaehtoiset ja valinnaiset oppiaineet yms. asiat. Kunnan tai oppilaitoksen opetussuunnitelmassa taas ilmaistaan se, millä tavoin nuo tavoitteet aiotaan saavuttaa (...). Tuloksellisuuden arviointi on toiminta-ajatuksen ja keskeisten tavoitteiden (tulosalueiden) toteutumisen arviointia. (Hirvi 1991b)

Jos laadunarviointia tarkastellaan yhtenä kes-keisenä koulutuksen toimintaan kohdistuvana viimeaikaisena reformina, on ajateltavissa, että toiveajattelun rationalismi toimii myös täällä. Siinä suunnataan huomio arvioinnin tavoitteisiin, mene-telmiin ja vaikutusten evaluointiin samalla sulkeis-taen ulos toiminnan yhteiskunnalliset, poliittiset ja kulttuuriset kontekstit ja ei-aiotut vaikutukset. Aivan kuten koulureformien perinteessä, vas-tuullisuuden, rationaalisuuden ja dynaamisuuden vaikutelman luominen on eri osapuolien intres-seissä.

ARVIOINNIN 'SIVUVAIKUTUKSET'

On ongelmallista käyttää käsitettä 'sivuvaiku-tukset', sillä se antaa ymmärtää, että arvioinnissa ja sen vaikutuksissa todellakin olisi kyse yksinker-taisista kausaalisuhteista. Juuri tässä mielessä Dahler-Larssen (2003) huomauttaa, ettei ole ole-massa välitöntä suhdetta arvioinnin päämäärien ja sen aikaan saamien vaikutusten välillä. Vaiku-



tus ja vaikuttavuus -käsitteiden ero on myöskin nähtävissä tästä näkökulmasta. Marketta Raja-vaaran (2006, 33) mukaan ero on vaikuttavuuden käsitteen teleologisessa luonteessa: se viittaa paitsi selvittelyn kohteeseen myös päämäärään, jota tavoitellaan. Sivuvaikutuksista puhuminen on siis järkevää vain arvioinnin tavoiterationaalisen laatuparadigman puitteissa. Ei olekaan yllättävää, että nämä kysymykset nousevat esiin arviointiteollisuuden ammattilaisten kirjoituksista. Seuraavassa tarkastelemme lyhyesti kolmea olennaista ongelma-aluetta, joihin *European Evaluation Society*n silloinen presidentti Frans L. Leeuw (2002) viittasi jäähyväispuheessaan *Challenges to a Growth Industry*. Nämä ovat arvioinnin kustannukset ja byrokratian kasvu, arvioinnin järjestämättömät vaikutukset sekä arvioinnin vaikutukset luottamukseen. Tarkastelemme näitä kolmea aluetta yliopistolaitoksen kentällä.

Kustannukset ja byrokratia

Hallintabyrokratia ja tähän kytkeytyvät kustannukset ovat kiistaton ongelma. Kyse on standardeja asettavan, valvovan ja arvioivan sekä asioihin puuttuvan hallintabyrokratian todella huomattavasta kasvusta (ks. Salter & Tapper, 2000, 72). Leeuw (2002, 5) kiinnittikin huomiota puheessaan niihin "valtavien resurssien, joita auditoinnin yhteiskunnassa käytetään laadunvarmistukseen ja joiden välittömimpänä seurauksena on byrokratisoitumisen lisääntyminen".

Yksi selkeimmistä tutkimustuloksista on, että erilaiset arviointi- ja laatuohjelmat näyttäsivät siirtävän valtaa yhtäältä yliopiston sisällä laitoksilta keskushallinnolle ja rehtorinvirastolle, toisaalta korkeakoulujen ulkopuolelle ministeriöön ja talouselämälle. Niinpä hallinnon määrälliseen kasvuun sisältyy myös 'laadullinen' tendenssi: yliopistojen sisälle on syntymässä ryhmittymä, joka näkee yliopistojen kehittämisen enemmän 'tahojen' - ministeriön ja talouselämän - silmin kuin 'yliopistolaisen' silmin.

Hallinnon paisumisesta johtuvia kustannuksia vielä merkittävämpänä monet tutkijat pitävät niitä välittömiä ja epäsuoria piilokustannuksia, joita laadunvarmistuksesta ja sen auditoinnista aiheutuu. Tällaisia ovat erityisesti ne kustannukset, jotka muodostuvat organisaatioiden muokatessa itsestään arvioitavia (Power 2003).

Useat tutkijat näkevät kuitenkin kaikkein vaka-

vimmaksi huolenaiheeksi sen, että yliopistoa(kin) konstituoivaksi praksikseksi on muodostumassa sellainen managerialistinen hallinta, jossa yliopisto nähdään yrityksenä, jonka tuotteista, niiden markkinoinnista ja samalla koko instituution kannattavuudesta on vastuussa toimitusjohtaja-rehtori. Avaintoimija on suunnittelija, joka suunnittelee tutkimus-, opetus- ja kehittämissuunnitelmia ja erilaisia projekteja, joita toteuttamaan tutkijoita ja opettajia sitten osa- ja määräajaksi palkataan. Tiedosta tulee hallinnollisen käytännön kohde, ja se muuttuu kyseenalaistamisen ja ajattelun sijaan kaupattavaksi informaatioiksi.

Sosiaaliantropologi Jukka Siikala (2005, 30-31) maalaa painajaismaista, mutta alan tutkimuskirjallisuudessa toistuvaa kuvaa managerialistisesta hallinnosta:

"Hallinto on lakannut olemasta hallintohenkilöstön toimia perinteisessä mielessä, ja siitä on tullut hallinnan vaatimusten soveltamista kaikkiin yliopiston toimiin. Hyvin hallittuna yliopiston suorittavan portaan uskotaan tuottavan laadukkaampia tuotteita kysyntää vastaavasti. (...) Hallinta ei enää perustu akateemiseen käytäntöön, vaan eroaa siitä ja ottaa sen kohteekseen. (...) Seuraukset tieteen tekemiselle, tiedon tuottamiselle ja koko tiedeyhteisön luonteelle ovat yhtä dramaattisia kuin mannerlaattojen liikkeit. Ne organisoivat intellektuaalisen toiminnan tilan ja ajan kokonaan uudelleen tuottamalla toiminnan ohjaajat ja sen periaatteet ulkopuolelta, redusoidmalla tutkija-opettajat suorittavaan portaaseen, jonka suorituksia on hallinnoitava (...) organisoimalla joukkoja jatkuvasti uudelleen ja vaihtamalla tavoitteita yhä kiihtyvään tahtiin."

Järjettömät vaikutukset ja siviiliuhrit

Laadunvarmistuksen ja sen auditoinnin tahattomat sivuvaikutukset on otettava vakavasti. Leeuw (2002, 10) ynnää em. puheessaan:

"Arvioidessani tutkimuksia tällä kentällä viimeisen 10-15 vuoden ajalta (...), johtopäätökseni on, että hyvää tarkoittavallakin evaluaatiolla on aina tahattomia ja ei-haluttuja sivuvaikutuksia, jotka vaarantavat julkisten, yksityisten ja hybridisektoreiden tehokkuutta tai vaikuttavuutta. Jotkut vertaavat näitä järjestömiä vaikutuksia (*perverse effects*) siviiliuhreihin tai oheisvahinkoihin (*collateral damages*)."



Tällaisia 'järjettömiä vaikutuksia' on tutkimus kirjannut vakuuttavan joukon arvioinnin kentältä. Ne kaikki liittyvät arvioinnin varsinaisen tavoitteen kannalta epätarkoituksenmukaisiin reaktioihin, joita esim. Asko Karjalainen (2004) listaa kolmenlaisia: sekundaaritehtäviä luovia, suojautumista synnyttäviä ja sosiaalisia kustannuksia aiheuttavia sivuvaikutuksia. Toivottavasti seuraaville brittiläisen korkeakoulumaailman kentältä nouseville käsitteille ei milloinkaan tarvitse etsiä suomenkielisiä vastineita: "*paranoia shift; game playing; dual self-evaluation; questionnaire fatigue; hijacked feedback; cherry-picking; tunnel vision; suboptimization; myopia; measure fixation; ossification; dramaturgical performance; feeding the beast*" (ks. Harvey 2002; Smith 1995; Newton 2002; 2000).

Tahattomiin mutta vakaviin oheisvahinkoihin voi kuulua myös epävirallisen, informaalisen ja yliopiston omaan akateemiseen perinteeseen juurtuneen arviointi- ja laatukulttuurin rappeutuminen. Arvioinnin siviiliihureihin taas voidaan lukea uupuneet, turhautuneet ja voimattomuuden tunnetta potevat laitostason tutkija-opettajat ja esimiehet.

Suorastaan hätkähdyttävää ennustamisen taitoa osoitti Ilkka Niiniluodon johtama *Yliopisto 2015* -työryhmän visio runsaan kymmenen vuoden takaa (22.1.1996)¹⁴. Työryhmä hahmoteli yliopistolle kolme mahdollista tulevaisuuskuvaavaa. Kaksi näistä ("hyvinvointiskenaario" ja sen vastakohta, "alajasjoskenaario") edellytti yliopiston rahoituksen merkittävää kasvua tai laskua. Kolmannessa, "kovan kilpailun" vaihtoehdossa yliopistoa

"johtaa toimitusjohtaja-rehtori opetusministeriön asettamien liiketaloudellisten tulosperiaatteiden ja kustannuslaskelmien mukaisesti. Soveltamiskelpoinen taitotieto ja ammatillinen osaaminen ovat yliopiston päätuotteet. Opetus on organisoitu tulos tavoitteet toteuttaviksi tehokkaiksi opintoväyliksi. Huippututkijoille annetaan hyvät resurssit. Tulosten avulla hankitaan puolet tarvittavista varoista."

Tässä vaihtoehdossa pahimpina uhkakuvina nähtiin muun muassa

"yliopistojen autonomian kuihtuminen, teknokratia, perustutkimuksen näivettymisestä seuraava tieteen taantuminen, ajan hukkaaminen byrokraattisiin suoritus tehtäviin, tuloksen vinosuuntaus väärin asetettujen mittareiden takia, tutkimuksen ja ope-

tuksen tason lasku, yliopiston kiinnostavuuden lasku tieteellisen kutsumustyön tarjoajana, sisäisten erimielisyyksien kärjistyminen, yliopiston hengen pilaantuminen, vilpin ja epärehellisuuden lisääntyminen opiskelussa ja tutkimuksessa, kaupallisuuden lisääntyminen, mainostoimistojen ja median valta korkeakoulumaailmassa, opettajien ja opiskelijoiden stressi, vieraantuminen, akateemisen vapauden häviö ja humanismin rappio."

Luottamuksen rappeutuminen

Laadunvarmistus ja sen auditointi ovat vaarantamassa myös luottamuksen yliopistoyhteisössä, mistä seuraa uhka yhteisön jakautumiseen kahtia. Leeuw (2002) varoitti puheessaan siitä, että arvioinnista saattaa moderneissa luottamukseen, sitoutumiseen ja hyvämaineisuuteen perustuvissa yhteiskunnissa tulla luottamuksen tuhoaja (*trust-killer*). Erityisesti hän pohtii sitä, miten "alituisesti kasvavat tarkistusten, auditointien, evaluointien ja tutkintojen kerrostumat" soveltuvat verkostojen (*networks*), yhteistoimintavaltion (*collaborative state*) ja organisaatioiden välisten yhteistyösopimusten (*partnering arrangements*) maailmaan, joka rakentuu nimenomaan luottamukselle.

Tämän ongelman taustalla on se, että äärimmäisen harvoin laadunvarmistusvaatimukset nousevat laitosten, opiskelijoiden tai opettaja-tutkijoiden taholta. Lähes poikkeuksetta ne tulevat yliopiston ulkopuolelta (so. ministeriöstä) ja laitostason yläpuolelta (so. keskushallinnosta).

Jo pelkkä 'laatutyön' määrittely ja tuonti puheeseen tuottaa myös vastakohtansa eli 'ei-laadun', joka on kaikkea sitä, mikä ei kuulu 'laatutyön' piiriin. Se tulkitaan helposti epäluottamuksen osoitukseksi. Paljon tutkimusnäyttöä onkin siitä, miten arviointi lisää epäluottamusta, kun taas hyvin vähän siitä, että se lisää tai palauttaisi menetettyä luottamusta. Siikala (2005, 36) näkee epäluottamuksen kasvun nousevan suoraan uusjohtamisen perusluonteesta. Rakentaessaan kirjjanpidollisia ja audittoivia malleja yliopistojen sosiaalisen toiminnan koko kentästä managerialismi suorittaa ekonomian imperialistisen laajentamisen eräänlaiseksi 'ikään kuin' -sosiaalisuudeksi. Yliopisto siirtyy näin 'manageriaaliseen magiaan', jolloin joudutaan tekemisiin kaikkeen magiaan aina liittyvän 'noituuden ongelman' kanssa:

”Sen lähtökohtana on, että kehenkään ei voi luottaa. Siksi kaikkia on epäiltävä, ja epäily voidaan hälventää ainoastaan läpinäkyvyydellä, ennalta määritellyllä tulosvastuullisuudella ja viime kädessä siihen suhteutetulla tilivelvollisuudella.”

LAADUNARVIOINNIN VAIKUTUKSET OSANA UUTTA HALLINNAN TAPAA

Jos laadunarviointia tarkastellaan osana hallinnan tapaa, saa vaikutusten tutkimus teoreettisen viitekehyksen, joka saattaa avata edellisiäkin mielenkiintoisempia tutkimustehtäviä.

Seuraavassa jäsenämme arvioinnin vaikutuksia *governmentality*-tutkimuksen suunnasta, erityisesti Mitchell Deanin (1999; 2002; ks. myös Simola 2004; Kantola 2002) työhön nojaten.¹⁵ Tässä katsannossa arviointia ja sen vaikutuksia tarkastellaan osana uudenlaista hallinnan tapaa. Peruskysymys on se, miten meitä hallitaan ja miten me hallitsemme itseämme. Suhteessa arviointiin voidaan tällöin asettaa ainakin neljä alustavaa tutkimuskysymystä: millaisia järkevyyden lajeja (*episteme*), näkyvyyden muotoja (*theasis*), toiminnan tapoja (*technne*) ja toimijoiden identiteettejä (*ethos*) laadunarviointi tuottaa?

Epistemen voi ajatella muodostuvan siitä, millaisen järkeilyn varassa puhe arvioinnista etenee: mitkä ajatusmuodot ja tietomuodot, millainen asiantuntemus, millaiset strategiat sekä kalkyloinnin, rationaalisuuden ja totuuden muodot ovat arvioinnista puhuttaessa käytössä. Voidaan puhua arvioinnin päämäärien, pyrkimysten, kritiikin ja problematisoinnin järjestelmistä, käsitteellisestä sanastoista, jotka kuvaavat todellisuutta ja tuottavat siitä totuuksia ja tosiasioita. Analyysin tehtävänä on löytää näitä kielellisiä konstruktioita ja kysyä, mistä ja miten ne ilmaantuvat ja millä tavoin ne on tuotettu ja miten ne toimivat. Arviointia tulisi siis tutkia osana poliittista rationaliteettia, tietynlaisena älyllisenä koneistona, joka muokkaa todellisuudesta sellaista, että siihen voidaan soveltaa poliittista ohjelmaa.¹⁶ Historiallisesti kyse on muuttuvista totuuksien tai järkeilyn järjestelmistä. Tämä asioiden ja sanojen väliin jäävälle harmaalle vyöhykkeelle sijoituvista postiivinen tiedostamaton rakentuu olennaisella tavalla itsestäänselvyyksille, jotka hyvin tiedettyinä mutta huonosti tiedostettuina, yleisesti tunnustettuina, mutta harvoin tunnistettuina konstituoi ajatteluamme, puhettamme ja toimin-

taamme.¹⁷

Theasis viittaa erityiseen näkymisen ja näkemisen tapaan, joka on ominaista ja välttämättömiä tietyille hallinnan regiimille. Niinpä voidaan kysyä, millaisia keinoja käytetään toimijoiden ja toiminnan nostamiseksi arvioitaviksi, mutta myös miten toisia piiloitetaan ja tehdään näkymättömiksi. Kyse on arvioinnin visuaalisesta ja spatioalisesta ulottuvuudesta, vallan ja auktoriteetin diagrammista, jonka kautta arvioinnin kohde 'käsitetään' - käyttäksemme sanan vanhaa merkitystä konkreettisesta kiinnittämisestä. Huomion kohteeksi tulevat tällöin epäilemättä erilaiset numerot, indikaattorit, ranking-listat ja vertailut, joilla arvioinnin kohteita visualisoidaan.¹⁸

Technnellä puolestaan kiinnitetään huomiota siihen, millaisia ovat oikeina ja järkevinä pidetyt toimintatavat; millaisia välineitä ja keinoja (mekanismeja, toimenpiteitä, instrumentteja, tekniikoita, stategioita) arvioinnissa on käytössä. Arvioinnin välineiden ja niiden vaikutusten korostamisen voi nähdä kritiikkinä tavoiterationalistiselle ajattelulle, missä oletetaan yhteys tavoitteiden (joita perustellaan arvoilla, ideologioilla tai järjellä) ja niiden saavutamiseksi käytettävillä välineillä ja keinoilla (joiden ajatellaan olevan johdettu tavoitteista). Tekniikat ja menettelytavat saavat aikaan vaikutuksia niiden perustelusta riippumatta(kin).

Ethoksen ulottuvuudella kysytään, millainen yksilöllinen ja kollektiivinen identiteetti on toiminnan ja muutoksen kohteena; mitä statuksia, kapasiteetteja, attribuutteja ja orientaatioita pyritään vahvistamaan ja mitä kielteisiä ominaisuuksia taas karsimaan. On ajateltavissa, että tietyt arvioinnin muodot edistävät, herättävät, nostavat ja houkuttelevat esiin tiettyjä ominaisuuksia ja kapasiteetteja toimijoissa, ja että ne ovat "tehokkaita siinä määrin, kuin ne saavat toimijat näkemään itsensä näiden tiettyjen ominaisuuksien kautta" (Kantola 2002, 32). Taustalla on liberalistisen näkemys siitä, että kyse on "modernin, eettisen yksilösubjektin luonteesta ja muutoutumisesta" (Helén 1997, 33). Eetos ei ole pelkästään yksilön itselleen tuottama ominaisuus, "vaan ennemminkin käsite, jolla voidaan tarkastella yksilön ja yhteiskunnan vuorovaikutusta" (Kantola 2002, 32). Sitä voidaan pitää "yhteiskunnan ja minän kohtaupaikkana, minän rakentamisen paikkana, joka pistää yksilön tuottamaan itselleen tiettyyn paikkaan ja asemaan liittyvää eettistä tapaa olla ja ajatella. Samalla eetoksen kautta tulevat näkyviksi myös vallitsevat valta-

asetelmat, elämäntavat ja -arvot” (Helén 1997, 335).

Tässä hengessä Power (2003) näkeekin arviointiin kohdistuvassa tutkimuksessa tärkeimmän tutkimuskohteen olevan juuri tämän uuden auditoitavan subjektin sosiaalisen rakentumisen.

Auditoitava on epäilemättä mutkikas olento: samanaikaisesti ovela ja masentunut; taitava myöntöväisyyspeleissä, mutta samalla kyyninen ja loppen uupunut niihin. Häntä ärsyttävät tyhjää täynnä olevat löysät arviointitodistukset, mutta samalla hän pelkää laajennettuja paikallisia auditoitavuuksia. Hän karsastaa auditoijien keskinkertaisuutta, mutta valittaa samalla heidän voimattomuuttaan panna kuriin todella ”pahat pojat”. Hän suree ajan tuhlaamista tarkastuksen rituaaleihin, mutta hyväksyy sen, että tässä me saamme ”sen minkä ansaitsemmekin”. Hänen mielestään pätevä ja erinomainen kärsii vääryyttä yrittäessään tulla toimeen laadunvarmistuksen vaatimusten kanssa samaan aikaan kun epäpätevä ja laiska onnistuu väistelemään sen pahimmat kohtuuttomuudet. Hän kuulee virallisten dokumenttien erinomaisuusretoriikan, mutta elää taantuman todellisuudessa. Hän tekee kollegojensa tavoin muistiinpanoja kokouksissa ”kaiken varalta” ja hänellä on enemmän arkistokaappeja kuin muutama vuosi sitten. Hän tietää menneisyyden olleen kaukana onnelasta, mutta on epätoivoinen auditoinnin rautahäkissä. Hän tietää, että tulosvastuu ja asiakasdialogi ovat hyviä asioita, mutta ihmettelee, miksi kaikkien koulutuksessa vietettyjen vuosien jälkeen häneen ei enää asian tuntijana luoteta. (Power 2003, 199-200)

Anu Kantola (2002) kirjottaa *governamentaliteetti*-koulukuntaan kohdistuneesta kritiikistä. On esimerkiksi esitetty, että hallinnan tutkijoilta puuttuisi poliittinen ja taloudellinen teoria, joka antaisi substanssia heidän analyyseilleen. Tästä syystä hallinnan käsitteen alla tehdyt analyysit tuottavat usein kyllä kiinnostavia ja osuvia kuvauksia, mutta jäävät jotenkin irralleen konkreettisista historiallisista valtasuhteista ja yhteiskunnallisista rakenteista. Kantola (emt., 37-41) pitää Foucault’iin ja hallinnan tavan tutkijoihin kohdistettua kritiikkiä osin oikeutettuna. Esimerkiksi sosiaaliset, poliittiset ja taloudelliset rakenteet, hallinnan konkreettiset seuraukset ja vaikutukset, muutosten takana olevat väestöryhmät, voimat ja instituutiot sekä hallinnan muutoksen päämää-

rät näyttävät monesti jäävän vähälle huomiolle. Kantola (emt., 27) korostaa, että Foucault ei itse näe edistyneiden liberaalien yhteiskuntien hallintaa pelkästään vapauden ja itsehallinnan kautta tapahtuvana. Modernissa hallinnassa ovat aina erityisissä muodoissaan läsnä niin kurinpito kuin kuninkuuskin. Poliittisen vallan muodoissa on aina kyse ”tietyissä ajassa ja paikassa risteilevistä voimista ja vastavoimista sekä niiden kautta muodostuvista hallinnan muodoista”.

Hallinnan tavan käsitettä ei pidä nähdä niinkään yleisenä ja universaalina, vaan hypoteettisena, tiettyä historiallista tapausta kehystävänä käsitteenä. Hallinnan tapaan tai edes sen edellä hahmoteltuihin ulottuvuuksiin ei ole syytä lukkiutua ja kuvitella, että niillä olisi täsmälliset konkreettiset vastineensa todellisuudessa ja että ’hallinnan tapa’ olisi sellaisenaan olemassa. Hallinnan tapaa on mielekästä käyttää yhdenlaisena ’heuristisena laitteena’: se ei ole teoria, selitys tai järjestelmä, joka vapauttaisi tutkijan ajattelemasta. Se on ponnahduslauta, mahdollisuuksien luettelo ja havaintojen tuottamisen apuväline. (Simola 2004, 124-126; Simola et al 1998; Heikkinen et al 1999, av. 8)

Onkin koko ajan perusteltua kysyä Kantolan (2002, 52) tavoin, ”miten hallinnan rationaliteetit ovat kytkeytyneet yhteiskunnan rakenteisiin ja ovatko ne oikeuttamassa esimerkiksi markkinoiden, pääomien tai korporatiivisen poliittisen järjestelmän valta-asemia.” Palataksemme alussa esitettyyn Dahler-Larssenin politiikka-käsitteen erotteluun, voitaisi ehkä sanoa, että *governamentaliteetti*-näkökulma avaa erityisesti sekä arvioinnin toimintapolitiian (*evaluation policy*) ja poliittisuuden (*the political in evaluation*) näkökulmaa, kun taas arvioinnin politiikan (*evaluation politics*) tutkimiselle se on vain (tärkeä) lähtökohta.

YHTEENVETO

Tässä artikkelissa on etsitty vastauksia kahteen kysymykseen. Ensimmäkin, miten on ymmärrettävissä se, että arviointiin liittyvä yhteiskuntatieteellinen tutkimus on keskittynyt tuottamaan tietoa pikemminkin arvioinnin tilaajia ja suorittajia varten kuin lisäämään ymmärrystä arvioinnista yhteiskunnallisena ilmiönä. Erityisesti olemme kysyneet, miksi jo arvioinnin sellaiset vaikutukset, jotka liittyvät arvioinnin päämääriin, mutta ovat ei-aiottuja, ovat jääneet tutkimuksen kat-

vealueelle puhumattakaan näiden päämääriin liittymättömistä yhteiskunnallisista seuraamuksista. Toiseksi olemme pyrkineet hahmottelemaan sellaisia tutkimuskysymyksiä, jotka avaisivat juuri näitä tähän asti syrjään jääneitä arvioinnin ja tässä erityisesti laadunarvioinnin vaikutuksia.

Näyttäisi siltä, että tuijottaminen arvioinnin julkausuttuihin tavoitteisiin on tekemisissä arvioinnin poliittisen ja hallinnallisen, vallankäyttöön syvästi liittyvän luonteen kanssa. Arvioinnista on itsestään tullut oleellinen osa politiikkaa ja hallintoa. Arvioinnin tilaustutkimusluonteen takia arvioinnin ei-aiotut vaikutukset ovat tilaajien kannalta vähemmän toivottuja, epäkiinnostavia ja myös epämieluisia tuloksia. Arvioita toteuttavien arviointiammattilaisten professionaalisten intressien mukaista on keskittyä arvioinnin menetelmälliseen osaamiseen ja tilaajan vakuuttamiseen arviointitekniikan hallitsemisesta eikä pohdiskella toiminnan katve- tai piilovaikutuksia. Arvioinnin tilaajien tavoitteeskeisyys ja arvioinnin suorittajien menetelmäkeskeisyys vahvistavat sellaisen vallitsevan diskurssin voimaa, jota tässä on luonnehdittu toiveajattelun rationalismiksi. Huomion keskittyminen arvioinnin tavoitteisiin, menetelmiin ja vaikutusten evaluointiin sulkeistaa samalla sopivalla tavalla ulos sellaiset kiusalliset kysymykset, jotka kytkeytyvät toiminnan yhteiskunnallisiin, poliittisiin ja kulttuurisiin konteksteihin ja ei-aiottuihin vaikutuksiin.

Intressiryhmänäkökulmasta on perusteltua puhua jopa eräänlaisen 'laatu- ja arviointiteollisen kompleksin'¹⁹ muotoutumisesta, jossa yhtyvät markkinaliberalististen poliitikkojen ja virkamiesten sekä laatu- ja arviointiammattilaisten intressit. Näin ainakin osittain voisi olla ymmärrettävää se, miten sellaiset polttavat vaikuttavuuteen liittyvät ongelmat kuten arvioinnin kustannukset ja byrokratian kasvu, arvioinnin järjettömät, päämäärien kanssa ristiriidassa olevat vaikutukset ja 'siviiliuhrit' sekä luottamusta rappeuttavat vaikutukset ovat jääneet laajemman mielenkiinnon ulkopuolelle. Kannattaa huomata, että edellä mainituista kriittisistä huomautuksista huolimatta Leeuwin (2002) puhe oli kokonaisuudessaan positiivinen ja eteenpäin suuntautuva.

Governmentality-tutkimuksen viitekehuksesta käsin taas saattaisi olla hyödyllistä aloittaa tutkimus kysymällä, millaisia järkevyyden lajeja, näkyvyyden muotoja, toiminnan tapoja ja toimijoiden identiteettejä arvioinnin kautta tuotetaan ja myös millaisia tuhotaan. Tätä kautta saatet-

taisiin päästä käsiksi sellaisiin olennaisiin yhteiskunnallisiin ja poliittisiin kysymyksiin kuin miten arviointi ja sen vaikutukset liittyvät taloudellisiin ja poliittisiin rakenteisiin, muutoksiin ja valtasuh- teisiin.

VIITTEET

¹ Suomen kieli on tunnetusti varsin eriytymätön politiikkaan, valtaan ja hallintaan liittyvien käsitteiden suhteen (ks. esim. Pekonen 2005; Hyvärinen et al 2003). Seuraavassa käytämme yhteiskunnallinen -käsitettä kattamaan kaikki Dahler-Larsenin englantinkieliset politiikka-käsitteet.

² Vrt. Rinne2001; 2002a; 2002b; 2004; 2005; Rinne, Kallo & Hokka 2004; Rinne & Koivula 2005; Kallo & Rinne 2006; Simola & Rinne 2004; Simola 2005; Rinne & Simola 2005; Simola 2006

³ evaluation state, ks. esim Neave, 1988; New Public Management ks. esim. Power, 1999; kilpailuvaltio, ks. esim. Heiskala 2006; academic capitalism ks. esim. Slaughter & Leslie 1997.

⁴ Tällä emme tarkoita sitä, etteikö esim. hyvinvointivaltiossa normi- ja resurssiohjauksella olisi ollut omia arviointi- ja palautekäytäntöjään. Niiden merkitys ja paikka hallinnan järjestelmässä oli vain olennaisesti toisenlainen.

⁵ Ks. Ball 2001, Simola 2001, Kantola 2002.

⁶ a management support system ja an integral part of routinised regimes of organisational guidance and control; ks. Dahler-Larsen, 2000; Power 1999

⁷ <http://www.minedu.fi/julkaisut/koulutus/2004/tr06/tr06.pdf>

⁸ Esim. Leeuw (2002) käyttää termiä Evaluation Industry.

⁹ http://www.bologna-bergen2005.no/Docs/00-Main_doc/050221_ENQA_report.pdf

¹⁰ http://www.kka.fi/pdf/julkaisut/KKA_405.pdf

¹¹ Laadunvarmistuksen nykytila Helsingin yliopistossa HELSINGIN YLIOPISTO, Kehittämisosasto, laadunvarmistus Aimo Virtanen, 24.11.2005.

¹² Arviointi on todellakin tullut uudella tavalla ja voimallisesti koulutukseen. Havainnollinen esimerkki on vuoden 2004 Perusopetuksen opetussuunnitelman perusteet (POP 2004), jossa sana 'kasvatus' (jossain taivutusmuodossa) esiintyy vain 36 kertaa kun taas sana 'arviointi' mainitaan 380 kertaa.

¹³ Tyler Rationale on saanut nimensä Ralph W. Tylerin (1950) vaikutusvaltaisesta julkaisusta, jossa kaiken suunnittelun lähtökohdaksi esitettiin tavoitteiden asettaminen. Mallia kohtaan on kautta vuosien varsinkin kansainvälisessä keskustelussa esitetty runsaastikin kritiikkiä (ks. Simola, 2000; 2002), mutta se on säilyttänyt asemansa opetuksen suunnittelun paradigmana niin valtiollisella kuin yksittäisen opettajan tasolla.

¹⁴ www.helsinki.fi/strategia/2015.html

¹⁵ Deanin lisäksi Governmentality-koulukunnan muita keskeisiä tutkijoita ovat Nikolas Rose, Peter Miller, Bary Hindess ja Graham Burchell. Heistä monet kuuluvat 1989 perustettuun History of the Present -tutkimusverkostoon, johon nykyisin kuuluu satakunta tutkijaa Euroopasta, erityisesti Britanniasta, sekä Pohjois-Amerikasta ja Australiasta. Suomalaisista ainakin Risto Eräsaari ja Sakari Hänninen ovat aktiivisesti osallistuneet verkoston toimintaan.

¹⁶ Ks. Rose 1996. Hilikka Summa (1995, 147) puhuu rationaliteetista juuri tässä samassa hengessä kokonaisuutena, ”johon sisältyy sekä tietoa ja käsityksiä tosiasioista ja syy- ja seuraussuhteista, käsitteistöjä ja tapoja puhua asioista, että tavoitteistoja ja näkemyksiä toivotusta kehityksen suunnassa”.

¹⁷ Ks. esim. Foucault 2005; Popkewitz 1998; 2001; Simola 1995

¹⁸ Ks. esim. Rose 1999a; Desrosieres 2002

¹⁹ Laatu- ja arviointidiskurssissa käytetään huomiota herättävän paljon sotilaskielestä peräisin olevia termejä ja metaforia kuten esim. strategia ja taktiikka, oheisvahingot ja siviiliuhrit, sota ja tappajat jne. 'Laatu- ja arviointiteollisen kompleksin' metafora viittaa Dwight D. Eisenhooverin jäähyväispuheessaan 1961 kansakunnalle esittämän varoitukseen 'sotilaallisteollisen kompleksin' (military-industrial complex) vaikutusvaltaan sisältyvistä vaaroista. Sotilaallisteollisen kompleksin 'rautaisen kolmion' ajateltiin muodostuvan asevoimista, aseteollisuudesta ja kongressin sotilaallisia väliintuloja kannattaneista 'haukoista'. http://en.wikipedia.org/wiki/Military-industrial_complex . Luettu 9.6.2006

KIRJALLISUUSVIITTEET

- Ahonen P. (1996) Evaluaatiokokemuksista evaluaatiovakaumuksiin. Julkaisussa Laukkanen R, Stenvall K. Arviointi koulutus- ja tiedepolitiikassa, ss. 147-171. Tampere: Tampereen yliopisto, Hallintotiede A 9.
- Ahonen P. (1998) Jotta arviointi olisi jotain, se ei voi olla kaikkea. Hallinnon tutkimus 17 (4), 257-261.
- Ahonen, P. (2001). Evaluaatio-oppia edistyville. Hallinnon Tutkimus, 20, 102-117.
- Amaral, A., Fulton, O. & Larsen, I. M. (2003) A managerial revolution? In A. Amaral, V. L. Meek & I. M. Larsen (toim.) The higher education managerial revolution? Dordrecht, Boston & London: Kluwer academic publishers, 275-297.
- Ball, S. J. (2001). Globaalit toimintaperiaatteet ja kansalliset politiikat eurooppalaisessa koulutuksessa. In A. Jauhainen, R. Rinne & J. Tähtinen (Toim.), Koulutuspolitiikka suomessa ja ylikansalliset mallit (pp. 21-44). Turku: Suomen kasvatustieteellinen seura. Kasvatusalan tutkimuksia 1.
- Ball, S. J. (2003) The teacher's soul and the terrors of performativity. Journal of education policy, 18(2), 215-228.
- Ball, S. J. (2004). Suorituskeskeisyys ja yksityistäminen jälkihyvinvointivaltion koulutuspolitiikassa Kasvatus, 35(1), 6-20.
- Callahan, Raymond (1963) Education and the Cult of Efficiency. A study of the social forces that have shaped the administration of the public schools. New York.
- Cerych, L. & Sabatier, P (1986) Great expectations and mixed performance: The implementation of higher education reforms in Europe. Trentham: Trentham books.
- Collins, R. (1990). Changing conceptions in the sociology of professions. Teoksessa R. Torstendahl, & M. Burrage (Toim.), The formation of professions : Knowledge, state and strategy (pp. 11-23). London: Sage.
- Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newbury Park, CA: Sage Publications.
- Dahler-Larsen, P. (2000). Surviving the routinization of evaluation: The administrative use of evaluations in Danish municipalities. Administration & Society, 32(1), 70-92.
- Dahler-Larsen, P. (2001). From programme theory to constructivism. On tragic, magic and competing programmes. Evaluation, 7(3), 331-350.
- Dahler-Larsen, P. (2003). Det politiske i evaluering. Studies in Educational Policy and Educational Philosophy, E-Tidskrift, 2(1), 1-12.
- Dean, M. (2002). Liberal government and authoritarianism. Economy and Society, 31(1), 37-61.
- Dean, M. (1999). Governmentality. Power and rule in modern society. London: Sage.
- Desrosieres, Alain (2002) The Politics of Large Numbers: A History of Statistical Reasoning Harvard University Press.
- Foucault, M.(2000). Tarkkailla ja rangaista. Helsingissä: Otava.
- Foucault, M. (2005). Tiedon arkeologia. Tampere: Vastapaino.
- Harvey, L. (2002). Evaluation for what? Teaching in Higher Education, 7(3), 245.
- Harvey, L. (2004). War of the worlds: Who wins in the battle for quality supremacy? Quality in Higher Education, 10(1), 65-72.
- Heikkinen, S., Silvonen, J. & Simola, H. (1999). Technologies of Truth: Peeling Foucault's Triangular Onion. Discourse (20), 1, 141-157.
- Heiskala, R. (2005). Kansainvälisen toimintaympäristön muutos ja Suomen yhteiskunnallinen murros. Teoksessa R. Heiskala, & E. Luhtakallio (Toim.), Uusi jako : Miten suomesta tuli kilpailukyky-yhteiskunta? (pp. 14-42). Helsinki: Gaudeamus.
- Helén, Ilpo (1997) Äidin elämänpolitiikka . Helsinki: Gaudeamus.
- Hirvi, V. (1991a). Koulutuspolitiikan suuntaa täsmennettävä. Tiedotusvälineille annettu lausunto 12.8.1991. Moniste.
- Hirvi, V. (1991b). Puheenvuoro opetushallituksen opetus-suunnitelmaseminaarissa Tampereella 30.10.1991. Moniste.
- Hirvi, V. (1996). Koulutuksen rytminvaihdos. 1990-luvun

- koulutuspolitiikka suomessa. Helsinki: Otava.
- Hyvärinen, M. et al (toim.). (2003). Käsitteet liikkeessä : Suomen poliittisen kulttuurin käsitehistoria. Tampere: Vastapaino.
- Kallo, J. & Rinne, R. (2006) (toim.) Supranational Regimes and National Education Policies - Encountering Challenge. Kasvatusalan tutkimuksia. Suomen Kasvatustieteellinen Seura.
- Kantola, A. (2002). Markkinakuri ja managerivalta. Poliittinen hallinta suomen 1990-luvun talouskriisissä. Helsinki: Viestinnän laitos/Loki kirjat.
- Karjalainen, A. (2004) Koulutuksen laatujärjestelmän perusteet. http://tievie oulu.fi/arvioinnin_abc/laatu/koullaat.pdf. Lueutu 5.6.2006.
- Koivusalo, M. (2005). Foucault ja elämän hallitseminen. kristillisestä tottelevaisuudesta laskelmoiviin taloussubjekteihin. Tiede & Edistys, 30(4), 355-379.
- Leeuw, F. L. (2002) "Evaluation in Europe 2000: Challenges to a Growth Industry". Evaluation, 8(1), 5-12.
- Lundgren, U. P. (1972). Frame factors and the teaching process: A contribution to curriculum theory and theory on teaching. Stockholm: Almqvist & Wiksell. , 378.
- Lundgren, U., P. (2003). Political governing of the education sector: Reflecting on change. Studies in Educational Policy and Educational Philosophy: E-Tidskrif (1)from <<http://www.upi.artisan.se>>.
- Meek, V. L. (2003) Governance and management of Australian higher education: Enemies within and without. Teoksessa A. Amaral, V. L. Meek & I. M. Larsen (toim.) The higher education managerial revolution? Dordrecht, Boston & London: Kluwer academic publishers, 275-297.
- Neave, G. (1988). On the cultivation of quality, efficiency and enterprise: An overview of recent trends in higher education in Western Europe, 1986-1988. European Journal of Education, 23(1-2), 7-23.
- Newton, J. (2000). Feeding the beast or improving quality?: Academics' perceptions of quality assurance and quality monitoring. Quality in Higher Education, 6(2), 153-11.
- Newton, J. (2002) Views from Below: academics coping with quality. Quality in Higher Education, 8(1), 39-23.
- Pekonen, K. (2005). Suomalaisen hallitsemiskäsitteistön historiaa. Helsinki: Yliopistopaino.
- Popkewitz, T. S. (1991). A political sociology of educational reform. New York: Teachers' College press.
- Popkewitz, T. S. (1998). Struggling for the soul. The politics of schooling and the construction of the teacher. New York: Teachers College Press.
- Popkewitz, T. S. (2001). The production of reason and power: Curriculum history and intellectual tradition. Teoksessa T. S. Popkewitz, B. M. Franklin & M. A. Pereyra (Toim.), Cultural history and education. critical essays on knowledge and schooling (pp. 151-183). New York: Routledge Falmer.
- Power, M. (1994) The audit explosion. London: Demos.
- Power, M. (1999). The audit society: Rituals of verification. Oxford: Oxford University Press.
- Power, M. (2003). Evaluating the audit explosion. Law & Policy, 25(3), 185-203:
- Pressman, J. L. & Wildavsky, A. (1974) Implementation: How great expectations in Washington are dashed in Oakland; Or, why it's amazing that federal programs work at all. Berkeley: University of California Press.
- Rajavaara, M. (2006). Yhteiskuntaan vaikuttava kela : Katsaus vaikuttavuuden käsitteisiin ja arviointiin. Helsinki: Kansaneläkelaitos. Tutkimusosasto.
- Rinne, R. (2001) Koulutuspolitiikan käänne ja nuorten syrjäytyminen. Teoksessa A. Jauhiainen, R. Rinne & J. Tähtinen (toim.) Koulutuspolitiikka Suomessa ja ylikansalliset mallit. Turku: Suomen kasvatustieteellinen seura, Kasvatusalan tutkimuksia 1, 91-138.
- Rinne, R. (2002a) Binaarimallista Bolognan tielle: Erilliset ammattikorkeakoulut tulevat ja menevät. Teoksessa J-P. Liljander (toim.) Omalla tiellä. Ammattikorkeakoulut kymmenen vuotta. Helsinki: Edita, 80-106.
- Rinne, R. (2002b) Kansallisen koulutuspolitiikan loppu? Teoksessa R. Honkonen (toim.) Koulutuksen lumo. Retoriikka, politiikka ja arviointi. Tampereen yliopiston julkaisuja, 93-120.
- Rinne, R. (2004) Searching for the Rainbow. Changing the Course of Finnish Higher Education. Teoksessa I. Fägerlind & G. Strömqvist (ed) Forming Higher Education in the Nordic Countries. Paris: IIEP, Unesco, 89-136.
- Rinne, R. (2005) Globalisaatio, ylikansalliset järjestöt ja kansallisen yliopistopolitiikan taittuminen. Teoksessa S. Häyrynen (toim.) Kulttuurin arviointi ja vaikutusten välät. Kulttuuripoliittisen tutkimuksen edistämässä-tiön julkaisuja 12. Helsinki: Yliopistopaino, 69-102.
- Rinne, R. & Koivula, J. (2005) The Changing Place of the University and the Clash of Values Higher Education Management and Policy 17, 3, 91-124.
- Rinne, R., & Simola, H. (2005). Koulutuksen ylikansalliset paineet ja yliopistojen uusi hallinta. Tiede ja edistys, 30(1), 6-26.
- Rinne, R., Kallio, J. & Hokka, S. (2004) Too eager to comply? OECD education policies and the Finnish response. European Educational Research Journal 3 (2): 454-485.
- Rinne, R., Kivinen, O. & Kivirauma, J. (1984). Lähtökohtia peruskoulutuksen yhteiskunnallis-historiallisen muotoutumisen tutkimiseen. Turku: Turun yliopiston kasvatustieteiden tiedekunta. Julkaisusarja B:11.
- Rose, N. & Miller, P. 1992. Political power beyond the state. British Journal of Sociology 43, 173-205.
- Rose, N. (1999a) Powers of freedom. Reframing political thought. Cambridge: Cambridge University Press.
- Rose, N. (1999b) Governing liberty. Teoksessa R. V. Ericson & N. Stehr (toim.) Governing modern societies, 141-176. Toronto: University of Toronto Press.
- Salter, B., & Tapper, T. (2002). The external pressures on the internal governance of universities. Higher Education Quarterly, 56(3), 245.
- Scriven M. (1991) Evaluation thesaurus. Newbury Park: Sage.
- Segerholm, C. (2003). Researching Evaluation in National (State) Politics and Administration: A Critical Approach

- ach. *American Journal of Evaluation*, 24(3), 353-20.
- Sennett, R. (1998) *The corrosion of character: the personal consequences of work in the new capitalism*. New York: W. W. Norton.
- Siikala, J. (2005). *Argumentteja auditoinnista*. *Tiede & Edistys*, 30(1), 27-37.
- Simola, H. (1995). *Paljon vartijat. Suomalainen kansanopettaja valtiollisessa kouludiskurssissa 1860-luvulta 1990-luvulle*. Helsingin yliopiston opettajankoulutuslaitos. Tutkimuksia 137, Helsinki
- Simola, H. (2000). *Firmly Bolted into the Air: Wishful rationalism as a Discursive Basis for Educational Reforms?* Teoksessa Ball, S. J., (Ed.) (2000). *Sociology of Education. Major Themes, vol. IV* London: Routledge / Falmer, pp. 2112-2138.)
- Simola, H. (2001). *Koulupolitiikka ja erinomaisuuden eetos*. *Kasvatus*, 32(3), 290-297.
- Simola, H. (2002) *Ilmaan propattu - toiveiden rationalismi koulureformien diskursiivisena perustana*. Teoksessa R. Honkonen (ed.) *Koulutuksen lumo - Retoriikka, politiikka ja arviointi*. Pp. 55-73. Tampere: Tampere University Press.
- Simola, H. (2004). *Kohti kolmannen tasavallan koulutuspolitiikkaa. Uusi hallinnan tapa, koulutus ja syrjäytyminen suomessa*. Teoksessa J. Löfström, J. Rantala & J. Salminen (toim.), *Esseitä historiallis-yhteiskunnallisesta kasvatuksesta* (ss. 115-138). Helsinki: Historiallis-yhteiskuntatiedollisen kasvatuksen tutkimus- ja kehittämiskeskus.
- Simola, H. (2005). *Koulukohtaiset oppimistulokset ja julkisuus*. *Yhteiskuntapolitiikka*, 70(2), 179-187.
- Simola, H. (2006). *Kuusi teesiä laadunvarmistuksesta*. *Tieteessä Tapahtuu*, 23(3), 44-49.
- Simola, H., & Rinne, R. (2004). *Laatuvallankumous, arviointiteollisuus ja korkeakoulut*. *Kasvatus*, 35(3), 330-341.
- Simola, H., Heikkinen, S. & Silvonen, J. (1998). *Catalog of possibilities: Foucaultian history of truth and education research*. Teoksessa T. S. Popkewitz & M. Brennan (Toim.), *Foucault's challenge: Discourse, knowledge, and power in education* (pp. 64-90). New York: Teachers College Press.
- Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Baltimore: The Johns Hopkins University Press.
- Smith, P. (1995). *On the unintended consequences of publishing performance data in the public sector*. *International Journal of Public Administration*, 18, 277-310.
- Summa, H. (1995) *Tulosohjaus hallinnan rationaliteettina ja tekniikkoina. Arvioita tulosohjauksen valtionhallintoon jättämistä jäljistä*, *Hallinnon tutkimus*, 14(2), 143-151.
- Teddlie, Charkes & Reynolds, David (2000) (ed.) *The International Handbook of School Effectiveness Research*. London: Falmer Press.
- Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia a century of public school reform*. London: Harvard University Press.
- Uusikylä, P. (1999). *Politiikan ja hallinnon arviointi*. Teoksessa R. Eräsaari, T. Lindqvist, M. Mäntysaari & M. Rajavaara (Toim.), *Arviointi ja asiantuntijuus* (pp. 17-30). Helsinki: Gaudeamus.
- Tyler, R. W. (1950). *Basic principles of curriculum and instruction*. Chicago: The University of Chicago Press.
- Valtonen, P. (1997). *Koulutuksen vaikuttavuustutkimuksen tila Suomessa*. Teoksessa R. Raivola, P. Valtonen & M. Vuorensyrjä (Toim.), *Koulutus, yhteiskunta, menestys. suomen akatemia koulutuksen vaikutusohjelman tutkimuksia* (pp. 9-53). Helsinki: Edita.
- Weick, K. (1976). *Educational organizations as loosely coupled systems*. *Administrative Science Quarterly*, 21(1), 1-19.
- Vedung, E. (1997). *Public policy and program evaluation*. New Brunswick (N. J.): Transaction Publishers.
- Vedung, E. (2003). *Arviointiaalto ja sen liikkeelle panevat voimat*. *Stakes, FinSoc Työpapereita 2/2003*. Helsinki: Stakes.
- Vedung, E. (2004) *Arviointiaallon muodot ja suunnat*. *Yhteiskuntapolitiikka* 69(3), 242-250.
- Virtanen, P. (2001). *Arviointiosaamisen ulottuvuudet julkisella ja kolmannella sektorilla*. *Hallinnon Tutkimus*, 20(2), 144-152.

Vaikuttavuuden tietokulttuurit

Marketta Rajavaara

ABSTRACT

Outcome Knowledge and Epistemic Cultures

Finnish public services and welfare practices are under pressure for greater accountability and increased attention to effectiveness, performance measurement and evaluation. While evaluation practices are usually understood as autonomous epistemic cultures, there is an alternative way of imagining their role: evaluation could be seen as an essential element of different epistemic cultures, including academic disciplines and administrative and professional cultures. The aim of the article is to bring into discussion the role of epistemic cultures in the production and utilization of outcome knowledge. Some of the most common "styles of reasoning" concerning effectiveness are presented

SANAN VALTA

"Tämä käsite ei merkitse mitään. Käytä sitä niin usein kuin mahdollista. Tämä ohje on liitetty postmodernismiin, mutta se soveltuu yhtä hyvin nykyisten hallinnon käytäntöjen ja kehittämisoppien tarkasteluun." (Wilkinson & Wilmott 1995, 1.)

Adrian Wilkinson ja Hugh Wilmott (1995) tarkastelevat näin laadun käsitettä: se ilmaantui julkisiin palveluihin 1980-luvulla, ja sen ympärille alkoi kietoutua nopeasti monenlaisia merkityksiä, käytäntöjä, oppeja, uudistuspyrkimyksiä ja voimainkoetuksia. Arvioinnin ja vaikuttavuuden käsitteet ovat samanlaisia. Ne on omaksuttu laajasti julkiselle ja yksityiselle sektorille sekä järjestöjen toimintaan. Kuitenkaan näitä käsitteitä ei voida määritellä siten, että kaikki voisivat olla määrittelyistä edes suurin piirtein yksimielisiä (Ahonen 1998).

Nikolas Rosen (2000, 9) mukaan käsitteet voivat olla tärkeämpiä sen vuoksi, mitä ne tekevät kuin siksi, mitä ne merkitsevät. Ian Hackingin mukaan jokin käsite, kuten vaikuttavuus, voidaan ymmärtää sanaksi eri sijainneissaan. Niitä ovat lauseet ja kielenkäyttö, käsitteiden muodostajat, instituutiot, auktoriteetit ja valta-asetelmat sekä

materiaaliset asiat, joita käsitteen toimivuus edellyttää. Hacking puhuu organisoivista käsitteistä tarkottaessaan niiden kykyä muovata ihmisten toimintaa. Asioiden nimittämisestä tai luokittelusta jonkin käsitteen avulla muodostuu uudenlaisia olemisen ja tekemisen mahdollisuuksia. Vakiinnuttuaan käsitteet voivat jopa määrällä ihmisten elämää. Tällaista tietokäsitystä, jossa painotetaan nimeämisen käytäntöjen vuorovaikutusta nimeämisen kohteiden kanssa, Hacking kutsuu dynaamiseksi nominalismiksi. (Hacking 2002a, 2, 17-20, 34-35, 48, 68; Lehtonen 2003.)

Vaikuttavuuden käsitteeseen tutustuttiin julkisessa hallinnossa 1960- ja 1970-luvuilla suunnitelun käytäntöjä laajennettaessa. Sen läpimurto alkoi valtion ja kuntien hallinnossa 1980-luvun lopulla, kun se määriteltiin tulohajasta käytönotettaessa yhdeksi julkisen toiminnan tuloksellisuuden osatekijäksi. Vaikuttavuuden termi sisältyi tulokäsitteiden suositukseen, joka laadittiin kuntien toimintaan hieman aiemmin kuin valtionhallintoon (Suositus kunnallisen palvelutoiminnan tuloksellisuuden arvioinnissa käytettävistä käsitteistä ja niiden sisällöstä 1989). Vaikuttavuuden käsitteen käytössä ei säästely julkisten palvelujen uudistamisasiakirjoissa 1990-luvun talouslaman aikoihin.

Vaikuttavuuden käsite on levinnyt kaikkiin julkisen sektorin toimintoihin 2000-luvulla toiminnan ja talouden ohjauksen uudistusten sekä arvioinnin velvoitteiden lakisääteistymisen myötä. Valtion ja yliopistojen hallinnossa otettiin käyttöön yhteiskunnallisen vaikuttavuuden käsite (Tulosohjauksen terävöittäminen 2003, 24-27; Yhteiskunnallisen vaikuttavuuden arviointistrategian laatiminen 2004; Laki yliopistolain muuttamisesta 715/2004). Valtion tilinpäätösuudistuksessa tiukennettiin vaikuttavuustiedon hankinnan vaatimuksia: vuodesta 2005 alkaen ministeriöiden on esitettävä vuosittainen arvio hallinnonalansa yhteiskunnallisen vaikuttavuuden kehityksestä. Valtiontalouden tarkastusvirasto on jo huomauttanut tarkastuskertomuksissaan vaikuttavuustiedon puutteista hallinnonalojen tuloksellisuusraportoinnissa (esim. Sosiaali- ja terveysministeriön hallinnonalan tuloksellisuusraportointi eduskunnalle 78/2004).

Vaikuttavuuden käsitettä käytetään tällä hetkellä kaikkien hyvinvointivaltion toimintojen, palvelujen ja etuuksien ohjauksessa. Vaikuttavuuden arvioinnin kohteena voivat olla esimerkiksi sosiaalityö, lääkintähelikopteritoiminta,

yliopistojen alueellinen vaikuttavuus, maatalouden ympäristötuki, jätevero, poliisien päällystökoulutus, kirjastotoiminta, hankerahoitukset, mielenterveyskuntoutus, aikuiskoulutus, työvoimapolitiikka, potilaiden hoitoketjut, tietoverkot, opettajankoulutus, tietostrategiat, lastensuojelu, reumakirurgia, taidemuseotoiminta, teknologia, Kelan etuudet, dementia kuntoutus, tutkimustoiminta, perhekeskustoiminta ja kielikoulutus (Google-haku 5.4.2006, käsiteparilla "vaikuttavuuden arviointi" yhteensä 33 900 osumaa). Kyse ei ole vain toimijoiden arvioinnin halusta vaan myös vaatimuksista: lainsäädännössämme oli vuoden 2006 alussa ainakin 80 arviointiin velvoittavaa säädöstä. Ne koskevat vaikutusten ohella laadun, toiminnan, vaarojen, suunnitelmien toteutumisen, palvelun ja hoidon tarpeiden, oppisuuritusten, tuotteiden ja henkilöiden arviointia. Arviointivelvoitteet on säädetty pääosin 1990- ja 2000-luvuilla. (<http://www.finlex.fi/fi/>, luettu 2.1.2006.)

Vaikuttavuuden käsite kietoutuu monenlaisiin arvioinnin tarkoituksiin, käytäntöihin ja sanastoihin. Vaikuttavuutta arvioidaan ja kehitetään esimerkiksi arviointikapasiteettien vahvistamisen (evaluation capacity building), tieteellisiin näyttöihin (evidence-based) perustuvan politiikan ja käytäntöjen kehittämisen, tulohajauksen ja tilivelvollisuusarvioinnin (performance management), auditoinnin (auditing), riskien hallinnan, tietojohdamisen ja tiedonhallinnan (knowledge management) sekä valtaistamisen (empowerment) nimissä. Samalla vaikuttavuuden arvioinnin käytännöt ovat levinneet tutkijoiden keskuudesta kaikkien julkisen sektorin työntekijöiden toiminnan yhteyteen (Rajavaara 2006). Vaikuttavuuden avulla hallinnoinnista on tullut keskeinen keino yhteiskuntapolitiikan toteutuksessa ja työntekijöiden toiminnan moraalisisessa sääntelyssä. Voitaisiin puhua vaikuttavuushallinnasta eli pyrkimyksistä ohjata yhteiskuntaa ja hyvinvointivaltiota vaikuttavuustiedon avulla. Tässä artikkelissa pohditaan vaikuttavuuden arviointia hallinnon, politiikan, tutkimuksen ja ammattilaisten tietokulttuurien kysymyksenä. Miten erilaisissa tietokulttuureissa järkeillään vaikutuksia ja vaikuttavuutta? Mitä tietokulttuurien moninaisuus merkitsee tiedonmuodostuksen sekä vaikuttavuushallinnan kannalta?

TIETOKULTTUURIT

Auditointia tutkineen Michael Powerin (1997a, 4) mielestä arvioinnin laajentuminen ilmentää yhteiskunnallisten toimijoiden kulttuurisia käsitteitä siitä, miten yhteiskuntaa tulee kehittää ja hallita sen ongelmia. Vaikuttavuuden käsitteen läpimurto voidaan näin ollen ymmärtää tietynlaisten ajattelutapojen ja käytäntöjen murtautumisena ihmisten yhteiskunnalliseen mielikuvitukseen. Mielikuvituksen teemaa kehitellyt politiikan filosofi Cornelius Castoriadis (1987; 1997) painotti sitä, ettei yhteiskunnallisia toimintoja ja instituutioita voida ymmärtää vain rationaalisten ja tiedollisten elementtien avulla. Hän korosti yksilöiden ja yhteisöjen luovaa kuvittelukykyä, joka ilmenee ihmisten kielessä ja sosiaalisissa käytännöissä. Castoriadis kutsui kuvittelukyvyksi merkitysten artikuloitumisia, jotka eivät vastaa sellaisenaan ”rationaalisia” tai ”todellisia” elementtejä, mutta jotka todellistuvat niiden luomisen kautta. Sosiaalisesti tai yhteiskunnalliseksi ihmisten kuvittelukyvyyn tekee hänen mielestään se, että sen aikaansaamat merkitykset ovat olemassa kollektiivien ja yhteiskunnan välityksellä. Se, että yhteiskunnallisten instituutioiden ja käytäntöjen olemassaolo perustuu keskeisesti kuvittelukykyyn, ei Castoriadis mukaan tarkoita sitä, etteivät ne olisi todellisia. Pikemminkin kuvittelukyvyyn vuoksi ne ovat ”enemmän todellisia kuin todellisia” (Castoriadis 1987, 197).

Vaikuttavuus voi näyttää monoliittiselta käsitteeltä, joka miltei itseksensä organisoi toimintaa. Tosiasiassa vaikuttavuuden käsitteen sananvalta tulee todelliseksi vasta ihmisten yhteisöllisen mielikuvituksen ja toiminnan välityksellä. Vaikuttavuutta selvitetään ja kehitetään monenlaisissa tietokulttuureissa. Niitä voidaan kutsua tietokeskeisiksi tai episteemisiksi käytännöiksi tai kulttuureiksi. Tieteentutkija Karin Knorr Cetinan mukaan ne ovat käytäntöjä, joissa muodostetaan ja oikeutetaan tietoa. Niissä ei toimita rutiininomaisesti vaan luovasti, mielikuvitusta ja rakentavaa järkeilyä käyttäen. Knorr Cetinan mielestä tieteen käytännöt ovat episteemisistä käytännöistä ensiarvoisimpia. Tietoyhteiskunnan kehityksessä muiden kuin tieteen käytäntöjen merkitys on kuitenkin kasvanut. Tietoyhteiskunta ei merkitse vain lisääntyvien asiantuntijoiden, teknologien ja spesialistitulkintojen maailmaa. Se on erilaisten tietokulttuurien ja -käytäntöjen läpikäymistä yhteiskunta, jossa muodostuu monenlai-

sia rakenteita ja mekanismeja tiedon tuotannolle, välitykselle ja käytölle. Tietokeskeisten käytäntöjen tärkeys yhteiskunnassa johtuu siitä, että niiden välityksellä ohjataan muita yhteiskunnallisia käytäntöjä. (Knorr Cetina 1999, 5-8; 2001, 175-184.)

Arvioinnin käytännöt leviävät yli kansallisvaltioiden rajojen, ja ne ylittävät myös organisaatioiden, ammattiryhmien sekä tieteen- ja hallinnonalojen tietokulttuurien rajoja. Konkreettisesti vaikuttavuuden arviointi muotoutuu paikallisissa tiedonmuodostuksen ja hallinnan käytännöissä. Tiedonhalun kohteena vaikuttavuudella on niin kutsutun rajaobjektin piirteitä: se vaikuttaa ensi tuntumalta yhteiseltä käsitteeltä erilaisille tietokulttuureille, mutta tosiasiaa sen ympärille rakentuu monenlaisia merkityksiä, toimintoja ja sosiaalisia suhteita. Tällaiset rajaobjektit voivat toimia yhtä aikaa sekä toiminnan yhteensovittamisen että erilaisten intressien tyydyttämisen välineenä. Rajaobjekteille on tyypillistä, että ne ovat samanaikaisesti globaaleja ja paikallisia. (Star & Griesemer 1989.)

Yhdysvaltalaiset tutkijat ovat yrittäneet jäsentää arvioinnin ”yleistä logiikkaa” (Scriven 1997; vrt. Schwandt 1997). Oletuksena on, että arviointi voitaisiin ymmärtää esimerkiksi joidenkin piirteiden perusteella omanlaiseksi tietokulttuuriksi tai -käytännöksi. Nämä pyrkimykset liittyvät siihen, että Yhdysvalloissa ja Kanadassa arviointi on yliopistollinen oppiaine, toisin kuin meillä. Meillä ei ole oikeastaan olemassa ”arviointitutkijoita” vaan eri alojen asiantuntijoita ja tutkijoita, jotka saattavat olla kiinnostuneita arvioinnista. Akateemisista aloista arvioinnin kysymykset ovat esillä esimerkiksi hallintotieteissä, taloustieteissä, psykologiassa, yhteiskuntapolitiikassa, sosiaalipolitiikassa, sosiaalityössä, sosiologiassa, kasvatustieteissä, lääketieteissä ja hoitotieteissä. Liiketaloustieteissä sekä teknologia- ja insinööritieteissä kehitettyjen arviointilähestymistapojen hyödyntäminen on lisääntynyt julkisella sektorilla 1980-luvun lopulta alkaen. Oikeustieteissä on kiinnitetty huomiota arvioinnin kysymyksiin muun muassa julkisen sektorin tehokkuusvaateiden lisääntymisen johdosta sekä globalisaation ja Euroopan unionin muutettua lakien laadinnan toimintaympäristöä (Pöysti 1999; Tala 2001; 2005).

Toisin sanoen arviointia voidaan jäsentää kahdella tavalla. Ensinnäkin voidaan ajatella arviointi jollain tavoin omintakeiseksi tai itsenäiseksi tie-

tokäytäntöksi (evaluation as epistemic practice). Toinen tapa ajatella asiaa on, että katsotaan arviointia sisältävän monenlaisiin tietokulttuureihin julkisessa toiminnassa tai laajemmin yhteiskunnassa (evaluation in epistemic practices). Kun arviointia sisältyy monenlaisiin tietokulttuureihin ja sen on laajentunut kaikenlaisen hallinnon ja ammatillisen toiminnan osaksi, voi olla hyödyllisempää pohtia arviointia erilaisissa tietokulttuureissa kuin arviointia omana tietokulttuurinaan.

Ei voida olettaa, että vaikuttavuuden arvioinnin haasteet ymmärretään samalla tavoin eri tietokulttuureissa. Vaikuttavuuden tietokulttuurit ovat erilaisia paikallisesti, eri tieteenaloilla sekä erilaisissa hallinto- ja asiantuntijaympäristössä. Tietokulttuurit voivat tukea toisiaan ja kamppailla keskenään, ja lisäksi kunkin tietokulttuurin sisällä vallitsee monenlaisia vaikutustiedon haluja, käytäntöjä ja käsityksiä pätevästä tiedosta.

TIETOKULTTUURIT JA VAIKUTTAVUUSTIEDON OIKEUTUS

Tietokulttuurien erot ilmenevät jo suhtautumisessa vaikuttavuustiedon oikeutukseen. Lääketieteellisen vaikuttavuustiedon tarvetta ei yleensä kyseenalaisteta sen paremmin lääketieteen asiantuntijoiden kuin muidenkaan keskuudessa. Emme hyväksy epävarmuudella leikittelyä terveysasioissa, joten vaikuttavuustutkimus ja hoitosuosituksia koskevat kiistat ovat olennainen osa lääketieteitä ja terveydenhuollon toimintaa. Terveystutkimuksessa hankitaan vaikutustietoa monenlaisten tutkimusasetelmien kautta, mutta satunnaistettu kokeellinen tutkimus (RCT, Randomised Controlled Trial) on perusihanteena. Voidaan kuvata jonkin suorituksen vaikutusta eli seurannaisvaikutusta (effect). Haluttaessa selvittää tämän suorituksen vaikutuskykyä laboratorio-oloissa tai ideaalisessa hoitotilanteessa, ollaan kysymyksissä, joista käytetään käsitettä efficacy. Kun tarkastellaan, miten suoritus on toteutettavissa tehokkaasti käytettävissä olevien inhimillisten ja taloudellisten voimavarojen ja saavutettujen hyödyllisten vaikutusten suhteen, tullaan vaikuttavuuden (effectiveness) ilmiöihin. (Mälkiä ym. 2004, 6.) Se edellyttää yhteiskunnallisiin käytäntöihin pureutumista, jolloin kokeelliset menetelmät eivät ole yleensä mahdollisia.

Muilla yhteiskuntapolitiikan alueilla ei efficacy-luonteisten vaikutusten tarkastelu ole mahdol-

lista, sillä yhteiskunnallisia ilmiöitä ja käytäntöjä ei voida viljellä laboratorioissa ja siirtää niistä takaisin yhteiskuntaan. Siten vaikuttavuuden arvioinnin vaatimukset saavat aikaan hämmennystä useilla aloilla, ja vaikuttavuuden kysymysten oikeutus on epävarma asia. Vaikuttavuustiedon tarpeellisuuteen näytetään suhtautuvan ainakin neljällä tavalla. Ensinnäkin hallinnossa ja päätöksenteossa saatetaan rutiinomaisesti perätä vaikuttavuustietoa, eikä tunnisteta tällaisen tiedon epävarmuuksia, sen hankinnan vaikeuksia tai käytön rajoituksia. Toiseksi halutaan ”vastatietoa”: vaikuttavuuden näkökulma yritetään tuoda esiin esimerkiksi sosiaaliturvan hallinnoinnissa vallitsevaksi tulleen tuottavuuden näkökulman peittoamiseksi (Sinfield 2000; Sosiaaliturvan suunta 2005-2006, 2006). Kolmanneksi vaikuttavuuden kysymykset halutaan torjua, sillä pelätään hyvinvointipalveluja ja -etuuksia purettavan tällaisen tiedon perusteella. Neljänneksi vaikuttavuus katsotaan vain retoriseksi tai kuvitteelliseksi asiaksi (vrt. Sulkunen 2005). Arviointikäytännöt eivät ole kuitenkaan kuvitteellisia vaan ne ovat levinneet joka paikkaan.

TIETOKULTTUURIT JA JÄRKEILYTYYLIT

Erilaisissa tietokulttuureissa järkeillään vaikuttavuutta eri tavoin, ja nämä järkeilyt ovat muodostuneet historiallisesti. Eri tapoja järkeillä vaikuttavuutta voidaan tarkastella esimerkiksi sillä tavoin kuin Ian Hacking on hahmottanut erilaisten tieteellisten ja hallinnollisten järkeilyjen historiallista muodostumista. Hacking on ottanut ideansa kehittämäänsä analyyttiseen työkaluun, jota hän kutsuu järkeilytyyliksi (styles of reasoning), tieteenhistorioitsija A.C. Crombielta. Toisin kuin Crombie, hän ei puhu ajattelutyyleistä vaan korostaa sitä, että järkeilytyylit ovat enemmän kuin ajattelua: niillä on myös materiaallinen puoli, joka ilmenee välineissä, laitteistoissa ja käytännöissä. Järkeilytyylit ovat tutkimuksen, luokittelun ja nimeämisen käytäntöjen kokonaisuuksia, jotka syntyessään suuntaavat tiedonmuodostusta uusiin kohteisiin ja kysymyksiin. (Hacking 2002a, 180-181; 2004, 6-7.)

Hackingin mukaan jokaisen järkeilytyylin sisällä kehittyvät todistamisen ja osoittamisen kriteerinsä, ja ne määrittävät totuusehdot tiedon alalle, jolla niitä sovelletaan. Siten järkeilytyyli on enemmän kuin menetelmien kokoelma, jonka

avulla hankitaan uusia tosiasioita. Järkeilytyylit eivät sinällään ole objektiivisia, vaan ne määrittävät ehdot sille, mikä katsotaan objektiiviseksi. Kun jokainen järkeilytyyli luo omat totuuskriteerinsä, niistä tulee itsensä oikeuttavia. Olennaista on siksi tarkastella tapoja, joilla järkeilytyylit tarjoavat vakaata tietoa sekä niiden sisältämiä objektiivisuuden standardeja. (Hacking 2002a, 180-181; 198; 2002b, 4.)

Jokainen järkeilytyyli tuo mukanaan uusia asioita, kuten uudenlaiset tiedon halun kohteet, uudet käsitykset evidenssistä, lausumat, jotka ovat ehdolla tosiksi tai epätosiksi, säännönmukaisuudet ja mahdollisuudet (Hacking 2002a, 189). Jonkin erityisen järkeilytyylin järkeä itseään ei voi kumota, vaan kaikenlainen kumoaminen tapahtuu sen sisällä. Uuden järkeilytyylin ilmaantuminen ei välttämättä kyseenalaista toisen pätevyyttä jollain toisella alueella. Järkeilytyylit luovat keston ja pysyvyytensä erilaisin vakautuskeinoin, ja olennaista on myös tiedonmuodostuksen institutionaalinen vakauttaminen. Se, että erilaiset järkeilytyylit voivat elää rinnakkain, erottaa Hackingin käyttämän käsitteen Kuhnin paradigman käsitteestä. (Lehtonen 2003, 6.) Suomessa tunnetaan myös hyvin Egon Guban ja Yvonna Lincolnin (1989) arvioinnin historian sukupolvitarkastelu. Sukupolven metafora on myös eri asia kuin järkeilytyyli: sukupolvet syntyvät ja kuolevat, ja uudet sukupolvet tulevat tilalle. Järkeilytyyli ei välttämättä katoa minnekään, vaikka sen merkitys voi vaihdella ja vakautuksen mekanismit voivat heikentyä.

Vaikuttavuuden arviointia tehdään monenlaisien järkeilytyyliin perustalta. Tämä johtuu jo siitä, että vaikuttavuuden arviointia kohdennetaan monenlaisiin asioihin, kuten kokonaisiin instituutioihin, yhteiskunnallisiin uudistuksiin, hankkeisiin, interventioihin, palveluihin tai etuuksiin. Toiseksi vaikuttavuuden arviointi sisältää arvottamista, ja sitäkin voidaan tehdä monella tavalla. Kolmanneksi vaikuttavuuden arviointia tehdään sekä tieteellisin että ei-tieteellisin menettelyin. Neljänneksi vaikuttavuuden arvioinnin käyttötarkoitukset suuntaavat tapoja jäsentää ja selvittää vaikutuksia. Tässä tarkastellaan joitakin yleisimpiä vaikuttavuuden järkeilytyylejä ja niiden asemaa tietokulttuureissa.

VAIKUTTAVUUDEN JÄRKEILYTYYLEJÄ

Tavoitteita mittaava järkeilytyyli. Tavoitelähtöistä vaikuttavuuden arviointia alettiin kehittää Suomessa 1960-luvulla, jolloin se vastasi aikakauden keskitetyn suunnitteluohjauksen tapaa. Toimintojen tavoitteet olivat pitkään julkisessa hallinnossa tavallisin vaikuttavuuden arvioinnin lähtökohta. Tavoitelähtöistä arviointia tehdään edelleen suunnittelun käytännöissä ja tulosohtuuskulttuureissa. Tosin 1960-1970-luvuilla aikaansaadut suunnitteluinstituutiot on pääosin purettu. Tavoitelähtöisessä vaikuttavuuden arvioinnissa selvitetään sitä, miten hyvin toiminnalle asetetut tavoitteet on saavutettu. Tavoitteiden saavuttamisen aste on toiminnan hyvyyden tai onnistuneisuuden kriteeri. Taustalla on rationaalisen päätöksenteon malli: suunnittele ensin ja tee sitten. (Sinkkonen & Kinnunen 1994, 82-83.)

Tavoitelähtöisen vaikuttavuusjärkeilyn etuna on selkeys, täsmällisyys, konkreettisuus, yksinkertaisuus ja ymmärrettävyys. Siihen turvautuminen on mahdollista, kun on toiminnan päämäärät ovat selkeät ja mitattavia, ja kaikki asianosaiset ovat niistä yhtä mieltä. (Sinkkonen & Kinnunen 1994, 83-84; vrt. Meklin 2001, 99-100; Vuorela 1990, 40-45.) Julkisessa hallinnossa ja yhteiskuntapolitiikassa selkeä tavoitteellisuus toteutuu kuitenkin vain harvoin, ja arvioinneissa on vaikeaa tukeutua tavoitteisiin niiden määrittelemättömyyden epämääräisyyden, abstraktisuuden, laajuuden, ristiriitaisuuden tai heikon mitattavuuden vuoksi. Toiminnoille ei usein voida tai haluta asettaa yksityiskohtaisia tavoitteita. Julkiset organisaatiot eivät muutenkaan voi valita toimintansa tavoitteita ja keinoja, vaikka niiden vaikutusmahdollisuudet ovat ehkä kasvaneet (Vartola 2004, 99). Toinen tavoitejärkeilyn ongelma liittyy siihen, keiden tavoitteiden avulla vaikuttavuutta tarkastellaan. Tavoitteenasettelu on arvopohjaista ja subjektiivista. Usein tavoitteet otetaan laeista, asetuksista ja muista virallisista normeista. Tähänkin menettelyyn sisältyy ongelmia: lait vanhentuvat tai ne voivat olla epätarkoitusten mukaisia ja ristiriitaisia kansalaisten tarpeiden kannalta. Kolmanneksi tavoitemallin heikkoutena on keskittyminen toivottuihin vaikutuksiin ja sivuvaikutusten unohtaminen (Sinkkonen & Kinnunen 1994, 84-85).

Evert Vedungin (2004, 16-17) mielestä tavoitelähtöistä arviointia on jopa liikaakin moitittu: edustuksellisessa demokratiassa se voi olla usein

tarpeellista. Sen avulla kansalaiset ja päätöksentekijät voivat seurata, edetäänkö asioissa asetettujen päämäärien suuntaisesti. Esimerkiksi jos tavoitteena on työllisyysasteen parantaminen tiettyyn prosenttiin, tämän tunnusluvun muutos ilmaisee kaikille yksinkertaisesti, miten asiassa onnistutaan.

Tarpeita mittaava järkeilytyyli. Tarvelähtöinen vaikuttavuuden arviointi on tavoitteita mittaavan arviointijärkeilyn aikalais- ja sukulaismuoto. Siihen on turvauttu esimerkiksi yhteiskuntasuunnittelussa sekä sosiaalipolitiikan tutkimuksessa, mutta sen merkitys näyttää vähentyneen. Tarvelähtöisessä vaikuttavuuden arvioinnissa etsitään vastausta siihen, vastaako ohjelma, palvelu tai etuus asiakkaan tarpeita. Sen tarkasteluun voidaan paneutua kansalais- tai asiakaslähtöisesti tai järjestelmä- tai asiantuntijakeskeisesti. Lisäksi voidaan hyödyntää esimerkiksi yhteiskuntatutkimuksessa kehitettyjä tarveteorioita. (Sinkkonen & Kinnunen 1994, 93; vrt. Vuorela 1990, 50-52.)

Tarvelähtöinen järkeilytyyli voi olla myös monitahoarvioinnin taustalla, josta puhutaan usein monitoimija-arviointina tai stakeholder-mallina. Siihen on tukeuduttu esimerkiksi työelämän ja organisaatioiden kehittämisessä sekä kuntien palvelujen arvioinnissa. Siinä arviointiin osallistetaan kaikki asiaan liittyvät sidosryhmät. Näitä voivat olla asiakkaat, sukulaiset, johto, keski-johto, asiakaspalvelutyöntekijät, asiakasjärjestöjen edustajat, veronmaksajat ja päätöksentekijät. Arvioinnista vastaava taho huolehtii siitä, että kaikkien tahojen näkemykset tulevat huomioon-otetuiksi. (Sinkkonen & Kinnunen 1994, 93-96; Vartiainen 1994.) Tätä mallia on perusteltu osallistuvan tai deliberatiivisen demokratian näkökulmista. Yhteiset keskustelut nähdään hyödyllisiksi, sillä niiden avulla eri tahot voivat tarkastella ja arvioida uudelleen omia näkemyksiään ja tavoitteitaan. Toisaalta malli on sosiaalisesti mutkikas ja voimavaroja vaativa. Arviointiin osallistuvien yksimielisyyttä ei usein muodostu, jolloin arvioinnin johtopäätösten teko voi olla vaikeaa. (Vedung 2004, 26-28.)

Kokeellinen järkeilytyyli. Kun vaikuttavuus samaistetaan vaikutuksiin, määritellään vaikuttavuus erotukseksi toimimisen ja toimimattomuuden välillä eli jonkin intervention toteuttamisen tai toteuttamatta jättämisen välillä. Näin määriteltyinä vaikutukset tarkoittavat olemassa olevan tilan muuttamista, säilyttämistä tai estämistä.

(Meklin 2001, 107-108.) Tällainen tapa ymmärtää vaikuttavuus edellyttää ajallista vertailua sekä kokeellisia tai puolikokeellisia tutkimusasetelmia. Hankkeissa käytetään koe- ja kontrolliryhmiä tai tehdään tilastollista vertailua. Niissäkin voi olla kyse tavoitelähtöisestä arvioinnista: kokeellisen intervention tavoitteet on usein ennalta määritelty (ks. Vuorela 1990, 44-45).

Kokeellisen arvioinnin historiaa voidaan jäljittää 1400-luvulle saakka. Ann Oakleyn (2000, 145-146) mukaan satunnaistetut kokeelliset tutkimukset yleistyivät lääketieteissä 1930- ja 1940-luvulla. Julkiseen hallintoon ja yhteiskuntapolitiikkaan liittyvän kokeellisen arvioinnin kultakausi sijoittuu 1960- ja 1970-lukujen Yhdysvaltoihin. Suomessa kokeellista tutkimusta on tehty yhteiskuntapolitiikan alueella 1950-luvulta alkaen. Tunnetusti Pekka Kuusi oli sekä yhteiskuntasuunnittelun että kokeellisen tutkimuksen sekä niiden toisiinsa liittämisen puolestapuhuja. Alkoholipolitiikka oli ensimmäisiä kokeellisen yhteiskuntatutkimuksen sovellusalueita, ja 1970-luvulla kokeellista arviointia alettiin ottaa käyttöön esimerkiksi Kansaneläkelaitoksen sosiaali- ja terveysturvan tutkimuksessa.

Kokeellinen vaikuttavuustutkimus juontaa juurensa loogisesta empirismistä, jossa korostetaan ilmiöiden lainalaisuuksien selvittämistä ja yhteiskunta- ja luonnontieteiden keskinäistä samankaltaisuutta. Sitä tehdään laajasti lääketieteissä ja hoitokäytäntöjen tutkimuksessa. Yhteiskuntapolitiikassa ja julkisessa hallinnossa kokeellisella arvioinnilla on nähty olevan suppeasti sovellusmahdollisuuksia. Ensinnäkin yhteiskunnalliset olosuhteet soveltuvat vain poikkeustapauksissa kokeellisesti tutkittaviksi. Toiseksi kokeellisen järkeilyn menetelmin ei ole mahdollista saada tietoa monitahoisista, useista toimenpiteistä muodostuvista kokonaisuuksista. Kolmanneksi koe- ja kontrolliryhmien käyttöä on pidetty monissa tilanteissa epäeettisenä menettelynä. Neljänneksi kokeellisin menetelmin toteutetut hankkeet ovat kalliita ja pitkäkestoisia, ja ne toteutuvat etäällä yhteiskunnallisen päätöksenteon tilanteista. Viidenneksi kokeellisen vaikuttavuusjärkeilyn on katsottu olevan menetelmälähtöistä ja epäteoreettista sekä vähättelevän ihmisten oman toiminnan merkitystä. (Oakley 2000.)

Mekanismeihin keskittyvä järkeilytyyli. Lain-säädäntö, uudistusohjelmat, palvelut ja etuudet sisältävät julkilausuttuja ja julkilausumattomia oletuksia yhteiskunnallisista tilanteista, ongel-

mista ja riskeistä sekä niihin puuttumisen mahdollisuuksista. Ne perustuvat sekä tietoon että uskoihin. Palvelujen tai etuuksien saajien kannalta ei ole olemassa jotain keskimääräistä vaikuttavuutta. Pikemminkin interventiot, palvelut ja etuudet vaikuttavat eri tavoin erilaisten ihmisten tai väestöryhmien tilanteessa.

Tästä syystä on lisääntyvästi kiinnostuttu väliintulojen taustaoletuksista ja mekanismeista. Niiden selvittämiseen on paneuduttu esimerkiksi sosiaalipalvelujen, kuntoutuksen, psykoterapian ja sosiaalityön ammatillisissa kulttuureissa. Uudistusten, ohjelmien, palvelujen ja etuuksien "juonen" selvittelystä puhutaan esimerkiksi julki- ja käyttöteorioiden sekä interventio- tai muutosteorioiden avulla. Tällaiset arviointijärkeilyt ovat kiteytyneet esimerkiksi realistisen arvioinnin tai teoriaperustaisen arvioinnin lähestymistavoiksi (Robson 2001, 110-120; Rostila 2001; Sanderson 2003, 335-337; Dahler-Larsen 2005). Huey-Tsou Chen (1990) on kehittänyt myös teorialähtöistä (theory-driven) arviointia, joka on eri asia kuin teoriaperustainen (theory-based) arviointi. Chenin kehittelemä arviointiotte on tulevaisuuteen suuntautunut, ja siinä painotetaan normatiivisia teorioita. Sen sijaan teoriaperustaisessa arvioinnissa korostetaan toteutuneiden ohjelmien ja politiikkojen mekanismien selvittämistä.

Realistisen arvioinnin tehtävänä pidetään sen selvittämistä, mitkä ovat johonkin väliintuloon sisältyvät vaikutusmekanismit ja miten ne toimivat. Samalla kiinnitetään huomiota väliintulojen toimintaympäristöön eli olosuhteisiin, joissa vaikutuksia ilmenee. Toimintaympäristöä ja mekanismeja koskevia toimijoiden taustaoletuksia selvittelemällä halutaan tehdä päätelmiä siitä, millaisia mahdollisuuksia jollain toimintaohjelmalla, palvelulla tai etuudella on saavuttaa tavoitteet, jotka sille on asetettu. Mitä osuvammin intervention taustaoletukset vastaavat yhteiskunnallisia olosuhteita ja kohderyhmän tarpeita, sitä suuremmalla syyllä voidaan odottaa myönteisiä vaikuttavuustuloksia. Karkeasti ottaen kyse on kaavasta: toimintamekanismi + toimintaympäristö = vaikuttavuus. (Pawson & Tilley 1997, xv ja 53-82.) Realistisen arvioinnin toivotaan avaavan kokeellisten arviointien puutetta, niin kutsuttua arvioinnin "mustan laatikon" ongelmaa: mikä toimii kenellekin missä olosuhteissa, ja miksi jotkut vaikutukset tapahtuvat tai jäävät tapahtumatta.

Vuorovaikutteinen järkeilytyyli. Toimijoiden

vuorovaikutukseen keskittyviä arviointijärkeilyjä, kuten osallistavaa, kehittäväää, dialogista tai valtaistavaa arviointia on liitetty muun muassa osallistavan demokratian ja deliberatiivisen demokratian kehittämispyrkimyksiin (Ahonen 1999; Vedung 2004, 33-36; vrt. Fetterman 2001). Tälle alueelle on kehittynyt lukuisia arviointiotteita, joiden painotukset poikkeavat hieman toisistaan. Hankkeiden käytännön toteutuksessa painotuserot saattavat olla vähäisiä. Tietokulttuureita, joissa toimintaa on järkeily näiden lähestymistapojen avulla, ovat esimerkiksi kuntien palvelutoiminta, työelämän ja organisaatioiden kehittäminen, yhdyskuntien kehittäminen, sosiaalityö, kasvatus ja kuntoutus.

Näihin arviointijärkeilyihin tukeudutaan yleensä toiminnan kehittämishankkeiden ja itse-arviointien yhteydessä (ks. Seppänen-Järvelä 1999). Niihin tukeuduttaessa ei kiinnitetä pääasiallista huomiota toiminnan vaikuttavuuteen. Sen sijaan painotetaan arviointiin osallistuvien ihmisryhmien, kuten työntekijöiden, asiakkaiden tai kuntalaisten osallistumisen, vaikutusmahdollisuuksien tai itsemääräämisoikeuden lisäämistä. Vaikuttavuuden otaksutaan lisääntyvän erilaisten välillisten tulosten kautta. Jaakko Harkon, Riitta Haverisen ja Juha Koiviston (2005) kirjallisuuskatsauksen perusteella osallistava arviointi voi esimerkiksi lisätä työntekijöiden yleistä kyvykkyyttä kehittämistyöhön: arviointi voi parantaa palvelujen ja väliintulojen dynamiikan ymmärrystä, lisätä paikallista tietoa ja aktivoida toimijoiden vuorovaikutusverkostoja.

Tieteellisiin vaikuttavuusnäyttöihin perustuva järkeilytyyli. Näyttöperustaisessa vaikuttavuusjärkeilyssä tukeudutaan systemaattisiin katsauksiin, joihin kootaan vaikuttavuustietoa erilaisilta toiminnan alueilta. Tämä "evidenssiilike" (Oakley 2002, 279) on laajentunut 1990-luvulta alkaen, ja siinä halutaan saada aikaan vaikuttavuustietoon perustuvia toimintakäytäntöjä ainakin terveydenhuoltoon, sosiaalialalle, mielenterveys-työhön, kasvatukseen, terapiaan, työvoimapolitiikkaan ja kuntoutukseen. Lisäksi haluttaisiin lisätä yhteiskuntapoliittisen päätöksenteon näyttöperustaisuutta. Iso-Britanniassa pääministeri Tony Blairin hallitus on tukenut näyttöperustaisen politiikkojen ja käytäntöjen aikaansaamista, ja kehitystyön tueksi siellä on perustettu arviointi-tyksiköitä, kuten EPPI-Centre, EvidenceNetwork ja Social Care Institute (SCIE). Näyttöihin perustuvia politiikkoja ja käytäntöjä kehitetään myös

Yhdysvalloissa, Kanadassa, Australiassa ja Pohjoismaissa. (Koivisto 2005; Blackmore 2005; Laforest & Orsini 2005.)

Näyttöperustaisen vaikuttavuuden kehittämisen on lähtöisin lääketieteen tietokulttuureista, joissa sitä toteutetaan sekä kansallisena että ylikansallisena toimintana niin kutsuttuna Cochrane-yhteistyönä. Sosiaalialan, kasvatuksen, koulutuksen ja työvoimapolitiikan alueella toimii samantapainen Campbell -yhteistyöverkosto, jonka toimintaan osallistutaan Suomesta. Lisäksi meillä muun muassa kansalliseen sosiaalialan kehittämissuunnitelmaan (2004-2007) sisältyvän Hyvät käytännöt -ohjelman avulla yritetään edistää sosiaalialan menetelmien perustamista tutkittuun ja arvioituun tietoon.

Auditoiva järkeilytyyli. Eri tieteenalojen, ammattiryhmien tai hallinnon tietokulttuureissa voi esiintyä mieltymyksiä turvautua tietynlaisiin arviointijärkeilyihin, mutta auditoiva järkeily ylittää tietokulttuurien rajat. Toisaalta se ei ole erityisen suosiossa julkisen hallinnossa työskentelevien asiantuntijoiden keskuudessa. Sen laajeneminen liittyy uuden julkisjohtamisen, niin kutsutun "new public management" (NPM) -ohjaukseen omaksumiseen julkisessa toiminnassa. Huomion kohteena on tullut manageristinen hallinnointi (management) eikä niinkään politiikka (policy). Uusmanagerismissa painotetaan taloudellisia yllykkeitä, tuotosten ja suoritteiden mittaamista, markkinaistamista, kilpailuttamista, määräaikaista sopimuksia sekä kustannusten leikkauksia. Työntekijöiden välittömästä valvonnasta halutaan siirtyä tulosten kautta tapahtuvaan valvontaan. Kansalaisten asemaa veronmaksajina ja hyvinvointietuuksien kuluttajina korostetaan ja vaaditaan julkiselta toiminnalta riittäviä vastikkeita veronmaksajien uhrauksille. (Esim. Meklin 2002; Bevir ym. 2003; Salminen 2004.)

Auditointijärkeily liittyy toimijoiden, kuten palvelujen tilaajien ja tuottajien, keskinäisen luottamuksen kysymyksiin ja tilivelvollisuuteen. Michael Power puhuu auditoinnin käytännöistä "todentamisen rituaaleina" (vrt. Strathern 2000). Powerin mukaan auditoinnissa on kyse kontrollin kontrollista eli siitä, onko jokin ohjaus- tai hallintajärjestelmä käytössä ja miten se toimii. Auditoinnissa ei keskitytä organisaation yksittäisiin toimintoihin. Etenkin liiketaloustieteissä kehitetyt, kokonaisvaltaiset lähestymistavat ovat tästä syystä tulleet keskeisiksi auditoinnin välineiksi. (Power 1994; 1997a; 1997b.) Vaikuttavuus sisältyy

olennaisesti uusmanageristiseen hallinnointiin ja sitä tukevaan auditoiva järkeilyyn. Epävarmaa on kuitenkin, missä määrin uusmanageristinen hallinto sisältää tosiasialisesti toiminnan vaikuttavuuden kehittämisen tai arvioinnin yllykkeitä. Juha Vartolan (2004, 112) mielestä manageristisessa johtamisessa toimintayksikköjen menestyksen ja tuotosten tavoittelu syrjäyttää vaikuttavuuden.

TIETOKULTTUURIT JA VAIKUTTAVUUSTIEDOLLA HALLINNOINTI

Vaikuttavuuden arvioinnin pohdinta tietokulttuurin ja järkeilytyylin käsitteiden avulla osoittaa sen miten moninaisesti asioihin vaikuttavuuden käsitteellä voidaan viitata. Vaikuttavuus voi liittyä kuvatuolisesti esimerkiksi tavoitteiden saavuttamiseen, tarpeiden tyydyttämiseen, koeryhmän tilanteen muutokseen, toimintamekanismeihin, yhteistoiminnallisuuteen ja vuorovaikutukseen, systemaattisiin näyttökatsauksiin tai tilivelvollisuuden tietojärjestelmiin. Edellä mainitut vaikuttavuuden arvioinnin järkeilytyylit eivät edes ole ainoita, joita on aikaa myöten kehittynyt hyvinvointivaltiollisiin tietokulttuureihin. Lisäksi on muodostunut esimerkiksi toimeenpanoon ja prosesseihin sekä verkostoihin ja niiden hallintaan liittyviä vaikutusjärkeilyjä. Tässä yhteydessä ei ole tarkasteltu keinoja ja instituutioita, joiden avulla vaikuttavuusjärkeilyjä on vakiinnutettu tietokulttuureihin. Näyttää kuitenkin siltä, että yliopistojen, valtion tutkimuslaitosten ja professionaalisen ohjauksen asema arviointitiedon vakiinnuttamisessa on muuttunut, kun näiden institutionaalisten vakautusmekanismien rinnalle on muodostunut monialaisia asiantuntijaverkostoja, pieniä toimintayksiköjä sekä "arviointiteollisuutta".

Vaikuttavuustiedon merkitystä hyvinvointivaltion hallinnoinnin tukena painotetaan tällä hetkellä runsaasti. Vaikuttavuustiedon avulla ohjautuvan julkisen organisaation kehittäminen on kuitenkin mutkikas asia. Tietokulttuurien ja järkeilytyylien moninaisuudesta johtuen vaikuttavuutta ei voida määritellä, selvittää tai kehittää vain yhdellä "oikealla" tavalla. Lisäksi vaikuttavuus on vain yksi tuloksellisuuden osatekijä: organisaatioiden onnistuminen edellyttää nähtävästi eri tuloksellisuuden osatekijöiden tasapainoa. Kun tehokkuutta korostetaan, kärsii vaikuttavuus. Jos

vaikuttavuutta painotetaan, toiminnan tehokkuus ehkä kärsii, mutta pitkän aikavälin tulokset saattavat olla paremmat.

Koska vaikuttavuustieto kiinnittyy monin tavoin politiikka-, hallinto- ja asiantuntijakulttuureihin, siitä voi helpolla tulla myös tietokulttuurien keskinäisten ja sisäisten voimainkoetusten kohde. Päädyttäisiin ehkä sekasortoon, jos valtion ja kuntien talousarviot alettaisiin laatia järjestelmällisesti toimenpiteiden vaikuttavuusnäyttöjen perusteella. Vaikutuksia koskevat kiistat kärjistyisivät kestäättömiksi, ja vaikuttavuustuloksia voitaisiin hankkia epäeettisin menettelyin. Kansalaisilla ei olisi oikeusturvaa, sillä vaikuttavuutta kyettäisiin osoittamaan niissä tietokulttuureissa, joilla on eniten arvovaltaa, kiinnostusta ja voimavaroja arviointiin. Esimerkiksi Ilpo Vilkkumaa pohtii tietokulttuurien merkitystä kuntoutuksen arvioinnin osalta seuraavasti: "Epätoivottava, joskin todennäköinen on tilanne, jossa lääketieteen tutkimusparadigman mukaisesti saadaan kiistämätön tulos, että jokin hoito- tai kuntoutustapa on hyödytön ja että siksi siihen ei kannata käyttää julkisia varoja (äärimmilleen vietynä ei edes julkisin varoin koulutettua lääkäriyövoimaa), mutta samaan aikaan jossain toisessa tieteellisessä paradigmassa (psykologiassa tai sosiaalitieteessä) sama hoito tai kuntoutus nähdään hyvin perusteltuna." (Vilkkumaa 1999, 88.) Suomessa ei vielä käydyä ankaria "vaikuttavuussotia", mutta tiedon pätevyys ehdot voisivat osoittautua keskeiseksi kiistakapulaksi hyvinvointivaltion tietokulttuureissa, jos vaikuttavuustiedolla hallinnointia vielä tehostettaisiin.

Vaikuttavuuden järjestytyliin yhdenmukaistuminen saattaisi olla yhtä lailla mahdollista: eniten yhteiskunnallista valtaa omaavien tietokulttuurien arviointimenettelyt laajenisivat. Erilaisiin vaikuttavuusjärjestytyliin ei nykyisinkään tukeuduta vain sillä perusteella, miten ne katsotaan soveltuviksi erilaisiin arvioinnin kohteisiin ja tarkoituksiin. Pikemminkin ratkaisuihin voivat vaikuttaa tietokulttuurien institutionaaliset ja sosiaaliset tekijät, niiden sisäiset ja keskinäiset voimasuhteet ja niissä omaksutut pätevän tiedon käsitykset. Kun vaikuttavuuden järjestytyliit siirtyvät alalta toiselle, ne kulkeutunevat vahvemmissa tietokulttuureista heikompiin ja vähemmän arvostettuihin. Esimerkiksi tieteellisiin näyttöihin perustuvien politiikkojen ja käytäntöjen kehittäminen on nostanut uudelleen näkyväksi lääketieteessä keskeisen kokeellisen tutkimuksen. Kokeellisten arviointien

mahdollisuuksia on alettu pohtia jälleen esimerkiksi työvoimapolitiikan tutkimuksessa (Hämäläinen & Uusitalo 2006).

Vaikuttavuustiedolla hallinnoinnin epävarmuuksista on keskusteltu runsaasti viime aikoina juuri tieteellisiin näyttöihin perustuvien politiikkojen ja käytäntöjen kehittämisspyrkimysten yhteydessä. Näyttöperustaisuuden vaatimus on nostanut pohdittavaksi visaisia kysymyksiä. Asiantuntijoilla on osoittautunut olevan kilpailevia kantoja siitä, mitä evidenssi tarkoittaa, millaisista asioista näyttöjä tulee hankkia, ja miten tieto kelpuutetaan vaikuttavuusnäytöksi. Lääketieteellisen ihanteen mukaisesti näytöksi haluttaisiin yleisimmin hyväksyä vain satunnaistetun kokeen avulla saatu tieto. On myös yritetty laittaa erilaiset näyttöperustat - satunnaistetuista koeasetelmista asiantuntijanäkemyksiin - paremmuusjärjestykseen. Lisäksi on vaadittu luopumista tällaisista tiedon pätevyyttä arvottavista asetelmista. (Oakley 2002; Arnkil & Seikkula 2005.)

Ei ole olemassa tuomaria sille, mikä on pätevää vaikuttavuustietoa. Vaikka tällaista tietoa halutaan hyvinvointivaltiollisen hallinnan tueksi, on pidettävä mielessä, että siihen sisältyy myös uudenlaisia epävarmuuksia. Ulrich Beckin (1992) mukaan siirtymä toiseen moderniin ja refleksiiviseen modernisaatioon on merkinnyt sitä, että tiedosta eikä niinkään tietämättömyydestä on tullut epävarmuuden ja riskien lähde. Yhtä lailla vaikuttavuuden tietoon sisältyy sekä arvokkaita ja toivottuja että arvaamattomia ja hallitsemattomia piirteitä.

Tiedon tuottamiseen ja käyttöön liittyvien epävarmuuksien vuoksi arviointitieto haluttaisiin saada yhä kiinteämmin ohjauksen ja sääntelyn kohteeksi (esim. Kohti julkisten palvelujen yhteistä arviointia, 2001). Kun arviointi nähdään laajasti erilaisten tietokulttuurien asiana eikä omana tietokäytäntönään, huomataan, ettei sen kokonaisvaltainen ohjaus ole mahdollista. Vaikuttavuustiedon hallinta on yhtä moniselkoinen asia kuin se, miten vaikuttavuustietoa käytetään hallinnassa. Ehkä juuri vaikuttavuuden tietokulttuurien runsaus ja monenkirjavat vaikuttavuusjärjestytyliit takaavat sen, että arviointi säilyy yhteiskunnassa renkinä eikä ota isännän paikkaa.

LÄHTEET

- Ahonen, Pertti: Jotta arviointi olisi jotain, se ei voi olla kaikkea. *Hallinnon tutkimus* 17 (1998): 4, 257-261.
- Ahonen, Pertti: Arviointi kansanvaltaisena käytäntönä. Julkaisussa Eräsaari Risto & Lindqvist, Tuija & Mäntysaari, Mikko & Rajavaara, Marketta (toim.) *Arviointi ja asiantuntijuus*, Gaudeamus, Helsinki 1999, 119-142.
- Arnkil, Tom Erik & Seikkula, Jaakko: Tuomariksi houkuteltu tiede. Julkaisussa Hänninen, Sakari & Karjalainen, Jouko & Lahti, Tuukka (toim.) *Toinen tieto. Kirjoituksia huono-osaisuuden tunnistamisesta*, Stakes, Helsinki, Gummerus Kirjapaino Oy, Saarijärvi 2005, 52-75.
- Beck, Ulrich: *Risk society. Towards a New Modernity*, Sage Publications, London, Newbury Park, New Delhi 1992.
- Bevir, Mark & Rhodes R. A. W. & Weller, Patrick: Traditions of governance: Interpreting the changing role of the public sector. *Public Administration* 81 (2003): 1, 1-17.
- Blackmore, Jill: Is it only 'What works' that 'Counts' in New Knowledge Economies? Evidence-based Practice, Educational Research and Teacher Education in Australia. *Social Policy & Society* 1 (2002): 3, 257-266.
- Castoriadis, Cornelius: *The Imaginary Institution of Society*, Polity Press, Cambridge 1987.
- Castoriadis, Cornelius: Magma: tutkielmia yhteiskunnan imaginaarista instituutioista, Artikkelit valinnut ja suomentanut Mikko Keinänen, Hanki ja jää, Helsinki 1997.
- Chen, Huey-Tsou: *Theory-driven evaluations*, Sage, Newbury Park 1990.
- Dahler-Larsen, Peter: *Vaikuttavuuden arviointi, Hyvät käytännöt, Menetelmäkäsikirja*, Stakes, FinSoc, Helsinki 2005.
- Fetterman, David M.: *Foundations of Empowerment Evaluation*, Sage Publications, Thousand Oaks 2001.
- Hacking, Ian: *Historical Ontology*, Harvard University Press, Cambridge, Massachusetts, London, England 2002a.
- Hacking, Ian: Inaugural lecture: Chair of Philosophy and History of Scientific Concepts at the Collège de France, 16 January 2001. *Economy and Society* 31 (2002b): 1, 1-14.
- Hacking, Ian: *The Taming of Chance*, Eighth printing, first published 1990, Cambridge University Press, Cambridge 2004.
- Harkko, Jaakko & Haverinen, Riitta & Koivisto, Juha: Alustava kirjallisuuskatsaus osallistavan arvioinnin vaikutuksista sosiaalityössä, Työpapereita 5, Stakes, Finsoc, Helsinki 2005.
- Hämäläinen, Kari & Uusitalo, Roope: Kannattaisi kokeilla. Kokeelliset menetelmät työvoimapolitiisten toimenpiteitten vaikutusten arvioinnissa, Työpoliittinen tutkimus 285, Työministeriö, Helsinki 2006.
- Kohti julkisen palvelujen yhteistä arviointia, Tuottavuudella tulevaisuuteen, Keskustelualoite, Suomen Kuntaliitto, Sisäasiainministeriö, Valtiovarainministeriö, Helsinki 2001.
- Koivisto, Juha: Tiedon kerääminen ja systematisointi sosiaalialalla, Katsaus viiden ulkomaisen arviointiinstituutin strategioihin. *FinSoc arviointiraportteja* 1, Stakes, Helsinki 2005.
- Knorr Cetina, Karin: *Epistemic Cultures. How the Sciences make knowledge*, Harvard University Press, Cambridge, Massachusetts, London, England 1999.
- Knorr Cetina, Karin: *Objectual practice*. Julkaisussa Schatzki, Theodore R. & Knorr Cetina, Karin & von Savigny, Eike (eds.) *The Practice Turn in Contemporary Social Theory*, Routledge, London, New York 2001, 175-188.
- Laforest, Rachel & Orsini, Michael: Evidence-based Engagement in the Voluntary Sector: Lesson from Canada. *Social Policy & Administration* 39 (2005): 5, 481-497.
- Laki valtion talousarviosta annetun lain muuttamisesta 19.12.2003 / 1216.
- Laki yliopistolain muuttamisesta 30.4.2004 / 715.
- Lehtonen, Turo-Kimmo: Historiallisesta ontologiasta. *Tiede & Edistys* 28 (2003): 1, 1-11.
- Meklin, Pentti: Tavoitteiden saavuttamisen arviointi kuntataloudessa. Julkaisussa Myllymäki A, Vakkuri J (toim.) *Tulos, normi, tilivelvollisuus*, Tampere University Press, Tampere 2001, 91-117.
- Meklin, Pentti: *Valtiontalouden perusteet*, Edita, Hallinnon kehittämiskeskus, Helsinki 2002.
- Mälkiä, Esko & Alaranta, Hannu, Heinonen, Ari & Häkinen, Arja & Kauranen, Kari & Kujala, Urho & Kuukkanen, Tiina & Nevala, Nina & Satka, Jari & Ylinen, Jari: *Fyysiseen aktiivisuuteen perustuvan fysioterapian ja kuntoutuksen näytön tarkastelua*. *Kuntoutus* 27 (2004): 1, 4-21.
- Oakley, Ann: *Experiments in knowing. Gender and Method in the Social Sciences*, New Press, New York 2000.
- Oakley, Ann: *Social Science and Evidence-Based Everything: the case of education*. *Educational Review* 54 (2002): 3, 277-286.
- Pawson, Ray & Tilley, Nick: *Realistic evaluation*. London, Thousand Oaks, New Delhi 1997.
- Power, Michael: *The Audit Explosion*, London, Demos 1994.
- Power, Michael: *The Audit Society: Rituals of Verification*, Oxford University Press, Oxford 1997a.
- Power, Michael: *From Risk Society to Audit Society*. *Soziale systeme* 3 (1997b): 1, 3-21.
- Pöysti, Tuomas: *Tehokkuus, informaatio ja eurooppalainen oikeusalue*, Forum iuris. Helsingin yliopiston oikeustieteellisen tiedekunnan julkaisuja, Helsingin yliopisto, Helsinki 1999.
- Rajavaara, Marketta: *Yhteiskuntaan vaikuttava Kela,*

- Katsaus vaikuttavuuden käsitteisiin ja arviointiin, Sosiaali- ja terveysturvan katsauksia 69, Kelan tutkimusosasto, Helsinki 2006.
- Robson, Colin: Käytännön arvioinnin perusteet, Opas evaluaation tekijöille ja tilaajille, Tammi, Helsinki 2001.
- Rose, Nikolas: Powers of Freedom, Reframing Political Thought, Cambridge University Press, Cambridge 2000.
- Rostila, Ilmari: Realistinen arviointitutkimus toimintamallin kehittämistyössä. Julkaisussa Järvikoski Aila & Härkäpää Kristiina & Nouko-Juvonen Susanna (toim.) Monia teitä kuntoutuksen arviointiin, Tutkimuksia 69, Kuntoutussäätiö, Helsinki 2001, 35-53.
- Salminen, Ari: Julkisen toiminnan johtaminen, Edita, Helsinki 2004.
- Sanderson, Ian: Is it 'what works' that matters? Evaluation and evidence-based policy making. *Research Papers in Education* 18 (2003): 4, 331-345.
- Schwandt, Thomas A.: Evaluation as Practical Hermeneutics. *Evaluation* 3 (1997): 1, 69-83.
- Schriren, Michael: Evaluation: The State of the Art. Conference in Stockholm, Sweden, April 24-26, 1997, Evaluation as a Tool in the Development of Social Work Discourse. Centre for Evaluation of Social Services, The National Board of Health and Welfare, Stockholm 1997.
- Seppänen-Järvelä, Riitta: Luottamus prosessiin, Kehittämistyön luonne sosiaali- ja terveysalalla, Tutkimuksia 104, Stakes, Helsinki 1999.
- Sinfield, Ardian: Managing Social Security for What? And for Whom? EISS Conference in Göteborg, 6-10 september 2000, moniste 2000.
- Sinkkonen, Sirkka & Kinnunen, Juha: Arviointi ja seuranta julkisella sektorilla, Kuopion yliopiston julkaisuja, Yhteiskuntatieteet 22, Kuopion yliopisto, Kuopio 1994.
- Sosiaali- ja terveysministeriön hallinnonalan tuloksellisuusraportointi eduskunnalle, Tarkastuskertomus 78, Valtiontalouden tarkastusvirasto, Helsinki 2004.
- Sosiaaliturvan suunta 2005-2006 (2006) Sosiaali- ja terveysministeriön julkaisuja 1. Helsinki: Sosiaali- ja terveysministeriö.
- Star, Susan Leigh & Griesemer, James R.: Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology. *Social Studies of Science* 3 (1989):19, 1907-1939.
- Strathern, Marilyn: Introduction: new accountabilities. Julkaisussa Strathern, Marilyn (ed.) *Audit cultures, Anthropological studies in accountability, ethics and the academy*, London, New York: Routledge, London, New York 2000, 1-17.
- Sulkunen, Pekka: Vaikuttavuus tutkimuksen rivien välissä. Julkaisussa Räsänen, Pekka & Anttila, Anu-Hanna & Melin, Harri (toim.) *Tutkimus menetelmien pyörteissä*, PS-kustannus, Jyväskylä 2005, 39-52.
- Suositus kunnallisen palvelutoiminnan tuloksellisuuden arvioinnissa käytettävistä käsitteistä ja niiden sisäl- löstä, Yleiskirje A 23/1989, Kunnallinen työmarkkinalaitos, Helsinki 1989.
- Tala, Jyrki: Lakien vaikutukset, Lakiuudistusten tavoitteet ja niiden toteutuminen lainsäädäntöteoreettisessa ajattelussa, Oikeuspoliittisen tutkimuslaitoksen julkaisuja 177, Oikeuspoliittinen tutkimuslaitos, Helsinki 2001.
- Tala, Jyrki: Lakien laadinta ja vaikutukset, Edita, Helsinki 2005.
- Tulosohjauksen terävöittäminen, Työryhmämuistioita 9, Valtionvarainministeriö, Helsinki 2003.
- Vartiainen, Pirkko: Palveluorganisaatioiden tuloksellisuusarviointi, teorit ja käytäntö, Päivähoitoon, erityisesti päiväkoteihin kohdistuva monitahoarviointi, Acta Wasaensia No 40. Hallintotiede 2, Vaasan yliopisto, Vaasa 1994.
- Vartola, Juha: Näkökulmia byrokraatiaan, Tampereen yliopistopaino Oy, Juvenes Print, Tampere 2004.
- Vedung, Evert: Evaluation models and the welfare sector. Julkaisussa Julkunen, Ilse (ed.) *Perspectives, models and methods in evaluating the welfare sector - a Nordic approach*. FinSoc, Working Papers 4, Stakes, Helsinki 2004.
- Wilkinson, Adrian & Wilmott, Hugh: Introduction. Julkaisussa Wilkinson, Adrian & Wilmott, Hugh (eds.) *Making Quality Critical, New Perspectives on Organizational Change*, Routledge, London, New York 1995, 1-32.
- Vilkkumaa, Ilpo: Kuntoutuksen tuloksellisuus. Julkaisussa Eräsaari, Risto & Lindqvist, Tuija & Mäntysaari, Mikko & Rajavaara, Marketta (toim.) *Arviointi ja asiantuntijuus*, Gaudemus, Helsinki 1999, 71-89.
- Vuorela, Terho: Toiminnan arvioinnin tutkimusotteiden kehitys, Valtionhallinnon kehittämisskeskus, Vapokustannus, Helsinki 1990.
- Yhteiskunnallisen vaikuttavuuden arviointistrategian laatiminen, Valtion tulosohjaus- ja tilivelvollisuus uudistus, TulosTILVE. Vaikuttavuusryhmä 20.2.2004, Valtiovarainministeriö, Helsinki 2004.
- Yliopistolaki 27.6.1997 / 645.



Yleinen ja erityinen viitekehys arvioinnista

Petteri Paasio

ABSTRACT

A general and a specific framework of evaluation

A general and a specific framework of evaluation is formulated in order to comprehend the expanding evaluation activities. On the general level, evaluation is seen as a basic human activity which can be separated from knowledge and action. Evaluation can be regarded as transdisciplinary, where its vital role is to have the expertise in making value judgements.

In democratic societies evaluation has a specific meaning which implies answering the question: what must be evaluated. It is argued that the outcomes of social programs and institutions are the main objects of evaluation. In the specific framework of evaluation, evaluation is defined as the evaluation of effectiveness, because the value of social programs and institutions is explicitly their outcomes. An exact definition of effectiveness is presented. It makes possible to distinguish impact assessment from the evaluation of effectiveness, because causal analysis is only one part of effectiveness. In many cases the causal analysis has nothing to do with effectiveness.

The most promising applications of the specific theory of evaluation are presented. The first one is program theory by Chen, which provides conceptual tools to study the essential question: how to connect prescriptive professional action in organizations to the descriptive changes taking place in the world. The latter means outcomes. The second application is realistic evaluation, which contains a powerful conceptual framework that enables to understand causal relationships in a meaningful way.

JOHDANTO

Suomessa on 1990-luvun lopulta alkaen tapahtunut hyvinvointipalvelujen ja laajemmin koko julkisen sektorin kentässä poikkeuksellinen arvioinnin ekspansio. Tämä ilmiö näkyy ainakin neljällä alueella.

Ensiksi on helppo havaita, että arviointia koskevaa kirjallisuutta julkaistaan kiihtyvällä vauhdilla. Arvioinnin edistämiseksi on perustettu lukuisia hankkeita, jopa instituutioita. Arvioinnista myös keskustellaan vilkkaasti erilaisissa seminaareissa.

Toiseksi erilaisia arviointeja tehdään yhä enemmän ja arviointiraportteja julkaistaan niin runsaasti, että on mahdotonta pysyä niistä ajan tasalla edes rajoitetulla toimialalla. Näyttäisi jopa siltä, että hankkeen tai toiminnan arvon mitta on se, että siitä on tehty arviointiraportti.

Kolmanneksi yhteiskunnallisella kentällä toimivissa organisaatioissa, pienistä järjestöistä valtiollisiin instituutioihin, sijoitetaan mittavia resursseja toiminnan tuloksellisuuden arviointikäytäntöihin osana organisaatioiden johtamisjärjestelmiä. Organisaation arvon mitta näyttäisi olevan se, että sillä on strategia, jonka toteutusta arvioidaan.

Neljäs huomio liittyy vaikuttavuuden merkitykseen. Suomessa on tapahtunut hyvin nopea ajattelutavan, tai ainakin puhutavan, muutos sen suhteen, mitä tulisi ensisijaisesti arvioida. Niinkin äskettäin kuin vuonna 2002, Tampereen yliopiston rehtori pohti pitkään avaussanoissaan kansainvälisessä Evaluation for Practice -konferenssissa kysymystä, onko vaikuttavuuden arviointi mahdollista. Ympäri maailmaa Tampereelle matkustaneiden ja vaikuttavuusarviointien tuloksia esittelemään valmistautuneiden konferenssivieraiden helputukseksi hän päätyi pohdiskelussaan lopputulokseen: vaikuttavuus-

den arviointi voi olla mahdollista.

Tänä päivänä vaikuttavuus nähdään laajasti keskeisenä arviointikäsityksenä samaan aikaan kun vaikuttavuutta pidetään toiminnan ja organisaatioiden tärkeimpänä päämääränä ja tulosaluena.

Vajaassa kymmenessä vuodessa tapahtunut massiivinen arvioinnin ja muutaman viimevuoden kuluessa tapahtunut vaikuttavuudesta puhumisen ekspansio ovat olleet ilahduttavaa kehitystä. Tätä kehitystä tarkastellessa joutuu väistämättä kyseenalaistamaan joitakin vahvoja ennakkoletuksia. Esimerkiksi yleinen käsitys siitä, että julkinen sektori on hidasliikkeinen ja muutosta hylkivä, ei välttämättä pidäkään paikkansa. Suomessa ainakin olemme hyvin lyhyessä ajassa ottaneet kiinni sitä kuilua, joka vallitsi arviointikäytännöissä meidän ja kansainvälisen kärjen välillä.

Tapahtuneella nopealla muutoksella on myönteisiä vaikutuksia, mutta vielä enemmän positiivisia mahdollisuuksia. Asialla on kuitenkin kääntöpuolensa. Touhukas puuha arvioinnin ympärillä voi jäädä merkityksettömäksi kuohuksi, jos ei samanaikaisesti rakenneta kestäväää ymmärrystä arvioinnin ytimestä: mistä arvioinnissa on kysymys yleisellä tasolla ja missä ovat sen haasteet erityisellä tasolla.

Kyseessä ei luonnollisestikaan ole vain suomalainen ilmiö. Kirjansa ensimmäisillä sivuilla Pawson ja Tilley (1997) viittaavat kansainvälisessä katsannossa samantyyppiseen ilmiöön. Heidän mukaansa arviointia voi verrata nopeasti kasvaneeseen ja kehittyneeseen murrosikäiseen nuoreen. Toisaalla on lupaus ja mahdollisuus arvioinnin kaikkivoipaisuudesta. Toisaalla tulokset ja näytöt eivät ole kehuttavia. ”Maailma lepää hänen jalkojensa alla. ... Resurssit odottavat käyttöä. Odotukset ovat korkealla. Mutta hän ei ole kuitenkaan täysin varma siitä, kuinka pääsisi ylös sohvaltaan.” (Pawson & Tilley 1997, 4.)

Tällä hetkellä merkittävä este suomalaisen arviointikäytäntöjen kehittymiselle ihmisten hyvinvointia, organisaatioiden tuloksellisuutta ja hyvinvointiammattilaisten onnistumista edistäväksi tekijäksi on arvioinnin ja erityisesti vaikuttavuuden arvioinnin teorian ja käytännön perinteen puuttuminen tai sen ohuus. Vaikka nykyisin on suosittua jäljittää joitain vaikuttavuuden arvioinnin piirteitä hyvinkin kauas historiaan (esim. Raja-vaara 2006), osoittaa tämä heikon historian tajun lisäksi itse ongelmaa: ei oikein hahmoteta mistä

puhumme, kun puhumme vaikuttavuuden arvioinnista. Se, että arviointiymmärrys, -käytännöt ja -teoria ovat nykyisinkin vain vastikään kehittymään lähtenyt hauras taimi, tulee selkeästi esille siitä, että meillä ei ole keskeisille arviointikäsitteille suomenkielistä vastinetta lainkaan tai käsitteelle ei ole vakiintunutta vastinetta, vaan jokaisella kirjoittajalla on sille oma käänös.

Arvioinnin ja vaikuttavuuden arvioinnin ekspansion tässä vaiheessa olisi tarpeellista luoda teoreettinen ja käsitteellinen perusjäsenitys sille, mistä me puhumme silloin kun puhumme arvioinnista ja vaikuttavuudesta. Thomas Kuhnin (1996) paradigman käsite viittaa siihen toiminnan ja tutkimuksen välttämättömään jaettuun traditioon, joka mahdollistaa toiminta-alan tuloksellisuuden. Paradigman keskeinen merkitys on, että jokaisen alan edustajan ei tarvitse rakentaa alansa uudelleen peruseriaatteista ja perustella jokaisen käytetyn käsitteen oikeutusta, vaan hän voi keskittyä varsinaiseen luovaan alan ongelmanratkaisuun.

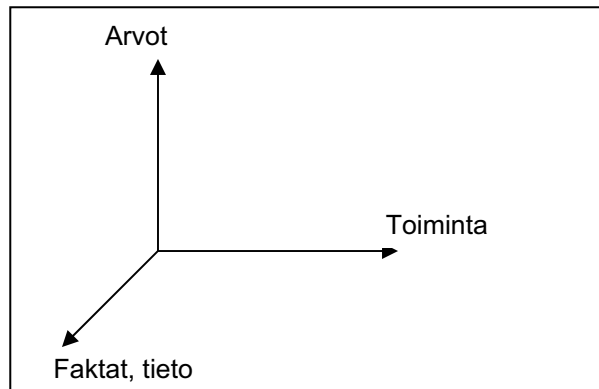
Tämän artikkelin tavoitteena on tarjota viitekehys arvioinnin yleiselle tunnistamiselle ja sen merkityksen ymmärtämiselle osana inhimillistä elämää sekä tarkastella vaikuttavuutta arvioinnin erityisenä viitekehyyksenä.

I OSA: ARVIOINNIN YLEINEN VIITEKEHYS

Arviointi on toimintaa, jossa yhdistetään kaksi eri elementtiä: faktat ja arvo. Asia voidaan kuvata yksinkertaisella kuviolla.

Mitä itse asiassa teemme, kun arvioimme? Arviointi on määritelmän mukaisesti aina ensisijaisesti arvon antamista tai määrittelyä. Vaikka tutkimuksessa käytettäisiin kuinka sofistikoituneita tieteellisiä analyysimenetelmiä, se ei ole arviointia, mikäli siinä ei tehdä arvoarvostelmaa. Arvoarvostelma tarkoittaa perusteltua kannonottoa asian hyvydestä tai huonoudesta.

Oheisen kuvion tarkoitus on havainnollistaa arvioinnin kolmiulotteista rakennetta perusinhimillisenä toimintana. Arvioinnin kehittymättömyys havainnollistuu kolmessa perusesimerkissä, jossa kolmiulotteinen tila on kutistunut kaksiulotteiseksi. Fanaatikko tai uskon soturi ei piittaa vähääkään tiedosta ja faktoista vaan nojaa toimintansa yksinomaan oman vakaumuksensa ja arvoarvostelmiensa oikeutukseen. Vastaavasti byrokraatti ei weberiläisen määritelmänsä mukai-



Kuvio 1. Arviointi yleisellä tasolla.

sesti puutu lainkaan arvokysymykseen vaan toimii mahdollisimman tehokkaasti tietämystään hyödyntäen muualla asetettujen arvopäämäärien puolesta, olipa kyseessä vaikkapa tietyn ihmisryhmän eliminointi. Kolmanneksi nykyaikainen filosofi on keskittynyt tietämyksensä ja käsitteellisten välineittensä avulla hienovaraiseen arvo-teorioiden analysointiin, ilman että tällä arvo- ja tietoekspertiisillä on mitään kytköstä käytännön toimintaan.

Ammatillisen arvioijan haaste liittyy siihen, että hän ei ole yksin tällä arvon määrittelyn kentällä. Inhimillinen elämä ei ole mahdollista ilman arviointia, koska arviointi on varhaisin ja perustavin inhimillinen käytäntö ja päättelyn muoto (Mathison 2005). Kun tarkastelemme ihmisten arkista toimintaa, on melko helppo havaita, että merkittävä osa kaikkien ihmisten vuorovaikutuksesta ja päättelystä liittyy erilaisten asioiden arvon määrittelyyn alkaen säätilasta päättyen organisaation ja yksilöiden onnistumiseen. Tällä perusinhimillisellä arkisella arvioinnilla on myös hyvin kohtalokkaat vaikutukset yksilöiden, ryhmien, organisaatioiden ja kansakuntien kannalta. Kaikkien kriisien ja konfliktien yhteinen rakennetekijä näyttäisi olevan epäonnistunut arviointi tai erimielisyys toiminnan tai asian arvosta. Tämän toteamiseksi ei tarvitse kuin tutkia minkä tahansa avioeron tai kansainvälisen konfliktin logiikkaa. Konkurssiin menneen yrityksen taustalta paljastunee virheellinen arvio yrityksen liiketoimintaedellytyksistä eli viimekädessä nykytoiminnan taloudellisesta arvosta. Vastaavasti kaikki inhimilliseen kasvuun, kehitykseen ja oppimiseen

perehtyneet tietävät kuinka ratkaisevassa asemassa merkittävälle saavutuksille ovat oikeaan osuneet arviot. Samoin minkä tahansa organisaation strategia on synonyymi arvioinnille siitä, mikä on arvokkainta organisaation menestyksen kannalta.

Ammatillisen arvioinnin tulisi olla erityisosaaja tällä arvioinnin ruuhkaisella kentällä.

ARVIOINTI TRANSDISCIPLIINÄ

Michel Scrivenin (1991) esittämä ajatus arvioinnista transdisipliininä viittaa juuri tähän arvioinnin ominaisuuteen. Transdisipliinillä tarkoitetaan sellaista itsenäistä tieteen tai tietämyksen alaa, joka tuottaa muiden tieteiden tai tietämyksen alojen käyttöön teoriaa, metodologiaa ja välineitä, joiden avulla ne voivat onnistua omassa tehtävässään paremmin. Kaikilla toiminnan ja tieteen aloilla tehdään loogisia päätelmiä ja laskutoimituksia, mutta transdisipliinit logiikka ja matematiikka ovat erikoistuneet näiden inhimillisen päättelyjen sääntöjen ja periaatteiden täsmälliseen analysointiin. Scrivenin mukaan arviointi on vastaavanlainen transdisipliini. Toisin sanoen arvioinnissa on kysymys teoriasta, metodologiasta ja välineistä, jotka ovat samanlaisia, riippumatta siitä arvioidaanko uimahyppyä, organisaation tuloksellisuutta tai väitöskirjan arvoa.

Arvioinnissa transdisipliininä korostuu kyky tehdä päteviä arvoarvostelmia. "Kaikkein yksinkertaisimmassakin tapauksissa on monia loogisia ja tieteellisiä haasteita arvojen selkeyttämiseksi

arvioinnissa. Näihin kuuluu: tunnistaa ja poistaa (i) epäjohdonmukaisuudet yksilöllisissä arvotuksissa (ii) arvojen väärinkäsitykset ja väärin-tulkinnat (iii) arvojen taustalla olevat epätodet tosiasioita koskevat olettamukset. (iv) halujen ja tarpeiden erottaminen toisistaan. Seuraavien ongelmien hallinta: (v) miten varmistamme, että olemme tunnustaneet kaikki olennaiset arvon tai ansion ulottuvuudet ja (vi) kuinka löydämme sopivat mittarit niille (vastakohtana sille, että vain nimeämme abstraktit ominaisuudet, joita tiedämme haluavamme). Näiden ulottuvuuksien painoarvon punnitseminen jollain tavalla, joka mahdollisimman tarkasti heijastaa tarkoituksi-amme. (viii) Kaikkien asettamiemme tai paikantamiemme standardien validiteetin varmistaminen.” (Scriven 1991, 5-6).

Voimme esittää yleismääritelmän arvioinnista:

Arviointi on perustavaa inhimillistä toimintaa, jossa kyetään luotettavalla tavalla yhdistämään arvot tietoon tosiasioista.

Miksi arvot muodostavat niin mittavan haasteen arvioinnille? Miksi ammatillisissa arvioinneissa on hyvin tavallista sivuttaa koko arvo-analyysi ja ottaa esim. toiminnalle asetetut tavoitteet itses-tään selvytenä, vaikka pikainen arvoanalyysi osoittaisi, että tavoitteet ovat niin hämärät ja ristiriitaiset, että niiden toteutumista ei voi arvi-oida? Tai miksi on niin kovin suosittua keskustella ja kirjoittaa arvioinnista sen sijaan, että tehtäisi-arviointia? Seuraavassa pyritään tarjoamaan muutama viite niistä suunnista, joihin saattaisi löytyä vastauksia tähän arvioinnin uskottavuuden kannalta suureen kysymykseen.

ARVOFOBIA

Scriven (1991, 375-378) on diagnostisoinut ilmiön, jolle hän on antanut nimen arvofobia. Scrivenin mukaan arvofobia on irrationaalinen arvioinnin pelko, joka usein ilmenee kohtuuttoman vahvana vastenmielisyytenä arviointia kohtaan tai sen vastustamisena. Sitä ei tule sekoittaa täysin normaaliin arvioinnin kohteena olemiseen liittyvään ahdistukseen tai huonojen arviointimenetelmien perusteltuun vastustamiseen.

Arvofobiaa pyritään yleensä selittelemään jol-lain retoriseen kärjistyksen perustuvalla myy-tillä arvioinnin sopimattomuudesta tilanteeseen. Niillä on hyvin vanha perinne: ”älä tuomitse,

ettei sinua tuomittaisi” on raamatullinen edeltäjä tieteen arvovapauden opille ja on yhtä epäjohdonmukainen sitä kannattavien henkilöiden toi-minnan suhteen (koska se sisältää kielteisen arvioinnin niistä, jotka arvioivat). Tarina Paratiisista on myös paljastava: juuri hyvän ja pahan tiedon puun hedelmät ovat kiellettyjä koska täl-lainen tieto on varattu jumalille; Aatamilta ja Eevalta taas oli kielletty siitä syöminen, vaikka he eivät olisi kyenneet erottamaan oikeaa ja väärää ilman sitä. Myöhemmin oikeus määritellä kaik-kein vakavimmat virheet siirtyi jumalilta papeille.

Arvofobia on yleensä kaikkein voimakkainta ryhmämuodossa, jossa näkemyksen yksilöllinen sopimattomuus voidaan kätkeä ryhmäajattelun taakse. Se voidaan naamioida uskomuksella, että on papin velvollisuus tunnistaa ja polttaa noidat ja kerettiläiset, jotka kyseenalaistavat ole-massa olevat arvon ja sopivaisuuden standardit. Laajalle levinnyt hyväksyntä arvovapaan tieteen opinkappaleesta on ryhmäajattelun ilmiö, joka hyödyntää arvofobiaa järjellisyuden kustannuk-sella.

Se, että arvioinnissa on kysymys syvästi inhi-millisestä toiminnasta, tulee esille seuraavasta Scrivenin määritelmän osasta.

”Etiologisesti puolustusmekanismi, jossa uhkaa vastaan hyökätään, on osa arvofobikon aggres-siivisen reaktion perustaa ja jo sinällään epä-ammattillista toimintaa. Mutta osa arvofobiasta juontuu syvemmälle haluttomuuteen kohdata epämiellyttäviä tosiasioita itsestä vaikka näin toi-mimisella olisi pitkänaikavälin hyötyjä. ... Arvofobia tai arviointiahdistus (lievempi muoto fobiasta) johtavat moniin häiriöihin arvioinnin suorittami-ssa” (Scriven 1991, 376)

Arvofobia ei ole vain amatööriarvioijien tai arviotavien ongelma. Arvofobia on hyvin yleistä myös ammatillisten arvioijien ja arvioinnin tilaa-jien joukossa. Arvofobian vaikutuksista amma-tilliseen arviointiin Scriven mainitsee seuraavia esimerkkejä:

- Pateettinen sen takaaminen, että arviointi-prosessi järjestetään ”vain henkilöstön autta-miseksi, ei kritisointiseksi”
- Palvelun vaikutusten arvioinnin korvaaminen toimeenpanon seurannalla.
- Patologinen pyrkimys välttää arviointi-käsit-teen käyttöä käyttämällä jotain oletettavasti vähemmän uhkaavaa vastinetta (auditointi, kartoitus, analyysi) sen sijaan, että otettaisiin

ammattillinen vastuu arvioinnin tekemisestä ja arvioitavana olemisesta.

- Arvioinnin tekijöiden liiallinen kanssakäyminen arviointikohteen henkilöstön kanssa (verukkeilla jotta ahdistus vähenisi, toteutus paranisi tai relevanssi lisääntyisi), joka odotetusti tuottaa älyllisesti vaatimattomia arvioiteja (syyllistämisen välttämisen kautta) ja suhteettoman määrän hyvin suopeita arvioiteja. Arvioijat, jotka käyttävät ”pehmeitä menetelmiä” kuten responsiivista arviointia, usein onnittelevat itseänsä pelkoreaktioiden välttämisestä. Mutta tämän ’menestyksen’ salaisuus on tietenkin pelon poistaminen poistamalla arvioinnin ydin, mikä merkitsee luotettavuuden, vastuullisuuden ja objektiivisuuden heikkenemistä.

OPPIMINEN

Scriven ei puhu paljoakaan siitä, kuinka arvo-fobian välttämiseksi tulisi toimia. Käytännön arviointikokemus antaa tähän kuitenkin jotain suuntaviivoja. Ensimmäinen asia on arvo-fobian tunnistaminen todellisena ja yleisenä asiana. Sen jäljille pääsee yksinkertaisimmin miten yksilöt, ryhmät tai organisaatiot käyttävät käsitettä ”hyvä”. Kirjoittaja on osallistunut lukemattomiin kehittämistilaisuuksiin, jotka alkavat sanoilla: ’Te teette jo nykyisin hyvää työtä, mutta...’ Kyseessä on tyypillinen arvo-fobian ilmaus: jos kerran tarkoitus on kehittää toimintaa, olisi syytä uskaltaa sanoa syy tälle. Toinen askel irti arvo-fobiasta on se, että aina kun jokin todetaan hyväksi, kyetään ilmaisemaan perustelut väittämälle. Eli toinen askel on yksinkertaisesti rohkeus kohdata tosiasiat silmästä silmään. Ammatillisen arvioinnin kannalta tämä on ratkaiseva kysymys, koska juuri sen avulla pitäisi pystyä vastaamaan perustellusti arvoarvostelmaan.

Vaikka oppivan organisaation teoriassa ja käytännössä ei yleensä puhuta paljoa arvioinnista tai arvoista, ovat alan klassikot päätyneet hyvin samankaltaisiin tuloksiin. Esimerkiksi Peter Senge (1994, 150-159) kuvaa kaiken kehityksen ja oppimisen keskeisenä voimana sananmukaisesti luovaa jännitettä, joka muodostuu rehellisen nykytilan tunnistamisen ja päämäärän tai vision (eli arvokkaan asiantilan) välille. Tämä luova jännite eli kehityksen moottori on mahdollista kadottaa kahdella tavalla. Joko ei uskalleta rehellisesti

tunnistaa nykytilannetta tai sitten ei kyetä näkemään mitään arvokasta tulevaisuuden asiantilaa. On helppo nähdä, että arvo-fobia tuottaa tehokkaasti nämä molemmat mahdollisuudet.

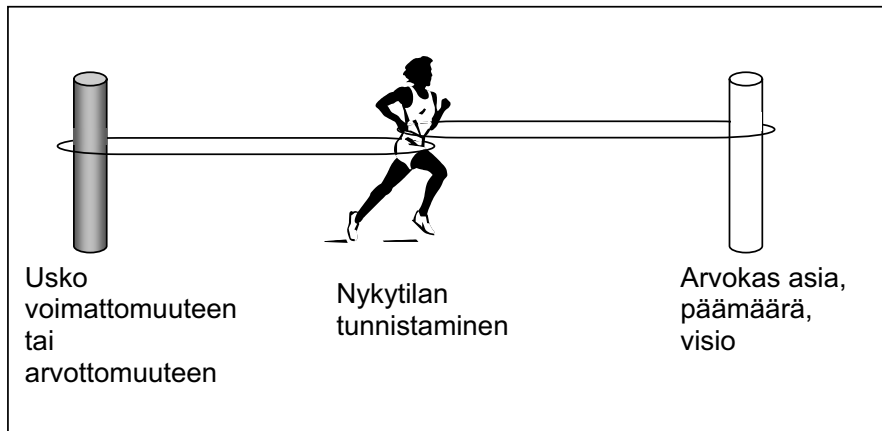
Arvo-fobian ymmärtämisen kannalta Sengen esityksen tekee erityisen mielenkiintoiseksi se, että vaikka nykytila ja arvot tunnistettaisiin, kuvioon astuu vastavoima, jonka voima lisääntyy samassa suhteessa kun muutos tai kehitys etenee kohti arvokkaita asioita tai päämääriä. Kuten oheisesta Sengen kuvioista käy ilmi, tämä vastavoima on usko omaan voimattomuuteen tai arvottomuuteen. Tämä ajatus antaa potentiaalisen selityksen arvo-fobian olemassaololle. Jos yksilönä uskon tai ryhmänä uskomme meidän perimmäiseen arvottomuuteen, on luonnollista, että kaikin voimin pyrimme estämään arvioinnin.

Myös Argyris ja Schön (1996, 99-100) ovat kiinnittäneet huomiota siihen, että organisaation oppimisen keskeisenä esteenä ovat defensiiviset käytännöt. Defensiivisten käytäntöjen tarkoitus on suojata yksilöitä kokemasta noloja tai uhkavia tilanteita samalla kun ne estävät yksilöitä tai organisaatioita tunnistamasta näiden uhkien syitä, jotta ongelmien korjaamiseksi voitaisiin ryhtyä toimiin. Yksinkertaisimmillaan defensiivissä käytännöissä on kysymys siitä, että ryhmätilanteissa ihmiset pyrkivät viimeiseen asti välttämään perustavimpien ajatustensa ilmaisemista, ja näin käy mahdottomaksi oppia uusia ajatuskulkuja.

Johtopäätöksenä voi todeta, että Scrivenin arvo-fobian käsite on sinänsä tärkeä. Arvo-fobiassa ei ehkä olekaan kysymys patologisesta tai irrationaalisesta käyttäytymisestä, vaan se on ymmärrettävä muutoksen ja oppimisen dynamiikkaan kuuluvana perusvoimana. Näin sen vaikutus on läsnä jokaisessa arvioinnissa ja arviointitilanteessa, mikäli defensiivisyyspakkoa ei ole osattu tai ymmärretty käsitellä ja vähentää.

ARVOT

Perustava selitys arvioinnin arvo-ulottuvuuden aikaansaamille ongelmille saattaisi löytyä siitä, että meille näyttäisi tuottavan suuria vaikeuksia ymmärtää, mitä arvot tosiasiallisesti ovat, lukuun ottamatta yhtä arvoa: taloudellista arvoa. Maapallon suurin arviointimekanismi on epäilemättä markkinat, joilla globaalisti saavat mitkä tahansa asiat työvoimasta, tavaroihin ja palveluihin täs-



Kuvio 2 Kaksi muutoksen voimakenttää (Senge 1994, 157)

mällisen taloudellisen arvonsa. Mikä on ymmärryksemme muista arvoista?

Radikaaleimman selityksen arvojen tuottamaan ongelmaan on tarjonnut Alasdair MacIntyre (2004). Hänen mukaansa olemme yksinkertaisesti menettäneet kyvyn ymmärtää mitä arvot ovat. Viimeiset sata vuotta on länsimaisessa ajattelussa ollut lähes täydellisessä monopoliasemassa emotivismiksi kutsuttu käsitys arvoista. ”Emotivismin mukaan kaikki arvottavat arvostelmat ja erityisesti moraaliset arvostelmat ovat ainoastaan ilmaisia preferensseistä, asenteista ja tunteista.” (MacIntyre 2004, 29.)

Emotivismin voittokulku on ollut kiistatonta, koska näin on saatu ihmisten markkinakäyttäytymistä selittävä viitekehys. Kokonaisia tieteenaloja on rakentunut tämän selittämisen ympärille. Esimerkiksi taloustieteessä tai sosiologiassa ei arvoja voi muulla tavalla ymmärtää. Emotivismi perustuu kuitenkin karkeaan väärinkäsitykseen. Arvot näkyvät yksilöiden yksittäisinä preferensseinä ja mielipiteinä tai tuntemuksina, mutta ne ovat eri asioita kuin nämä yksittäiset tuntemukset. Sen sijaan voidaan esittää, että:

Arvot ovat yksilöiden ja ryhmien elämän mielekkyyttä ja tarkoitusta jäsentäviä perustavia merkitysrakenteita, jotka vastaavat viimekädessä kysymyksiin, mikä elämässä on merkityksellistä ja mitä on hyvä elämä.

Viitteen arvojen merkityksestä ihmisten ja yhteisöjen elämässä saamme kaukaa historian hämästä, tarkemmin sanottuna 2300 vuoden takaa.

Faidroksessa (246a-) Platon esittää sielun kolmijaon vertauskuvana siivellisen parivaljakon. Ajuri edustaa järkeä ja tietoa, mutta ihmisen liikkeen kannalta sillä ei ole mitään voimaa. Voima ja liike tulee kahdesta hevosesta. Toinen hevonen edustaa arkisia haluja, tarpeita ja toimeliaisuutta niiden suhteen. Nyt kysymys kuuluu, mikä on tämä toinen hevonen? Siivellisen parivaljakon vaellus taivaalla kohti tuttuutta ja kauneutta on kiinni juuri tästä hevosesta. Platon kutsuu tätä sielun kolmatta osaa nimellä thymos. Se kääntetään usein kiihkeydeksi, jolloin menetetään mahdollisuus ymmärtää mitä käsitteellä tarkoitetaan. On itse asiassa täysin MacIntyren teorian mukaista että käsitettä ei löydy: jos emotivismi on läpäissyt ajattelumme ja kieleemme, ei sen kanssa ristiriidassa olevalle ilmiölle voi löytyä kuin täysin emotivistisia käsitteitä.

Francis Fukuyama (1992) on analysoinut laajasti thymos-käsitettä. Hän tiivistää thymoksen merkityksen seuraavalla tavalla. Käsitteellä viitataan ”siihen ihmisen psyyken osaan, joka tuntee tarvetta arvojen määrittelyyn - ensisijaisesti arvostamaan ihmistä itseään, mutta myös määrittelemään arvoja häntä ympäröiville ihmisille, teoille ja esineille. Tästä persoonallisuuden osasta ovat lähtöisin ylpeyden, kiukun ja häpeän tunteet, eikä sitä voida sulattaa yhtäältä haluihin eikä toisaalta järkeen.” (Fukuyama 1992, 211.)

Johtopäätöksenä voidaan todeta, että arvot näyttävät olevan inhimillisen elämän keskeinen merkitysrakenne ja arvojen määrittely tärkeä osa inhimillistä toimintaa. Arvot selkeästi eroavat toi-

saalta yksittäisistä preferensseistä tai valinnoista ja toisaalta erilaisista tunteista vaikka ne usein ilmenevätkin vahvoina mielipiteinä tai voimakaina tunteina. Voi vain ajatella kuinka toisen näköiseltä ihmiskunnan historia näyttäisi, jos inhimillistä toimintaa eivät olisi ohjanneet sellaiset keskeiset arvot kuin totuus, vapaus, kauneus tai oikeudenmukaisuus. Oma empiirinen kysymyksensä on sitten se, että minkälainen on meidän arvokenteemme. Empiirisesti testattuja vastauksia tähän kysymykseen löytyy mm. edellä viitatus Chris Argyrisin tuotannosta.

ARVIOINTI JA ETIIKKA

On suorastaan hämmästyttävää kuinka vähän kansainvälisessä arviointikirjallisuudessa on keskusteltu arvioinnin ja etiikan välisistä suhteista muuten kuin arvioinnin etiikkaa koskevassa merkityksessä. Arviointi ja etiikka ovat kuitenkin rakenneyhtäläisiä: molemmissa on kysymys arvoarvostelmien tekemisestä tai arvon määrittelystä.

Historiallisessa katsannossa arvioinnin perusteokset arvo-ulottuvuuden näkökulmasta näyttäsivät olevan Aristoteleen Nikomakhoksen etiikka, joka käsittelee ihmisen hyvää elämää ihmisenä sekä Poliitiikka, joka käsittelee hyvän elämän yhteisöllisiä ehtoja, hyvää elämää poliiksessa eli kaupunkivaltiossa.

Aristoteles toteaa arvioinnin perustavan periaatteen nasevasti: ”Koska nyt käsillä oleva tutkimus [Nikomakhoksen etikka] ei tapahdu teoreettista tietoa silmällä pitäen, kuten jotkut muut tutkimukset (sillä emme nyt tutki hyvettä tietääksemme pelkästään mitä hyve on, vaan tullaiksemme hyväksi - muutenhan tästä ei olisi mitään hyötyä)” (1103b26-29).

Päätän artikkelin ensimmäisen osan kokoaavaan johtopäätökseen Aristoteleen periaatteen mukaisesti:

hyvällä elämällä yksilöllisellä ja yhteisöllisellä tasolla on vahva kytkös yksilöiden ja yhteisöjen arviointikäytäntöihin, arviointikykyyn ja käsityksiin arvioinnista.

Arvioinnin etiikan sekä arvioinnin ja hyvän elämän välisen suhteen lisäksi arvioinnilla ja etiikalla on kuitenkin myös kolmas kytkös. Siinä kysytään, mitä on arvioitava eli minkä asioiden arviointi on eettinen velvollisuus. Jos tällainen

ala löytyy, kuinka tällä kentällä on arviointi toteutettava, jotta voisimme tulla hyväksi tämän alan edustajiksi. Artikkelin toinen osa pyrkii vastaamaan tähän kysymykseen osoittamalla, että tällainen ala löytyy. On olemassa viitteitä siitä, millaisilla arviointikäytännöillä voi tulla tämän alan hyväksi ammattilaiseksi.

II OSA: VAIKUTTAVUUS - ARVIOINNIN ERITYISENÄ VIITEKEHYKSENÄ

Kun arvioinnilla on merkitys hyvän elämän kannalta yleisellä tasolla, on olemassa eettinen velvollisuus arvioida sellaista toimintaa, jolla on erityinen vastuu inhimillisestä hyvinvoinnista. Demokraattisissa yhteiskunnissa tämä velvollisuus on kaikella verovaroin rahoitettavalla toiminnalla eli sillä, mitä on totuttu kutsumaan julkiseksi sektoriksi.

Julkisen sektori on olemassa edistääkseen ihmisten hyvinvointia. Joissain tapauksissa tämä kytkentä ihmisten hyvinvointiin on välillistä.

Julkisen sektorin tehtävä ihmisten ja hyvinvoinnin edistäjänä on ilmeisen haasteellinen ja monimutkainen. Jotta tehtävässä voisi onnistua ja siinä kehittyä, on yhtä ilmeistä, että tarvitaan parhaita mahdollisia arviointikäytäntöjä ja menetelmiä.

Näin päädyimme lähtökohtaväittämään.

Julkisella sektorilla ammatillisesti korkeatasoisen arvioinnin tulee olla keskeinen ja systemaattinen osa toimintaa. Arvioinnin tärkein kysymys on, kuinka toiminnassa on onnistuttu ihmisten hyvinvoinnin edistämiseksi.

Samansuuntainen lähestymistapa näkyy myös käsityksissä, jolla arviointi ymmärretään. Mark, Henry & Julnes (2000, 17) määrittelevät arvioinnin hyvin yksinkertaisella tavalla. Heidän mukaansa arvioinnissa on kysymys ”assisted sensemaking that can support the ultimate goal of social betterment”. Arvioinnissa on heidän mukaansa kysymys menetelmällisesti tuetusta ihmisten luontaisesta ymmärryskyvystä sekä merkitysten selvittämisestä, jonka tarkoituksena on yhteiskunnallisen hyvinvoinnin lisääminen.

JULKISEN SEKTORIN ARVO

Eräs tapa ymmärtää, miksi nimenomaan julkisella sektorilla on niin vaikea omaksua arviointia keskeiseksi osaksi käytäntöjä, on tarkastella julkisen sektorin toiminnan erilaisia arvoparadigmoja. Se tarkoittaa erilaisia tapoja määritellä julkisen sektorin toiminnan arvo sekä arvon edellyttämiä erilaisia arviointikäytäntöjä.

Julkisen sektorin historiasta ja nykyisyydestä voidaan tunnistaa ainakin kolme erilaista käsitystä, mihin julkisen sektorin arvo perustuu (Paasio 2003, 65-66). Nämä ovat:

- Intention eli pyrkimyksen hyvyys
- Toiminnan lainmukaisuus
- Palvelutuotannon tehokkuus

Intention eli pyrkimyksen hyvyyden historiallisena lähtökohdana on hyväntekeväisyys. Toiminnan arvo määräytyy toiminnan päämäärän kautta. Tällöin arvioinnin on kohdistuttava ensisijaisesti sen varmistamiseen, että kaikki toimijat ovat omaksuneet tämän saman hyvän päämäärän. Tämä lähestymistapa toiminnan arvoon on hyvin laajassa käytössä nykyisinkin. Sen eräs modifikaatio on panostaminen henkilöstön koulutukseen. Ajatellaan, että tietyn koulutusstandardin omaava henkilöstö hoitaa ammatillisen osaamisensa avulla tehtävät parhaalla mahdollisella tavalla. Toinen suosittu muoto on asettaa organisaatiolle yleviä tavoitteita. Silloin ajatellaan, että henkilöstö niistä inspiroituneena toteuttaa ne. Arviointi molemmissa tapauksissa on sitten varsin yksinkertaista: joko tarkistetaan henkilöstön koulutustaso tai kerätään todisteita siitä, kuinka organisaatio on yleviä tavoitteitaan toteuttanut. Intention hyvyyden arviointi ei kuitenkaan vastaa toiminnan varsinaista arviointia, koska hyvä intentio voi edesauttaa onnistumista, mutta se ei missään olosuhteissa sitä takaa. Yleensä suurimmat katastrofit ja kärsimykset syntyvät toiminnassa, joka on tehty ns. hyvässä uskossa eli hyvää tarkoittaen.

Lainmukaisuus toiminnan arvona on eräänlainen sovellus hyvästä intentiosta. Siinä lainsäätäjä määrittelee jotain hyvää tarkoitusta varten tietyn proseduurin, jonka toteuttamista valvotaan. Arvioinnin kannalta tämä on huomattavasti kehittyneempi versio, koska lainsäädäntöön liittyy automaattinen palautejärjestelmä. Kuka tahansa, jonka mielestä määritelty proseduri ei

ole toteutunut, voi saattaa asian arvioitavaksi. Tässä järjestelmässä arvioinnin siis hoitavat tuomioistuimet ja muut vastaavat muutoksenhaku- ja valitustenkäsittelyinstanssit. Toiminnan arvona lainmukaisuus on välttämätön mutta ei riittävä ehto. Asioita voi hoitaa laillisesti ja samalla erittäin huonosti.

Historiallisesti tuorein julkisen sektorin arvoparadigma on käsitys, että julkisen sektorin arvo määräytyy palvelutuotannon tehokkuuden kautta. Tämä on kahta edellistä kehittyneempi arvioinnin viitekehys, koska toiminnan arvo liittyy siihen, kuinka resursseja käytetään. Julkisella sektorilla tämä on olennainen kysymys kahdesta syystä: siinä toimitaan veronmaksajien rahoilla ja resurssit eivät voi kasvaa rajattomasti. Tämä viitekehys edellyttää myös, että kustannukset osataan kohdistaa suhteellisen tarkasti erilaiseen tekemiseen tai erilaisiin asiakkuuksiin. Tämän arvoparadigman ongelma on, että mikään ei takaa, että palvelujen käytöllä olisi jotain vaikutusta asiakkaan hyvinvointiin. Osa palveluista voi olla asiakkaan hyvinvoinnin kannalta hyödyttömiä, jolloin niiden tehokaskin tuottaminen on resursien haaskausta. Osa palveluista voi olla asiakkaan hyvinvoinnin kannalta jopa haitallista, jolloin niiden tehokas tuottaminen eettisesti tuomittavaa.

Yhteenvetona voi todeta, että hyvä intentio, lainmukaisuus ja palvelujen saatavuus sekä tehokas toteuttaminen ovat sinänsä arvokkaita asioita, ja niiden toteutumista voi mielekkäästi arvioida. Ne eivät kuitenkaan kuvaa julkisen sektorin toiminnan varsinaista arvoa. Julkisen sektorin arvo on se, missä määrin toiminnalla kyetään edistämään ihmisten hyvinvointia.

Julkisen sektorin varsinainen arvo on, missä määrin toiminnassa kyetään edistämään ihmisten hyvinvointia. Toisin sanoen julkisen sektorin toiminnan tärkein arvo on ihmisten hyvinvointiin kohdistuva vaikuttavuus.

ESIMERKKI VAIKUTTAVUUDEN ARVIOINNISTA

Malli esimerkki siitä, mistä vaikuttavuuden arvioinnissa on kysymys, on professori Harri Sintosen kehittämä 15D terveyteen liittyvän elämänlaadun mittarin ja sen käytön lääketieteellisten interventtioiden vaikuttavuuden ja kustannusvaikuttavuuden arvioinnissa (Sintonen 2001; Räsä-

nen, Sintonen ym. 2005; Kukkonen 2005). Mittari muodostuu 15 terveydellisen elämänlaadun mitattavasta ulottuvuudesta, joiden muodostama kokonaisuus voidaan tiivistää yhdeksi indeksiluvuksi. Mittarilla voidaan arvioida minkä tahansa lääketieteellisen intervention vaikuttavuutta ja kustannusvaikuttavuutta. Siis sitä kysymystä, miten eri interventioissa ihmisten yleinen terveydellinen hyvinvointi on kohentunut. Samalla voidaan vertailla sitä, miten suuriin terveydellisen hyvinvoinnin puutteisiin eri interventioita kohdistetaan.

Vaikuttavuuden arvioinnissa ei siis ole kysymys vain siitä, että katsotaan tuliko tavoite saavutettua. Jos polvileikkauksessa polvi ei tule kuntoon, ei toiminnalla ole luonnollisestikaan ole vaikuttavuutta. Vaikuttavuuden arvioinnissa on kuitenkin olennaista, miten ihmisten hyvinvointi muuttui. Eikä ensisijaisesti se, toteutuiko, mitä oli aikomus tehdä.

Käsillä oleva esimerkki on oivallinen myös sen takia, että se voi kitkeä syvään pinttynyttä väärinkäsitystä siitä, että vaikuttavuuden arvioinnissa olisi kysymys ensisijaisesti, ja terveydenhuollossa erityisesti, satunnaistetun koeasetelman käytöstä. 15D mittarin käytön esimerkit vaikuttavuuden arvioinnissa eivät ole perustuneet satunnaistettuun koeasetelmaan, vaan terveydenhuollon normaali toiminnan vaikuttavuuden arviointiin. Mittaria voisi luonnollisesti käyttää satunnaistetuissa koeasetelmissa, jos oltaisiin kiinnostuttu eri instituutioiden hoidon vaikuttavuuden eroista tai eri menetelmien vaikuttavuuden eroista. Kuitenkin vaikuttavuusnäkökulmasta on paljon tärkeämpää tietää olemassa olevan toiminnan ja niiden käyttämien mittavien resursien tosiasiallinen panos ihmisten hyvinvointiin. Täsmällisen syy - seuraussuhteen mittaaminen on yksi vaikuttavuuteen liittyvistä arviointikysymyksistä, mutta ei ainoa.

15D mittarin ehkä tärkein opetus vaikuttavuuden arvioinnille on, että vaikuttavuuden arviointi ei ole ensisijaisesti teknistä arviointispecialistien toimialaa. 15D kehittäminen vaikuttavuusmittariksi on edellyttänyt poikkeuksellista asiantuntemusta ihmisen terveydellisen hyvinvoinnin kokonaisuudesta. Vaikuttavuuden arvioinnin suurin haaste ei ole itse mittaaminen, vaan kokonaisvaltainen, jäsentynyt ja teoreettisesti kestävä ymmärrys siitä, mitä ollaan mittaamassa. Juuri tästä syystä ovat tärkeimmät vaikuttavuuden arvioinnin menetelmät monella alla laadullisia.

YLEISKATSAUS VAIKUTTAVUUDEN ARVIOINTIIN

Rossi, Lipsey & Freeman (2004) määrittelevät arvioinnin olevan nimenomaan vaikuttavuuden arviointia. Tästä huolimatta vaikuttavuuden arvioinnissa on useita olennaisia arviointikysymyksiä, jotka eivät suoranaisesti koske vaikuttavuutta. Mistä tämä paradoksi johtuu? Yritän seuraavassa tiivistää heidän monisatasivuisen kirjansa perusajatuksen.

Vaikuttavuuden arviointi nojautuu olennaisesti vaikuttavuuden aikaansaamiseen. Vaikka arviointi on aina ensisijaisesti vaikuttavuuden arviointia, joudutaan arvioinnissa esittämään kysymyksiä, jotka edesauttavat vaikuttavuuden aikaansaamista.

Varsinaisia vaikuttavuuden aikaansaamisen vaiheita ja niitä vastaavia arviointikysymyksiä ei ole kuin viisi. Nämä vaiheet ja kysymykset eivät koske vain uuden vaikuttavan toiminnan toteuttamista vaan, jokaisen olemassa olevan toiminnan tulisi kyetä niihin vastaamaan, voidakseen olla vaikuttavaa. Samalla kunkin arviointikysymyksen esittäminen ja siihen vastaaminen lisää olemassa olevan toiminnan vaikuttavuuden edellytyksiä.

Ensin pitää tietää kohdeväestön hyvinvoinnin ongelmat tai tarpeet. Tähän kysymykseen vastataan tarvearviointilla. Tarvearviointi on kuitenkin eri asia kuin palvelutarpeen arviointi. Tarvearvioinnin ensisijaisena kohteena on hyvinvoinnin toteutumisen tosiasiallinen dynamiikka ja ongelmat siinä. Toiseksi pitää pystyä eksplikoimaan se logiikka, miten olemassa olevalla tai potentiaalisella toiminnalla kyetään ongelmaan tai tarpeeseen vastaamaan ja minkälaisen lisääntyvän hyvinvoinnin toiminta logiikkansa perusteella aikaansaa tai mahdollistaa. Kyseessä on ns. vaikuttavuusteorian rakentamisen vaihe, jota artikkelissa käsitellään tarkemmin jäljempänä. Toiminnan logiikan tai vaikuttavuusteorian arviointi on yksi olennaisimmista vaikuttavuuden arvioinnin muodoista siitä yksinkertaisesta syystä, että toiminnan edellytykset olla vaikuttavaa paljastuvat siinä hyvin helposti ja pienin panoksien.

Kolmanneksi pitää kyetä toteuttamaan se, mitä on suunniteltu tehtävän. Tämä prosessien arviointi on yksi suosituimmista arvioinnin osa-alueista. Se on vaikuttavuuden arvioinnin kannalta olennaista siksi, että jos prosesseissa ei ole mitään yhtenäisyyttä, ei voida puhua minkään

asian vaikuttavuudesta eikä kyetä tunnistamaan vaikuttavia prosesseja. Prosessin arviointi on keskeinen alue myös siksi, että prosesseissa resurssit kohdentuvat tekemiseen ja tekeminen tuottaa jotain kokonaisuuksia asiakkaille. Perinteinen prosessianalyysin kaava input - prosessi - output on keskeinen osa vaikuttavuuden arviointia silloin kun ollaan kiinnostettu joko tuottavuudesta tai prosessien laadusta. Prosessien arvioinnin suuri suosio kertoo mahdollisesti vakavastakin ongelmasta. Mikäli prosessien arvioinnissa on tarkoitus vain selvittää, mitä tuli tehtyä, herää kysymys: eikö tähän kysymykseen vastaamiseksi jokaisessa organisaatiossa, toiminnassa ja hankkeessa ole aivan oma ammattiryhmä, jota kutsutaan esimiehiksi ja johtajiksi? Näin ollen, jos tämänkaltainen prosessien arviointi tehdään erillisenä ulkoisen arviointina, lienee jo arvioinnin tilaaminen selkeä osoitus siitä, että prosesseissa on vakavia ongelmia - erityisesti johtamisprosesseissa. Aito tilaus prosessien arvioinnille rakentuu kuitenkin vaikuttavuusnäkökulman kautta.

Neljänneksi pitää kyetä tunnistamaan, miten toimintaan osallistuneiden asiakkaiden hyvinvointi on tosiasiallisesti muuttunut. Neljäs arviointikysymys on vastauksen etsimistä toiminnan toteuttamisen vaatimukseen: miten toteutuneisiin toimintaprosesseihin osallistuneiden asiakkaiden hyvinvointi on muuttunut. Tälle arvioinnin osaluokalle on olemassa ytimekäs käsite: outcome assessment tai outcome evaluation. Tähän päivään mennessä suomenkielestä ei ole löytynyt vastinetta outcome käsitteelle. Tämä arvioinnin ja vaikuttavuuden arvioinnin ydinkohta tarvitsisi suomenkielisen selkeän vastineen. Pelkästään sen luominen olisi merkittävä panos suomalaisen arvioinnin ja vaikuttavuuden arvioinnin kehittämisessä. Joka tapauksessa tämän "asiakkaiden hyvinvoinnin muutoksen" arvioinnin sisällä on erilaisia arviointikysymyksiä. Tärkein kysymys on tunnistaa, mikä on ollut tämä muutos. Toinen kysymys on, missä määrin tämä muutos on johdettu nimenomaan interventiosta, johon asiakas on osallistunut. Tätä jälkimmäistä kutsutaan vaikutus arvioinniksi eli impact assessment.

Viidenneksi ja viimeiseksi tulee kyetä tunnistamaan ja analysoimaan niitä kustannuksia, mitä eri "asiakkaiden hyvinvoinnin muutokset" ovat kuluttaneet. Tämä on kustannus-vaikuttavuuden arviointia, mikä on keskeinen osa vaikuttavuuden arviointia. Missään luonnollisessa järjestelmässä ei voi sanoa vaikuttavaksi interventiota vaikka

siihen osallistuneiden asiakkaiden hyvinvoinnin muutokset olisivat merkittäviä, jos tämän intervention toteuttaminen kuluttaa kaikkien muiden interventioiden resurssit. Lopputuloksena olisi yleinen hyvinvoinnin heikentyminen.

Arvioinnin ja vaikuttavuuden arvioinnin logiikka kulkee myös vastakkaiseen suuntaan. Mikäli todetaan, että interventio ei ole ollut vaikuttavaa, on vain kolme vaihtoehtoa. Ensimmäinen vaihtoehto on, että ei ole tehty sitä mitä suunniteltiin. Jos prosessien toteutus on onnistunut, jää jäljelle kaksi vaihtoehtoa. Toinen vaihtoehto on, että toiminnan logiikka perustui väärin oletuksiin. Ei-vaikuttavuus on vahva näyttö edellisestä. Kolmas vaihtoehto on, että tarvearviointi oli epäonnistunut. Kenties ihmisillä ei yksinkertaisesti ollut sellaisia tarpeita tai ongelmia, joita heillä oli ajateltu olevan.

Tämä yksinkertainen arviointilogiikka tekee ymmärrettäväksi myös edellisen luvun kolme julkisen sektorin arvoparadigmaa. Intention hyvyys perustuu joko tietyn tarpeen tunnistamiseen tai jonkin toiminnan päämäärän ympärille. Toiminnan lainmukaisuus on taas tyypillinen prosessin laatumittari: onko toimittu tietyn prosessivaiheituksen mukaisesti. Palvelutuotannon tehokkuus on toinen prosessimittari, jossa tarkastellaan panosten suhdetta tuotoksiin.

Tässä esitetyn vaikuttavuuden arvioinnin yleisjäsenyyksen keskeinen idea on, että eri arviointikysymykset saavat merkityksen vaikuttavuuskysymyksen kautta. Esimerkiksi Suomessa tällä hetkellä suosittu prosessin kehittäminen ja arviointi ovat hyödyttömiä puuhia, mikäli ei kyetä tunnistamaan tosiasiallisia asiakkaita, heidän tarpeitaan ja sitä asiakas hyvinvoinnin muutosta, johon prosessilla pyritään.

OUTCOME

Edellä viitattiin outcome käsitteen keskeiseen asemaan vaikuttavuuden kannalta ja siihen tosiasiaan, että suomalainen vaikuttavuuden arviointi ja toiminnan vaikuttavuuden edistäminen edellyttäisi yksiselitteistä käsitettä tälle "asiakkaiden hyvinvoinnin muutokselle". Tämä käsiteongelma on sekä kielellinen, suomalaisen toiminnan kontekstiin liittyvä, että osaksi teoreettinen.

Teoreettinen osa asiaa on se, että käsite perustuu systeemi ajatteluun. On pystyttävä erottamaan toisistaan systeemi ja siitä erillinen ympä-

ristö. Tässä tapauksessa prosessi tai toiminta on systeemi, jossa käytetään erilaisia resursseja, joista syntyy erilaisia tuotteita. Sekä tarve tai ongelma että outcome ovat tästä systeemistä erillisiä asioita. Tarvearvioinnissa ei olla kiinnostuttu palvelujärjestelmän tarpeista (lisää sitä ja tätä) vaan ihmisten tarpeista. Outcome on näiden samojen ihmisten tosiasiallisessa hyvinvoinnissa tapahtunut muutos. Esimerkiksi asiakastytyväisyys ei ole outcome koska asiakastytyväisyys on prosessin laadun mittari. Jos tämä ihmisen asiakkuuselämässä koettu tyytyväisyys johtaa johonkin muutokseen hänen varsinaisessa elämässä, on tämä muutos outcome.

Outcome ei ole tulos tai lopputulos. Suomen kielen perussanakirjan (2004, 351-352) mukaan tuloksella tarkoitetaan seikkaa tai asiantilaa, joka jollakin toiminnalla saadaan aikaan. Se on siis aikaansaannos, saavutus tai seuraus. Suomalaisen julkisen sektorin todellisuudessa nämä tarkoitetut aikaansaannokset liittyvät lähes poikkeuksetta kaikkeen muuhun kuin outcome alueelle. Suomalaisessa todellisuudessa tulos on siis harvoin outcome. Poikkeuksen tekevät ne harvat organisaatiot, joissa tulostavoitteet liittyvät eksplisiittisesti vaikuttavuuteen.

Outcome käsitteen suomenkieliseksi vastineeksi voisi asiakasseuraus olla ajattelemisen arvoinen vaihtoehto. Vielä täsmällisempi ilmaus olisi asiakassysteemin muutos.

VAIKUTTAVUUS JA VAIKUTUS

Vaikuttavuuden arvioinnin esteenä näyttää olevan vaikuttavuuden ja vaikutuksen käsitteiden sekoittaminen toisiinsa. Missä tahansa kausaaliketjussa eli syy - seuraus -ketjussa, on syy elementin suhde seuraukseen yhtä kuin vaikutus. Luonnontieteiden keskeinen tutkimuskenttä on kausaalisuhteiden selvittäminen ilman, että tällä on mitään tekemistä vaikuttavuuden kanssa. Emme siis puhu tsunamin vaikuttavuudesta vaan sen vaikutuksista. Vastaavasti inhimillisen toiminnan kentällä käsitellään luontevasti kausaalisuhteita esimerkiksi rikosoikeudessa ilman, että tällä on mitään tekemistä vaikuttavuuden kanssa. Ihmisen arkisen elämän eläminen ei ole myöskään mahdollista ilman luotettavia kausaalisuhdeoletuksia.

Vaikuttavuudella tarkoitetaan sellaista ihmisten hyvinvointiin tähtäävän toiminnan ominaisuutta

tai suorituskykyä, jossa keskeinen elementti on asiakassysteemin hyvinvoinnin muutoksen (outcome) systemaattinen havainnointi ja tunnistaminen sekä tämän havainnoinnin perusteella todennettu muutos. Vasta tämän jälkeen eli toissijaisena kysymyksenä vaikuttavuus viittaa siihen, että havaittu muutos (outcome) on nimenomaan toiminnan tai intervention vaikutusta (impact).

Vaikuttavuuden vaikutuselementin selvittämiseksi on olemassa loputon määrä erilaisia tutkimus- ja vertailuasetelmia, joiden valinta on riippuvainen toisaalta vaadittavan evidenssin vahvuudesta, toisaalta muutosilmion luonteesta.

VAIKUTUSTEN ENNAKKOARVIOINNIT

Suomessa on erityisesti poliittisten päättäjien suosiossa ollut viimeaikoina erilaisten asioiden vaikutusten ennakkoarvioinnit. Hallinnollisina käytäntöinä meillä ovat vakiintuneet ympäristövaikutusten arviointi ja sosiaalisten vaikutusten arviointi. Uusimpia tulokkaita on mm. talousarvion sukupuolivaikutusten arviointi. Näillä vaikutusten arvioinneilla ei pääsääntöisesti ole mitään tekemistä vaikuttavuuden arvioinnin kanssa.

Vaikutusten arviointien ajatus on, että kun taloudellisen arvon ohjaamalla toiminnalla, esimerkiksi rakennushankkeella, on vaikutuksia toisiin yhteiskunnan arvokkaina pidettyihin asioihin tai toimintaan, on järkevää selvittää ja arvioida, mitkä nämä vaikutukset ovat. Miksi tämänkaltaisten vaikutusten selvittäminen on arviointia? Se on arviointia siksi, että siinä tarkastellaan kahden eri arvokategorian mukaista toimintaa tai ilmiötä ja toisen arvokategorian mukaisen toiminnan kausaalivaikutukset arvioidaan toisen arvokategorian mukaisin kriteerein.

Vaikutusten arviointia voi tehdä ennakkoarviointina, jälkikäteisarviointina tai molempina. Järkevimmillään sitä tehdään sekä ennakko- että jälkikäteisarviointina, jolloin tosiasiallisten vaikutusten seuranta auttaa kehittämään etukäteisarvioinnin kriteerejä.

Vaikutusten arviointiin liittyy kuitenkin kaksi vakavaa riskiä arvioinnin näkökulmasta. Jos vaikutusten arviointi jää pelkäksi ennakkoarvioinniksi ja se muuttuu hallinnolliseksi rutiiniksi, jossa sovelletaan täysin sopimuksenvaraisia kriteerejä ilman erityistä empiiristä näyttöä, ei ole enää mielekästä puhua arvioinnista. Kyseessä on pikemminkin ennakkotarkastus. Emme yleensä

puhu sähkötarvikkeiden, lelujen tai pornofilmiä vaikutusten ennakoarvioinnista, vaan niiden ennakkotarkastuksesta.

Toinen vakava riski liittyy arviointivoimavarojen kohdentamiseen ja käyttöön sekä arvioinnin rehellisyyteen ja integriteettiin. Julkisen sektorin arvo on sen tosiasiallinen vaikuttavuus ja arviointiponnistelut tulisi ensisijaisesti kohdistaa tosiasiallisen vaikuttavuuden arviointiin. On vakavia merkkejä siitä, että tosiasiallisen vaikuttavuuden arvioinnin sijasta voimavaroja aletaan suunnata hallinnolliseen ennakoarviointiin. Kun ennen byrokratia tuotti loputtomat määrät mitä erilaisempia ja yksityiskohtaisempia suunnitelmia seurannan jäädessä lähes olemattomaksi, nyt nuo samat paperit tuotetaan vaikutusten ennakoarvioinnin nimellä ja pyritään antamaan vaikutelma, että emme vain suunnittele, vaan myös arvioimme.

On helppo nähdä kuinka hallintoa voidaan rajattomasti paisuttaa tai se voidaan hukuttaa mitä erilaisimpien yksittäisten asioiden tai päätösten sivuvaikutusten oletuksiin perustuvaan ennakoarviointiin mitä erilaisimmilla elämäntilanteilla. Kaikkein vaarallisinta on, jos oletettujen sivuvaikutusten ennakoarvioinneilla pyritään antamaan kuva, että organisaatiossa tehtäisiin vaikuttavuuden arviointia.

VAIKUTTAVUUDEN TÄSMÄLLINEN MÄÄRITELMÄ

Seuraavassa määritellään vaikuttavuus eksaktisti Nugent, Sieppert & Hudsonin (2001) esitystä väljästi mukailien (myös Hudson 1997).

1. Vaikuttavuus on suhde, joka muodostuu kahden erillisen elementin välille. Mikään asia ei ole vaikuttavaa sinällään, vaan vain sen suhteen kautta, mikä sillä on yhtälön toiseen muuttujaan. On siis kyettävä tunnistamaan ja tietämään jotain suhteen molemmista muuttujista.
2. Vaikuttavuussuhteen toista elementtiä tai muuttujaa kutsutaan tässä asiakassysteemiksi. Sillä tarkoitetaan mitä tahansa systeemiä, joka on interventioista erillinen, mutta jonka toimintakyvyssä, terveydessä, hyvinvoinnissa jne. pyritään aikaansaamaan positiivista muutosta. Asiakassysteemi voi olla yksilö, perhe, yhteisö tai väestö mutta se voi

olla myös ei-vain-ihmisistä koostuva järjestelmä, kuten hermosto tai rahoitusjärjestelmä. Olennaista on, että asiakassysteemi kyettään jollain tarkkuusasteella tunnistamaan, ja sellainen on myös olemassa.

3. Jotta vaikuttavuutta voisi olla olemassa, on kyettävä havaitsemaan asiakassysteemin hyvinvoinnin muutos eri ajanhetkinä ja on kyettävä vertaamaan hyvinvoinnin (tai toimintakyvyn jne.) tasoa eri havaintojen perusteella. "Ei ole, eikä voi olla sellaista asiaa kuin vaikuttavuus, ennen kuin on olemassa mitattavaa positiivista muutosta asiakkaan hyvinvoinnissa." (Nugent, Sieppert, Hudson 2001, 3).
4. Toinen vaikuttavuussuhteen elementti on interventio, hoito, palvelut eli toiminta, jonka vaikuttavuudesta ollaan kiinnostettu. Ensimmäinen ehto interventioille on, että ne on kyettävä tunnistamaan kohtuullisella tarkkuudella. Interventioiden tunnistamisen kannalta on toivottavaa, että toiminnan tunnistaminen tapahtuu prosessin yhtäläisyyden tai erilaisuuden kautta.
5. Lopuksi tulisi kyetä erottamaan mikä tunnistettu interventio on kohdistunut mihinkin asiakassysteemiin eri ajan hetkinä mukaan lukien se, milloin interventio ei vielä ollut alkanut ja milloin se oli päättynyt.

Näiden ehtojen jälkeen saadaan yksinkertaisen ja eksaktin vaikuttavuuden määritelmä:

Vaikuttavuus on

$$\{ \text{Asiakassysteemin hyvinvoinnin muutos eri ajan hetkinä} \\ = \\ \text{toteutettujen interventioiden funktio} \}$$

Ja sama toisin ilmaisten:

$$V = \{ \Delta \text{ASH}_i = f(I_i) \}$$

jossa

V vaikuttavuus

Δ muutos

ASH asiakassysteemin hyvinvointi tietyllä ajan hetkellä

I tietyntyyppinen aikana käytetyt interventiot

OHJELMATEORIA VAIKUTTAVUUS- TEORIANA

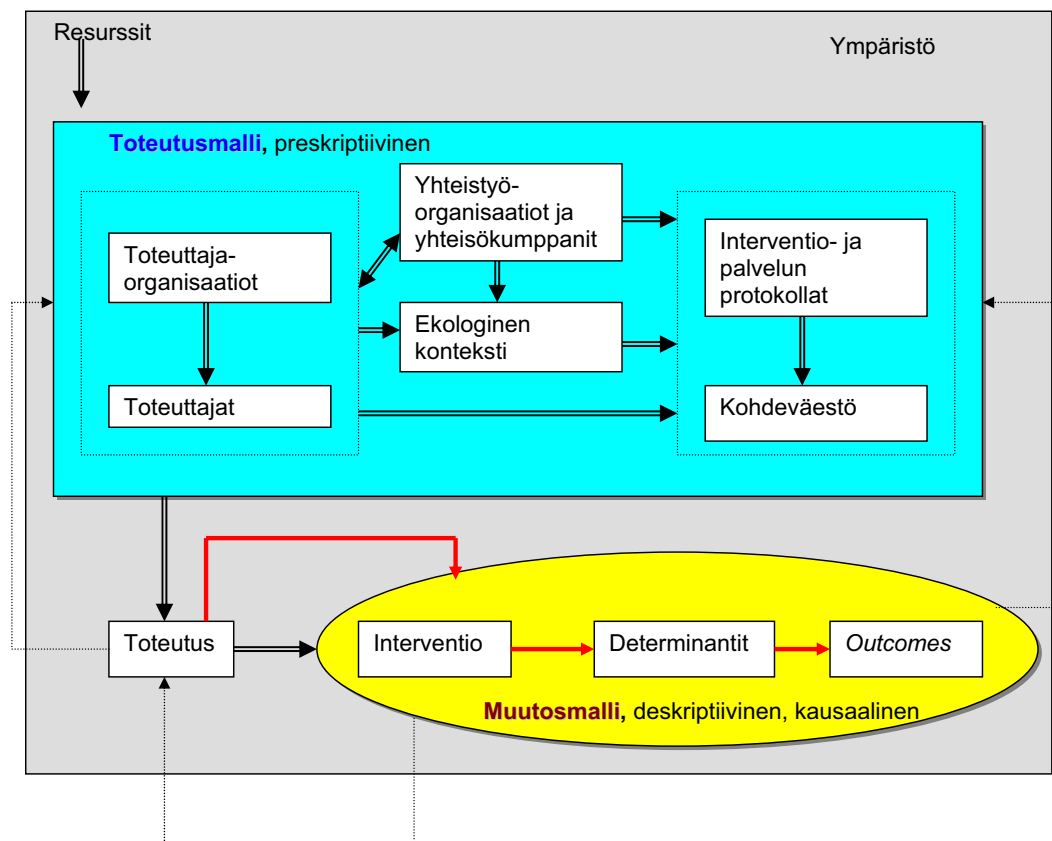
Vaikuttavuuden arvioinnin keskeinen osa-alue on se, mitä kutsutaan käsitteellä program theory. ”Ohjelmateoria” ei suomenkieliselle lukijalle sano oikeastaan mitään ja käsitteen parempi käännös olisi vaikuttavuusteoria. Vaikuttavuusteoriassa on kysymys niistä oletuksista, joiden kautta vaikuttavuuden ajatellaan syntyvän todellisuudessa. Eli se vastaa kysymykseen, minkälainen on tuo edellisen määritelmän funktio.

Vaikuttavuusteoria muodostuu kahdesta eri osasta, muutosmallista ja toteutusmallista (Chen 2005). Muutosmalli kuvaa sitä kausaaliketjua, joka asiakassysteemin toiminnassa ja toimintaympäristössä aikaansaa toimintakyvyn, terveyden tai hyvinvoinnin muutoksen. Keskeinen osa muutosteoriaa ovat ne muutosvoimat tai muuttu-

jat, joiden välityksellä varsinainen hyvinvoinnin muutos tapahtuu. Mitkä ovat esimerkiksi pitkäaikaistyöttömän työllistymisen keskeiset estävät ja edistävät determinantit tai muuttujat.

Vaikuttavuusteorian toinen elementti on preskriptiivinen toteutusmalli, joka vastaa kysymykseen kuinka organisaatiossa on toimittava, jotta kyettäisiin parhaalla mahdollisella tavalla ja tehokkaimmin aikaansaamaan muutosta. Toteutusmalli ei ole perinteinen intervention prosessikuvaus vaan siinä kiinnitetään huomio keskeisiin rakenteellisiin ehtoihin, mitä menestyksekkäällä intervention toteuttamisella on. Toteutusmallin rakenteet ovat interventio-organisaatio, kumppanit, toimintaympäristö ja asiakassysteemi.

Oheisessa kuviossa toteutusmallin nuolet kuvaavat toiminnan ajallista ja prosessuaalisia suhteita, jossa edelliset vaiheet luovat perustan seuraaville. Muutosmallin nuolet taas kuvaavat



Kuvio 3 Vaikuttavuusteorian käsitteellinen viitekehys (Chen 2005, 31)

kausalisuhteita. Katkoviivat kuvaavat palautetta eli arvioinnin kulkua.

Vaikuttavuusteoria on keskeinen osa tosiasiasta vaikuttavuutta. Voi jopa sanoa, että vaikuttavuus on yhtä kuin vaikuttavuusteoriaan kuuluvien oletusten paikkansapitävyys. Näin ollen vaikuttavuusteorian arvioinnin eli sen eksplikoinnin ja testaamisen, tulisi olla keskeinen osa sekä uusien hankkeiden tai palvelujen suunnitteluprosessia että vakiintuneiden palvelujen vaikuttavuuden parantamista. Logiikka tämän taustalla on selkeä: epätosiin vaikuttavuusteorian oletuksiin perustuva palvelu tai hanke ei voi olla vaikuttavaa. Hämäriin oletuksiin perustuva voi satunnaisesti olla vaikuttavaa samassa määrin kuin se voi olla ei-vaikuttavaa ja haitallista. Vastaavasti vaikuttavuuden arviointi systemaattisena osana ammatillista ja organisaation toimintaa mahdollistaa vaikuttavuusteorian riittävän täsmällisen eksplikoinnin ja testaamisen. Siten se edesauttaa asiantuntemuksen ja tietopääoman karttumisessa.

Chenin muotoileman vaikuttavuusteorian erityinen ansio on, että siinä toteutus- ja muutosmallit kytkeytyvät tiiviisti yhteen. Toteutusmallin, eli organisaatorakenteen, yhteistyörakenteiden ja varsinaisen palvelujen tuottamistavan pitäisi aina määräytyä sen asiakassysteemin hyvinvoinnin dynamiikan mukaan, jota ollaan aktivoimassa. Tämä on jo sinällään merkittävä haaste erilaisille julkisen sektorin toimijoille, joissa pääsääntöisesti ajatellaan, että toteutusmalli on lähes sama riippumatta interventiosta. Vaikuttavuusteoria on olennainen parannus laajasti käytössä olevaan tasapainoisen mittariston rakenteeseen, jossa varsinkin hyvinvointipalveluissa asiakkaat ja vaikuttavuusosio jäävät sisällöllisesti hyvin ohueksi ellei sisällyksettömäksi.

REALISTISEN ARVIOINNIN AVAAMA MAHDOLLISUUS

Ray Pawson ja Nick Tilley julkaisivat vuonna 1997 teoksen *Realistic Evaluation*. Realistisen arvioinnin periaatteita soveltaa sosiaalityön ja terveydenhuollon käytännön arviointiin Mansoor A.F. Kazi kirjassaan *Realist Evaluation in Practice* (2003). Realistisen arviointi on merkinnyt ammatillisen käytännön tasolla eräänlaista valankumousta, koska sen kautta vaikuttavuuden arviointi on kyetty mieltämään keskeiseksi osaksi

amatillista käytäntöä (esim. Högnabba ym. 2005).

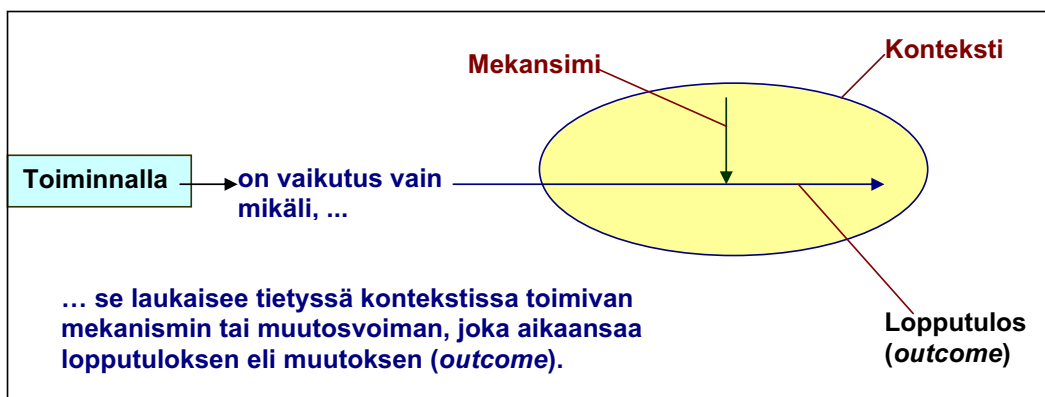
Mitä merkitystä ammatilliselle käytännölle voi olla Pawson & Tilley'n kaltaisesta tieteenfilosofian peruskysymysten ympärillä pyörivästä kirjasta? Realismi on vastaisku "mittarimiehille" eli empiristeille, jotka mittailevat vain tuloksia, mutta huomiota ei kiinnitetä todellisuuden kannalta ydinkysymykseen: kuinka näitä tuloksia voi parhaalla mahdollisella tavalla saada aikaan. Realismi on vastaisku konstruktionisteille, jotka ovat niin etäännyneet todellisuudesta, että sekoittavat puhumisen, diskurssit tai hienommin ilmaistuna puhunnan, todellisuuden kanssa, vaikka ympäröivässä todellisuudessa on todellista kärsimystä riippumatta siitä, mitä puhutaan.

Realistinen arviointi antaa yksinkertaiset ja käytännölliset käsitteelliset välineet vaikuttavuusteorian muutosmallin ymmärtämiseksi ja testaamiseksi. Ammatillisen käytännön näkökulmasta toiminnan ytimenä on tunnistaa, mikä on hyvinvoinnin muutosta ja kuinka se eri olosuhteissa syntyy.

Oheisessa kuviossa on esitetty realismin perusidea.

Kun realistisen arvioinnin lähtökohtia sovelletaan, esimerkiksi tapauskohtaiseen sosiaalityöhön, voidaan päätyä seuraavanlaisiin periaatteisiin, joissa ammatillinen arkipäivän käytäntö ja vaikuttavuuden arviointi nivoutuvat osaksi samaa prosessia.

1. On tiedettävä, miksi vaikuttavuuden arviointia tehdään.
2. On ymmärrettävä, miten muutos ihmistyössä syntyy (eli on ymmärrettävä syy- ja seuraussuhteita).
3. On tuotettava tietoa asiakkaan tilanteesta, tehtävä kattava tilannearvio (konteksti).
4. On tuotettava tietoa muutosta edistävästä ja ehkäisevästä voimista asiakkaan elämässä (mekanismi).
5. On tuotettava tietoa yhdessä sovitusta, tavoitteellisista ja toteutuneista interventioista tai toiminnasta.
6. On tuotettava tietoa tapahtuneesta asiakkaan tilanteen muutoksesta.
7. On osattava hyödyntää syntynyttä informaatiota vaikuttavuuden tietämiseksi ja toiminnan muuttamiseksi.



Kuvio 4 Generatiivinen kausalisuus (Pawson & Tilley 1997, 58)

LOPPUHUOMAUTUS

Viimeistellessäni tämän artikkelin lopullista käsikirjoitusta syyskuun alussa 2006 pidettiin Espoon Dipolissa kansainvälinen konferenssi, EPC 2006 European Productivity Conference. Kolmipäiväisen konferenssin ehkä merkittävin anti oli yhteisymmärrys Suomen korkeimman poliittisen johdon, keskeisten valtiollisten instituutioiden, ammattiyhdistysliikkeen, Euroopan Union toimijoiden, kansainvälisten tutkijoiden ja eri organisaatioiden edustajien kesken. Euroopalaisten ja suomalaisten ihmisten hyvinvoinnin kehitys on ratkaisevasti riippuvainen tulevien vuosien tuottavuuden, erityisesti julkisen sektorin tuottavuuden kehityksestä. Tuottavuusongelman ratkaisun keskeinen avain on nähdäkseni tuottavuuden tarkastelu vaikuttavuuden näkökulmasta.

Tuottavuus vaikuttavuuden näkökulmasta ei ole ensisijaisesti sitä, että terveyskeskuksen tai sosiaalipalvelutoimiston ovet käyvät tiuhaan, vaan sitä, että ihmisten hyvinvointi tosiasiallisesti paranee oven sulkeuduttua selän takana. Tuottavuusnäkökulmalla on myös merkittäviä yhteyksiä vaikuttavuuteen. Tärkeimpiä palvelusektorin tuottavuutta ja vaikuttavuutta lisänneitä oivalluksia on ollut se, että ihmiset eivät enää liiku palvelujen luokse, vaan palvelut tulevat ihmisten luokse.

LÄHTEET

- Argyris, Chris & Schön, Donald A.: Organizational Learning II. Theory, Method, and Practice, Addison-Wesley Publishing Company, Reading, Massachusetts 1996.
- Aristoteles: Nikomakhoksen etiikka. Suomentanut Simo Knuutila. Gaudeamus, Helsinki 1989.
- Aristoteles Poliitiikka. Suomentanut A. M. Anttila. Gaudeamus, Helsinki 1991.
- Chen, Huey-Tsyh: Practical Program Evaluation. Sage Publications, Thousand Oaks, CA 2005.
- Fukuyama, Francis: Historian loppu ja viimeinen ihminen. Suomentanut Heikki Eskelinen. WSOY, Porvoo 1992.
- Hudson, Walter W.: Assessment Tools as Outcome Measures in Social Work. Teoksessa: Mullen, Edvard J. & Magnabosco, Jennifer L. (eds.): Outcomes Measurement in the Human Services. NASW Press, Washington DC 1997.
- Högnabba, Stina & Julkunen, Ilse & Kainulainen, Sakari & Korteniemi, Pertti & Lindqvist, Tuija & Peitola, Petri: Steps into Realist Evaluation in Social Work in Finland. Teoksessa: Sommerfeld, Peter (ed.): Evidence-Based Social Work - Towards a New Professionalism?: Peter Lang, Bern 2005
- Kazi, A.F.: Realist Evaluationon in Practice. Health and Social Work. Sage Publications, London 2003
- Kuhn, Thomas S.: The Structure of Scientific Revolutions. 3rd edition. University of Chicago Press, Chicago 1996.
- Kukkonen, Jarmo: Terveysthuollon vaikuttavuuden arviointi rutiinisti kerätyn tiedon pohjalta. Kuopion yliopiston julkaisuja D. lääketiede 355. Kuopion yliopisto, Kuopio 2005.
- MacIntyre, Alasdair: Hyveiden jäljillä. Moraaliteoreettinen tutkimus. Suomentanut Niko Noponen. Gaudeamus, Helsinki 2004.
- Mathison, Sandra: Preface. Teoksessa: Mathison,

- Sandra (ed.): Encyclopedia of Evaluation.: Sage Publications, Thousand Oaks, CA 2005
- Mark, Melvin M & Henry, Gary T. & Julnes George: Evaluation.: Jossey-Bass, San Francisco 2000
- Nugent, William R. & Sieppert, Jackie D. & Hudson, Walter W.: Practice Evaluation for the 21st Century. Wadsworth/Thomson Learning, Belmont, CA 2001
- Newcomer, Kathryn E. & Hatry, Harry P. & Wholey, Joseph S.: Meeting the Need for Practical Evaluation Approaches: An Introduction. Teoksessa: Newcomer, Kathryn E., Hatry, Harry P., Wholey, Joseph S (Eds.): Handbook of Practical Program Evaluation. Second edition. Jossey-Bass, San Francisco 2004.
- Paasio, Petteri: Vaikuttavuuden arvioinnin rakenne ja mahdollisuus sosiaalialalla. FinSoc työpapereita 3/2003. Stakes, Helsinki 2003.
- Pawson, Ray & Tilley, Nick: Realistic Evaluation. Sage Publications, London 1997.
- Platon: Faidros. Suomentanut Pentti Saarikoski. Teoksessa Platon: Teokset kolmas osa. Otava, Helsinki 1979.
- Rajavaara, Marketta: Yhteiskuntaan vaikuttava Kela. Katsaus vaikuttavuuden käsitteisiin ja arviointiin. Sosiaali- ja terveysturvan katsauksia 69. Kela, Helsinki 2006
- Rossi, Peter H. & Lipsey, Mark W. & Freeman, Howard E.: Evaluation. A Systematic Approach. 7th edition. Sage Publications, Thousand Oaks, CA 2004
- Räsänen, Pirjo & Sintonen, Harri & Ryyänen, Olli-Pekka & Blom, Marja & Semberg-Konttinen, Virpi: Measuring cost-effectiveness of secondary health care: Feasibility and potential utilization of results. International Journal of Technology Assessment in Health Care, 21 (2005), (1): 22-31.
- Senge, Peter M.: The Fifth Discipline. The Art and Practice of The Learning Organizations. Currency Doubleday, New York 1994.
- Scriven, Michael: Evaluation Thesarus. 4th edition. Sage Publications, Newbury Park, CA 1991.
- Sintonen, Harri: The 15D instrument of health-related quality of life: Properties and applications. Ann Med. 33 (2001), (5): 328-336.
- Suomen kielen perussanakirja. Kolmas osa. Kotimaisen kielten tutkimuskeskuksen julkaisuja 55. Edita, Helsinki 2004.

Systemaattiset tutkimuskatsaukset vaikuttavuuden arvioinnin välineenä sosiaaalialalla

Juha Koivisto ja Riitta Haverinen

ABSTRACT

Systematic research reviews as a tool of effectiveness evaluation in social work and in social care

The article analyzes systematic research reviews as a new kind of tool to systematize effectiveness research knowledge in social work and in social care. The main objective of the article is to analyze the possibilities of systematic reviews as a tool to evaluate effectiveness of social interventions in a situation where there are many competitive research paradigms of effectiveness research. First, the background of the birth of effectiveness thematic is regarded. Then three main lines of effectiveness research are outlined, and the practices of some international evaluation institutions to produce systematic reviews are reviewed. After that two systematic reviews that systematize randomized controlled trials (RCT) are presented. At the end, some lessons learnt from the two reviews are presented and RCT as a research paradigm is critically analyzed. The article suggests that systematic reviews should systematize effectiveness research produced by different research paradigms.

JOHDANTO

Vaikuttavuustematiikka on noussut keskiöön eri hallinnonaloilla vuosituhannen vaihteessa. Terveystieteiden alalla vaikuttavuuskeskustelu heräsi jo hivenen aikaisemmin, sosiaaalialalla vaikuttavuus on kuuma keskustelun aihe juuri nyt. Käsitteet vaikuttavuus ja vaikutukset ovat saaneet hyvin erilaisia sisältöjä eri aloilla riippuen siitä, miten interventiota ja kausaalisuutta jäsenetään ontologisesti ja teoreettisesti. Kuitenkin, vaikutuksilla voidaan tarkoittaa laajemmin millaisia tahansa intervention tuottamia, eri tasoille ja aikajanelle eri tavoin asettuvia vaikutuksia. Meklinin (2001, 108-110) mukaan voidaan puhua määrällisistä ja laadullisista, välittömistä ja välillisistä, sekä lyhyt- ja pitkäkestoisista vaikutuksista. Lisäksi vaikutukset voidaan nähdä myönteisinä ja kielteisinä, odotettuina ja tavoiteltavina sekä odottamattomina sivuvaikutuksina. Vaikuttavuus voidaan taas ymmärtää kapeammaksi näkökulmaksi, jossa ollaan kiinnostuneita siitä, missä määrin ennalta määritellyt ja tavoiteltavat tulokset saavutetaan tietyn intervention avulla. Vaikuttavuuskeskustelun ominaispiirre on se, että vaikuttavuutta tarkastellaan ennen kaikkea asiakkaiden kannalta. Tällä tarkoitetaan interventioiden vaikuttavuutta esimerkiksi asiakkaan terveydentilan, olosuhteiden, elämäntilanteen tai hyvinvoinnin kannalta.

Tämä artikkeli keskittyy vaikuttavuustematiikka-

kaan erityisesti sosiaalialalla, jossa vaikuttavuuden arvioinnin peruskysymys asetetaan eri tavoin riippuen siitä, millaisen tutkimusparadigman läpi vaikuttavuutta tarkastellaan. Näyttöön perustuva tieto ja politiikka (evidence based practice and policy) -agenda on ollut sosiaalialalla lähtökohdiltaan vahvasti positivistiseen ja kokeelliseen tutkimusparadigmaan nojaavaa. Mallia on otettu terveydenhuoltoalalta, jossa satunnaistettu kontrolliasema on vaikuttavuustutkimuksen kultainen sääntö. Positivistisen vaikuttavuustutkimuksen peruskysymys on "mikä toimii". Tämä agenda on vahvistunut erityisesti Tanskassa ja Ruotsissa sosiaalialan arviointi-instituuttien työssä. Kummassakin maassa on alettu tuottaa kokeellista vaikuttavuustutkimusta systematisoivia ja tiivistäviä tutkimuskatsauksia. Usko koeasetelmaan ja sen pohjalta tuotetun tiedon hyödyllisyyteen sosiaalialalla on näissä maissa kova. Kokeellisen tutkimusparadigman rinnalle ja kritiikiksi on sosiaalialalla noussut erityisesti realistinen arviointi. Realistista arviointia edustavan Dahler-Larsenin mukaan (2005, 7) vaikuttavuuden arvioinnin peruskysymys on "mikä vaikuttaa mihin, miten, milloin ja millä edellytyksillä".

Tämä artikkeli jäsentää vaikuttavuustematiikkaa lähinnä systemaattisten tutkimuskatsausten viitekehityksessä, joista on tullut uudenlainen väline tuottaa vaikuttavuustietoa. Artikkelin päätaivoitteena on jäsentää ja selkeyttää tutkimuskatsausten mahdollisuuksia vaikuttavuuden arvioinnin välineenä. Tarkastelemme erityisesti sitä, missä määrin kokeellista tutkimusta systematisoivat katsaukset tuottavat luotettavaa ja siirrettävää vaikuttavuustietoa sosiaalialalla. Artikkelin rakenne on seuraava. Tarkastelemme ensin vaikuttavuustematiikan nousun taustaa. Sitten jäsenämme vaikuttavuustutkimuksen päälinjoja sosiaalialalla ja tuomme tämän pohjalta esille joidenkin eurooppalaisten tutkimuslaitosten tapoja tuottaa vaikuttavuustietoa systematisoivia katsauksia. Tämän jälkeen esitämme tiivistetyksi kaksi ryhmätyönä tekemää kokeellista vaikuttavuustutkimusta systematisoinutta tutkimuskatsausta. Erittelemme niissä omia kokemuksiamme ja valintojamme. Tutkimuskatsausten tekijät näkyvät lähdeluettelossa. Keskusteluosiossa nostamme esiin katsauksen tekemisen prosessin pulmia ja opetuksia sekä pohdimme sitä, millaista tietoa toimenpiteiden ja interventioiden vaikuttavuudesta katsauksissa tulisi koota ja systematisoida.

VAIKUTTAVUUSNÄKÖKULMAN NOUSUN TAUSTA

Vaikuttavuustematiikan ajankohtaistumisen taustalla on useita kehityskulkuja, jotka osin risteävät keskenään. Rajavaara (2006) on analysoinut vaikuttavuusnäkökulman esille nousun syitä. Hän viittaa Vedungiin, jonka mukaan (2003) vaikuttavuudessa on laajasti ottaen kyse yrityksestä ohjata julkista sektoria sekä yhteiskuntapolitiikan harjoittamista tilanteessa, jossa vahvat taloudelliset näkökohdat korostuvat. Aiemman radikaalin rationalismin näkökulmasta kokonaisvaltaisuus, tulevaan suuntautuminen sekä huolellinen suunnittelu menetti merkityksensä poliittishallinnollisessa päätöksenteossa ja omaksuttiin ns. julkisten interventioiden insinöörimalli, joka toimii kokeilujen avulla. Kokeilujen arviointi tieteellisesti ja tämän pohjalta parhaiten toimivan käyttöön ottaminen oli ajattelun taustalla, mutta tämä ajattelumalli näytti toimivan kuitenkin heikosti käytännössä.

Uuden julkisjohtamisen doktriinin (New Public Management) saadessa jalansijaa managerihallinto, taloudellisten ylykkeiden korostaminen, tulosten ja tuotosten mittaaminen, kilpailuttaminen sekä kustannustehokkuus osuivat hyvin yhteen vaikuttavuuden ja vaikutusten kanssa. Kustannustietoisuuden lisääntyessä vaikuttavuustietoisuus voisi myös olla kuluttajien ja kansalaisten kiinnostuksen kohteena. Tästä ei kuitenkaan ole merkkejä. Eniten korostetaan Value for money -ajattelua uuden julkisjohtamisen näkökulmasta. Uusi julkisjohtaminen on nostanut esille erilaisia toimijakoalitioita, joissa on mukana yksityisiä ja julkisia toimijoita sekä yrityksiä. Tässä käsitteellisessä yhteydessä vaikuttavuus voidaan nähdä tuloksellisuutta kuvaavien käsitteiden eräänä rinnakkaiskäsitteenä.

Myös kustannustietoisuuden sopeutuva jälkiekspansiivinen hyvinvointivaltion uudistustyö on lisännyt vaikuttavuuden teeman esilläpitoa ja samalla korostanut arviointien tarvetta ylipäätään. Samaa suuntaa toimivat yhteiskunnalliset hallinnan muutokset. Hallinnan käsite korostaa yhteiskuntapoliittisten ja taloudellisten kysymyksen muuttuneita rajoja ja vastuita. Siinä on kyse systeemien tarkastelusta, joka painottaa valtariippuvuuksia, vuorovaikutusta ja neuvottelukykyä eri instituutioiden kesken tulosten saamiseksi eli jotta asiat saataisiin tehdyksi.

Vedung (2003) näkee politiikan tutkijana arvi-

oinnissa ratkaisun demokratiavajeisiin, osallistuvan demokratian ja keskusteleavan kansanvallan kehityspyrkimyksiin. Kansanvaltaisuuden korostaminen ja keinojen kehittäminen on tuonut tilaa uudennlaisille arviointiotteille. Se on lisännyt tarvetta toimintatutkimuksellisille, osallistuville ja dialogisille menettelyille ja siirtänyt arviointia myös kansalaislähtöisen tiedonmuodostuksen suuntaan. Arvioinnin odotusten laajentumista kuvaa myös se, että sitä aletaan pitää työelämän, työn ja palveluprosessien uudistamisen välineenä sekä työntekijöiden osaamisen kehittämisen välineenä

Rajavaaran (2006, 33-34) mukaan vaikuttavuuden arviointia on kehitetty kaikkein innokkaimmin sellaisten hyvinvointipalvelujen alueella, jotka ovat luonteeltaan vuorovaikutuksellisia ja yksilöllisiä. Palveluissa korostetaan vaikutusten vuorovaikutteista syntyprosessia, jossa työntekijän ja asiakkaan keskinäinen vuorovaikutus tuottaa myönteisiä tai kielteisiä seurauksia tai vaikutuksia. Tämä on yksi keskeinen eri vaihtoehtoisia lähestymistapoja koskeva kysymys; miten nähdä intervention ja vaikutuksen suhde. Leiman (2004) on suhtautunut erittäin varauksellisesti vaikuttavuuden tutkimiseen koeasetelmilla vuorovaikutukseen perustuvassa työssä. Hän on tutkinut lähinnä psykoterapiaa ja esittää sen vaikutusten tutkimiseen vaihtoehtoisia lähestymistapoja.

Vahvoja kansainvälisiä vaikuttajia evidence based policy and practice -liikkeessä ovat koe- ja kontrolliasetelmien erinomaisuutta interventioiden vaikuttavuusnäytön tuottamisessa korostavat kansainväliset tutkimuslaitokset ja verkostot (mm. Cochrane ja Campbell -verkostot ja Exeterin yliopiston Evidence based practice -yksikkö). Nämä organisaatiot tuottavat kokeelliseen näyttöön perustuvaa tietoa interventioiden vaikutuksista. Kuitenkin, vaikka koeasetelmien mukaiseen näyttöön perustuvan luotettavan tiedon agenda on saanut suuren jalansijan lääketieteessä, sen asema monissa maissa erityisesti sosiaalialalla on ollut varsin vaatimaton.

VAIKUTTAVUUSTUTKIMUKSEN PÄÄLINJAT SOSIAALIALALLA

Vaikuttavuustutkimusta systematisoivien tutkimuskatsausten laatiminen on erityisen haasteellista tilanteessa, jossa on useita keskenään

kilpailevia ja yhteen sopittamia tutkimusparadigmoja.

Sosiaalialan vaikuttavuuden tutkimuksen lähestymistavat voidaan pelkistää arviointikirjallisuuden perusteella kolmeksi päätutkimuslinjaksi: 1) positivistiseen tiedekäsitykseen, 2) sosiaaliseen konstruktivismiin ja 3) edellisten yhdistämiseen tähtäävään tutkimuslinjaan (ks. Rossi & Freeman 1993; Patton 1997; Rajavaara 1999; Mäntysaari 1999; Robson 2001; Pawson & Tilley 1997). Tutkimuslinjojen sisällä voidaan erottaa useampia lähestymistapoja, mutta tarkastelemme niitä seuraavassa yksinkertaisuuden vuoksi kolmena linjana.

Tutkimuslinjat sisältävät erilaisia käsityksiä sosiaalisten interventioiden ontologiasta ja kausaalisuudesta. Niitä voidaan tarkastella Bruno Latourin (1992) mittarilla, jolla tämä on jäsentänyt modernin tieteen ontologiaa. Latour jäsentää modernin tieteen ontologiaa, jossa luonnon ja yhteiskunnan, materiaalsen ja sosiaalisen tai teknologian ja sosiaalisen välille vedetään selkeä raja. Tässä tapauksessa kahtiajako on intervention menetelmä ja sosiaalinen konteksti (Kuvio 1).

Positivistiseen tiedekäsitykseen perustuvan vaikuttavuustutkimuslinjan mukaan intervention menetelmällä on enemmän tai vähemmän kausaalista voimaa, joka työntää interventiota eteenpäin ja saa aikaan vaikutuksia asiakkaisissa, joilla on tarpeita ja ongelmia. Interventiomenetelmä selittää intervention tuloksen (ks. Kuvio 1, nuoli oikealle). Tämä tutkimuslinja ei keskity interventioprosessin, vaan interventiolla saavutettujen tulosten analysointiin. Tutkimuksen tehtävänä on tällöin tuottaa objektiivista tietoa siitä, millaisia vaikutuksia interventio on saanut aikaan asiakkaiden ongelmissa. Vaikuttavuustieto voi olla objektiivista, yleistettävää ja kumuloituvaa. Sosiaaliset ja subjektiiviset tekijät, kuten intervention antajan omat intressit tai persoonallisuus, eivät saa häiritä intervention kulkua. "Hyvyys" tai "huonous" ovat menetelmän ominaisuuksia. Parhaat menetelmät voidaan löytää ja identifioida. Menetelmän siirtäminen on toimivaksi osoitetun menetelmän kopioimista ja soveltamista muualla samanlaisten asiakkaiden kohdalla.

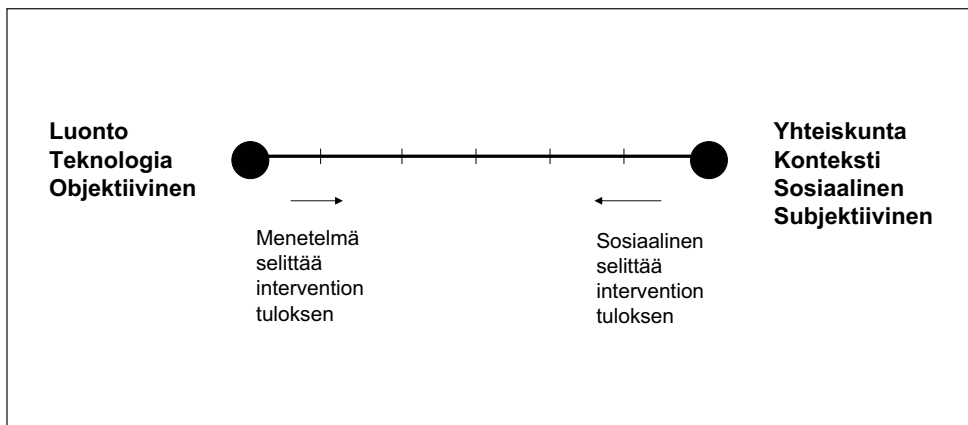
Kokeellinen, satunnaistettu koeasetelma on tämän tutkimuslinjan malliesimerkki (ks. Robson 2001; Campbell & Stanley 1963). Siinä intervention kohderyhmä, samoja ominaisuuksia omaaviksi oletetut henkilöt, esimerkiksi työttömät, jaetaan satunnaisesti (esim. arpomalla) koeryh-

mään ja kontrolliryhmään. Koeryhmälle tuotetaan interventio, esimerkiksi jokin aktivointitoimenpide, kontrolliryhmälle ei. Tämän jälkeen suoritetaan mittaukset tulosmuuttujilla ennen ja jälkeen intervention kummassakin ryhmässä ja katsotaan keskiarvoinen muutos tulosmuuttujien, esimerkiksi aktivoitumisen, suhteen. Satunnaistamisella uskotaan pystyttävän kontrolloimaan interventiota häiritsevien tekijöiden vaikutus. Kvasikokeellisessa koeasetelmassa kohdehenkilöitä ei jaeta satunnaisesti koe- ja kontrolliryhmiin, vaan jako tehdään esimerkiksi sukupuolen, iän tai koulutuksen perusteella. Tämän asetelman ei uskota tuottavan täysin objektiivista tietoa, koska häiritsevien sosiaalisten tekijöiden vaikutusta ei näkökulman mukaan pystytä täysin kontrolloimaan.

Sosiaaliseen konstruktivismiin perustuva vaikuttavuustutkimuslinja (Guba & Lincoln 1989; ks. myös Patton 1997) olettaa, että sosiaalinen konteksti, esim. asiakkaan olosuhteet tai intervention antajan persoonallisuus tai henkilökohtaiset intressit, ovat ratkaisevia tekijöitä sen kannalta, mitä interventiossa tapahtuu. Sosiaalinen selittää intervention tuloksen (ks. Kuvio 1, nuoli vasemmalle). Sosiaaliseen konstruktivismiin nojaavat lähestymistavat tutkivat ensisijaisesti interventio-prosessia ja sen sosiaalista kontekstia. Niiden mukaan intervention menetelmällä ei niinkään ole kausaalista voimaa. "Hyvyys" tai "huonous" määrittyvät relativistisesti kontekstista käsin. Toimivia menetelmiä ovat ne, jotka sopivat parhaiten olemassa oleviin sosiaalisiin olosuhteisiin ja rakenteisiin. Vaikuttavuudesta ei voi tuottaa

objektiivista yleistettävää tietoa, vaan tieto on tilannekohtaista, kulloisenkin kontekstin muokkaamaa. Menetelmän siirtäminen onnistuu, mikäli uusi sosiaalinen konteksti on riittävän yhtäläinen sellaisen aiemman kontekstin kanssa, jossa menetelmä osoittautui toimivaksi.

Realistinen arviointi on yritys yhdistää perinteisen ja konstruktivistisen sekä prosessi- ja tulosarvioinnin vaikuttavuustutkimuksen linjat (ks. Pawson & Tilley 1997). Realistisessa arvioinnissa suhtaudutaan sosiaaliseen realistin tavoin. Sosiaaliset rakenteet, kerrostumat ja olosuhteet ovat sen mukaan olemassa toimijoista riippumatta, vaikka ovatkin näiden tuottamia. Näkemys kausaalisuuteen on puolestaan generatiivinen: menetelmät tai interventiot eivät itsessään tuota muutosta, vaan vasta mekanismien välityksellä. Mekanismit voidaan ymmärtää interventiotoimien aikaansaamana vastakaikuna intervention kohteena olevissa, tietyissä olosuhteissa elävissä, subjekteissa. Esimerkiksi rikollistaustaisiin kohdistetut toimenpiteet voivat herättää heissä kuntoutumishaluja ja edesauttaa kaidalla tiellä pysymistä. Realistinen arviointi tutkii siis sitä, millaisia mekanismeja interventio synnyttää tietyissä olosuhteissa ja miten tulokset syntyvät mekanismien välityksellä. Se voidaan tulkita jonkinlaiseksi keskitien kulkijan lähestymistavaksi. Siinä kausaalisuus nähdään monisuuntaisempaa kuin pelkästään peräkkäisinä syy-seuraussuhteina. Muutokseen vaikuttavat ainakin sosiaalisen kontekstin tekijät, interventiotoimenpiteet ja näiden aikaansaamat mekanismit. Rea-



Kuvio 1. Modernin tieteen ontologia (Latour 1992)

listinen arviointi asettuu positivismiin ja soliaalisen konstruktivismiin välimaastoon selittäessään vaikutuksia niin interventiolla kuin kontekstillakin. Sen kanta interventiomenetelmien siirrettävyyteen on: ”tämä teoria toimii tässä suhteessa, näillä henkilöillä ja tällaisissa olosuhteissa” (Pawson 2001).

Sosiaalialan vaikuttavuuskeskustelussa toisaalta positivistinen tiedekäsitys ja toisaalta realistinen arviointi on merkittävimmin esillä tällä hetkellä. Tarkastelemme seuraavassa viiden ulkomaisen arviointi-instituutin tapoja laatia systemaattisia tutkimuskatsauksia ja sitä millaiseen tutkimusparadigmaan ne nojaavat.

SYSTEMAATTISTEN TUTKIMUSKATSAUSTEN TUOTTAMISPROSESSI

Halusimme tarkastella systemaattisia tutkimuskatsauksia uutena tapana sosiaalialalla koota ja tiivistää vaikuttavuustietoa. Siksi päätimme tehdä tutkimuksen ulkomaisten arviointi-instituuttien lähestymistavoista tuottaa ja tiivistää tietoa ja erityisesti vaikuttavuustietoa. Tarkastelimme viiden ulkomaisen arviointi-instituutin strategioita ja menetelmiä tuottaa systemaattisia tutkimuskatsauksia (Koivisto 2005). Instituutit olivat kansainvälisen Campbell Collaboration -verkon pohjoismainen, Tanskassa toimiva yksikkö Nordic Campbell, Ruotsin sosiaalishallituksen alainen sosiaalityön metodi-insituutti IMS (Institutet för utveckling av metoder i socialt arbete) sekä Englannissa toimivat riippumattomat EPPI-Centre (The Evidence for Policy and Practice Information and Co-ordinating Centre), kahdeksan tutkimusyksikön verkosto EvidenceNetwork ja Social Care Institute for Excellence (SCIE). Instituuttien työmuotona on nimenomaan kerätä ja systematisoida sosiaalialan vaikuttavuustietoa, ei niinkään tehdä primääritutkimusta.

Systemaattisia vaikuttavuus- ja muuta tietoa keräviä katsauksia kutsutaan tarkasteltavissa instituuteissa monilla eri nimikkeillä, mikä kuvaa terminologian vakiintumattomuutta. Niitä kutsutaan systemaattisiksi kirjallisuuskatsauksiksi, systemaattisiksi tutkimuskatsauksiksi, systemaattisiksi katsauksiksi ja systemaattisiksi tietokatsauksiksi. Käytämme tässä artikkelissa termiä systemaattiset tutkimuskatsaukset.

Systemaattiset tutkimuskatsaukset eroavat perinteisimmistä tutkimuskatsauksista siten, että

niissä määritellään etukäteen katsauksen kohde tarkasti, samoin avainkäsitteet sekä kriteerit, joilla tutkimuksia otetaan mukaan analyysiin tai jätetään tarkastelun ulkopuolelle. Tarkoitus on kartoittaa etukäteen määritellyn asetelman avulla kaikki aihealueen materiaali elektronisista tietokannoista ja muista lähteistä, käsihakuna ja henkilökohtaisten kontaktien kautta. Systemaattisten tutkimuskatsausten tekoa varten on lisäksi kehitetty www-pohjaisia työkaluja tiedon keräämistä ja varastointia varten. Nämä mahdollistavat saman katsauksen työstämisen eri puolilta maailmaa ja katsauksen myöhemmän päivittämisen sekä tietojen hyödyntämisen uusia katsauksia laadittaessa. Systemaattisten tutkimuskatsausten tuottamisen peruseriaatteita ovat kaikenkattava eksplisiittisyys ja läpinäkyvyys. Eksplisiittisyydellä ja läpinäkyvyydellä on tarkoitus varmistaa se, että katsausten hyödyntäjät näkevät, millaisen prosessin tuloksena katsauksen johtopäätöksiin on tultu.

Ketkä voisivat hyötyä systemaattisista tutkimuskatsauksista? Niitä tuotetaan sosiaalialalla asiakaspalvelutyötä tekeville ammattilaisille, palvelujen käyttäjille, kouluttajille ja päätöksentekijöille. Käyttäjien osallistaminen katsausten tuottamisprosessiin on noussutkin keskeiseksi katsausten tuottamisen periaatteeksi. Jokainen tarkastelemistamme instituutista pyrkii jollakin tavalla osallistamaan käyttäjät katsausten tuottamisprosessiin. Käyttäjää pyydetään esimerkiksi mukaan katsauksen aihealueen valintaan, kysymyksenasettelun määrittelyyn tai kommentoimaan aihealueen merkitystä ja relevanssia, osallistumaan laajemmin protokollan laadintaan, kommentoimaan katsauksen eri vaiheiden etenemistä, arvioimaan tuloksia ja niiden käytännön hyötyä tai sovellettavuutta sekä kirjoittamaan tiivistelmiä katsauksesta. Osallistamistategialle on ominaista kumppanuussuhteiden ja yhteistyöverkostojen rakentaminen potentiaalisten tiedonkäyttäjien kanssa.

Tarkastelemissamme instituuteissa katsausprosessi jäsennetään pääpiirteissään samantapaiseksi. Katsauksen alussa laadittavassa protokollassa, joka on katsauksen metodologinen suunnitelma, määritellään katsauksen aihealue, kysymyksenasettelu ja eri vaiheissa tehtävät metodologiset ja muut valinnat. Katsauksen tuottamisprosessin keskeisimmät vaiheet ovat katsauksen tarpeen ja kysymyksenasettelun määrittely, kaiken relevantin tutkimuksen ja muun materi-



aalin haku, materiaalin tarkempi valinta, tiedon kerääminen materiaalista ja materiaalin laadun arviointi, tiedon syntetisointi (meta-analyysi, narratiivi tai jokin muu) sekä raportin laatiminen (ks. myös Boaz ym. 2002).

Katsausten tuottamisprosessin erot tarkastelemme instituuttien kesken ovat lähinnä siinä, millaisella tutkimusotteella tuotettuja tutkimuksia ja mitä muuta mahdollista materiaalia katsauksiin hyväksytään. Campbell -verkosto noudattelee terveydenhuollossa vallalla olevaa positivistista näkemystä. Katsauksiin hyväksytään ensisijaisesti satunnaistettuun, kontrolloituun koeasetelmaan perustuvia tutkimuksia ja toissijaisesti ei-satunnaistettuun koeasetelmaan perustuvia tutkimuksia. Myös IMS:n ja EPPI-keskuksen lähökohdissa on Campbellin kaltaisia oletuksia, mutta ne eivät ole niin selkeästi eksplikoituja kuin Campbellillä. Ne ottavat vaikutteita myös realistisesta arvioinnista ja asettavat perusarviointikysymyksen muotoon ”mikä toimii, kenelle ja missä olosuhteissa”. SCIE edustaa instituuteista toista ääriä. SCIE kerää ensinnäkin erilaisten tutkimusasetelmien tuottamaa tietoa sosiaalityöstä ja toiseksi myös muunlaista materiaalia eri tiedon lähteistä, esimerkiksi policy -dokumentteja, käyttäjien kokemuksia ja sosiaalialan ammattilaisten näkemyksiä koskevaa materiaalia. Myös SCIE asettaa perusarviointikysymyksen realistisen arvioinnin tapaan. Evidence Networkin yksiköillä ei ole yhteisiä periaatteita sille, millaisia tutkimuksia tai muuta materiaalia katsauksiin hyväksytään. Verkoston kirjoituksissa on kuitenkin vahvasti esillä realistisen arvioinnin periaatteet.

Laadun arvioinnissa instituutit ovat keskittyneet ensisijaisesti systematisoitavan tutkimuksen metodologian laadun arviointiin. Yleisesti hyväksytyjä tutkimuksen laadun arvioinnin kriteerejä ei ole olemassa, joten tarkasteltavat instituutitkaan eivät aseta ehdottomia vaatimuksia käytettävien kriteerien osalta. Tutkimuksen laadun arviointi on keskittynyt pääasiassa tutkimusasetelman validiuden sekä tutkimuksen toistettavuuden ja tutkimustulosten yleistettävyyden arviointiin. Validiuden arviointi perustuu positivistisen tiedekäsityksen ajatukseen, jonka mukaan tutkimusasetelman tulee olla sellainen, että se estää vääristymät tutkimustuloksissa eli paikallisen kontekstin vaikutuksen. Kokeellisen tutkimuksen osalta arviointi on keskittynyt erityisesti siihen, missä määrin jako koe- ja kontrolliryhmiin oli satunnaistettu, oliko arvioija sokaistu ja mil-

lainen kato tutkimukseen osallistujien osalta oli. Laadullisen tutkimuksen arvioinnille ominaisia kriteereinä ovat vaatimus tutkimuksen kontekstin kuvauksesta ja tutkijan näkökulman huomioiminen.

Pawsonin ym. (2003) SCIE:lle jäsentämät sosiaalialan tiedon lajien arviointiin tarkoitetut yleiset laatukriteerit on avaus laadun arvioinnin laajentamiseksi yli tiedonalueiden sisäisten laatukriteerien. Esimerkiksi tutkimustiedon osalta tämä merkitsee sitä, että metodologian arvioinnin lisäksi arvioidaan tutkimuksen läpinäkyvyyttä, tarkkuutta, tarkoituksenmukaisuutta, hyödyllisyyttä, oikeudenmukaisuutta ja käytettävyyttä. Samoja kriteereitä voidaan alustavasti soveltaa arvioitaessa muitakin tiedon lajeja: organisatorista tietoa, sosiaalialan ammattilaisen tietoa, poliittisen yhteisön tietoa ja sosiaalipalvelujen käyttäjän tietoa.

Synteesivaiheessa Campbell -katsauksissa tähdätään ensisijaisesti vaikuttavuustutkimuksista kerättyjen tulosten eli interventioiden nettovaikutusten tilastolliseen meta-analyysiin. Muut instituutit tavoittelevat sekä meta-analyysijä että narratiivisia synteesejä tai näiden yhdistelmiä. Narratiivinen synteesi on sanallinen ja tematisoiva. Narratiivinen katsaus voi sisältää sekä kvantitatiivisia että laadullisia tutkimustuloksia. EPPI-keskuksen viimeaikaisissa katsauksissa on otettu tavoitteeksi meta-analyysin ja narratiivisen synteessin yhdistäminen omaksi synteeksiksi. SCIE erottaa toisistaan lisäksi alustavat (scoping) että arvioivat (evaluative) katsaukset. Alustava katsaus ei tutki kaikkea materiaalia. Niiden ensisijaisena tarkoituksena on selvittää tutkimuskysymyksen luonnetta, kartoittaa relevantin materiaalin laajuutta sekä tehdä karkea arviointi tiedon yhtenäisyydestä ja laadusta. Arvioivien katsausten tavoitteena on paikantaa kaikki relevantti materiaali, arvioida materiaalin sisältämän näytön vahvuus ja antaa toimintasuosituksia.

Jotta systemaattisen tutkimuskatsauksen tuottamisen prosessi konkretisoituisi, tarkastelemme seuraavassa kahta Stakesin tutkijoiden yhteistyönä toteuttamaa tutkimuskatsausta (Westman ym. 2005; Toljamo ym. 2005). Niissä noudatettiin tutkimuskatsausten tyypillisintä muotoa eli niissä analysoitiin systemaattisesti kokeelliseen vaikuttavuustutkimukseen perustuvia tutkimuksia. Tarkastelemme näiden kahden katsauksen tuottamisen prosessia, esille nousseita ongelmia ja saatuja tuloksia.



PERHEINTERVENTIOIDEN VAIKUTTAVUUS

Tavoitteet

Westmanin ym. (2005) katsauksen tavoitteena oli arvioida perheinterventioiden vaikuttavuutta lapsiperheiden sosiaalisten ongelmien ratkaisemisessa, vähentämisessä tai ennaltaehkäisemisessä kokoamalla systemaattisesti yhteen aiheeseen liittyvät satunnaistettuihin kokeellisiin ja kvasikokeellisiin asetelmiin perustuvat tutkimustulokset.

Katsauksen kohderyhmänä olivat sosiaalisista ongelmista kärsivät lapsiperheet, joissa oli vähintään yksi aikuinen ja vähintään yksi lapsi (0-17-vuotias). Perheisiin kohdistuvalla interventiolla katsauksessa tarkoitettiin ammatillisen työntekijän tekemää työtä tietyn menetelmän tai menetelmien avulla. Katsauksella haettiin vastausta seuraaviin kysymyksiin: millaisiin lapsiperheiden ongelmiin perheinterventioilla on vaikutusta, mitkä perheisiin kohdistuvat interventiot auttavat lapsiperheitä heidän sosiaalisten ongelmien ratkaisemisessa tai tilanteen parantamisessa, ja mikä on perheinterventioiden vaikuttavuusnäyttö.

Tiedonhaku

Tutkimusmateriaalin haun ensimmäisessä vaiheessa aihe rajattiin sosiaalityön perhetyöhön. Tämän perusteella tehtiin katsauksen alustava metodologinen suunnitelma ja etsittiin aiheeseen liittyviä aiempia tutkimuksia. Haun ensimmäisellä kierroksella haettavat käsitteet olivat sosiaalityö ja perhetyö tai perhe/lapsi/vanhemmat ja vaikuttavuus. Hakujärjestyksenä oli 1. sosiaalityö, 2. perhe tai lapset ja 3. vaikuttavuus. Kaikkia käsitteitä haettiin sekä suomeksi että englanniksi. Hakuvastuu jaettiin yhden tutkijan ja informaation kesken.

Sosiaalityö ja perhe osoittautuivat haun kannalta ongelmallisiksi käsitteiksi. Tiedonhaun tuloksena oli satoja viitteitä, joista vain murto-osa käsitteli perhetyön menetelmien vaikuttavuutta. Kaikissa sosiaalityön alueelle sijoittuvissa viitteissä ei pidetä tarpeellisena mainita asiasanana sanaa sosiaalityö, mikä hankaloitti hakua. Valintakriteerit täyttäviä artikkeleita löytyi yhteensä vain 11 kappaletta. Tämän jälkeen päätettiin, että tavoitteena on etsiä vaikuttavuustutkimuksia

perheisiin kohdistuvista interventioista - ei sosiaalityön menetelmistä. Sosiaalityö jätettiin hakusanoista pois.

Toisessa haussa haluttiin etsiä perhetyöhön liittyviä julkaisuja (perhetyö -hakusanat), joissa oli tutkittu intervention vaikuttavuutta (vaikuttavuus -hakusanat) kokeellisen tai kvasikokeellisen asetelman avulla (tutkimusmenetelmät -hakusanat). Haussa käsiteiden järjestys oli 1. perhetyö, 2. vaikuttavuus ja 3. tutkimusmenetelmät (esim. family work + effectiveness + randomized controlled trials). Käytetyt tietokannat olivat samat kuin ensimmäisessä haussa, yhteensä noin 20 kotija ulkomaista tietokantaa. Toisen haun tulos oli jonkin verran tarkempi kuin ensimmäisellä hakukierroksella, koska sosiaalityö -käsitteen monimerkityksisyys ei enää aiheuttanut ongelmia.

Tutkimusaineiston hakuprosessi oli kokonaisuudessaan melko ongelmallinen. Ensimmäiseksi tiedonhaun ongelmaksi osoittautuivat hakukäsitteistä johdetut hakusanat. Tiedonhaussa hakusanoina käytettiin erilaisia perheinterventioiden yläkäsitteitä. Toteutunut haku ei välttämättä ole kattava, koska usein tutkimusten pääkäsitteenä on jokin perheintervention alempi menetelmä tai ohjelma. Näiden etsiminen olisi kuitenkin ollut mahdotonta, koska interventioiden variaatio oli laaja. Toisena ongelmana oli tietokantojen teknisen toimivuuden erilaisuus. Usein tietokantojen kapasiteetti ei riittänyt moniosaisen haun tekemiseen. Useampiosaisia käsitteitä (esim. family therapy) saattoi käyttää useimmissa tietokannoissa hakusanoina ongelmitta, kun taas jotkut tietokannat tulkitsevat hakutermin kahdeksi eri sanaksi, jolloin ohjelma kykeni etsimään hakutermin osia vain erillään. Tämä tuotti runsaasti epärelevanttejä viitteitä hakutulosten joukkoon. Science Direct-tietokanta esimerkiksi tuotti tarkimmallakin hakulauseella yli 2000 viitettä. Yleinen ongelma tietokannoissa oli, että ne vääräristivät aihealueen tutkimuskuvaa angloamerikkalaiseksi.

Tutkimusten valintakriteerit ja analysointi

Katsaukseen valittiin vain kokeellisella (koe- ja kontrolliryhmä) tai kvasikokeellisella asetelmalla (luonnolliset samanaikaiset vertailuryhmät) tehdyt tutkimukset. Annetuilla hakusanoilla löydettiin yhteensä 2 745 viitettä, jotka käytiin yksittellen läpi inklusio- ja eksklusiokriteereiden

avulla.

Katsauksen sisäänottokriteereitä olivat seuraavat: 1) Tutkimusasetelma on kokeellinen (randomised) tai kvasikokeellinen (quasi-experimental), 2) intervention kohderyhmä on lapsiperhe, jossa on vähintään yksi aikuinen ja yksi lapsi (0-17-vuotias) ja vähintään yksi sosiaalinen ongelma (lapsen tai vanhemman/vanhempien ongelma), 3) intervention tavoitteena on lapsiperheen sosiaalisten ongelmien ratkaiseminen, vähentäminen tai ennaltaehkäiseminen, 4) interventio on ammatillisen koulutuksen saaneen työntekijän tekemä perheisiin kohdistuva interventio tai interventioiden yhdistelmä ja 5) tutkimuksen tavoitteena on perheintervention vaikuttavuuden arvioiminen. Poissulkemiskriteerit olivat seuraavat: 1) teoreettinen ja kuvaileva kirjallisuus, tutkielmat, projektiraportit, 2) erityiskohderyhmät tai ongelmatilanne kuten perheväkivalta, sairaus, seksuaalinen hyväksikäyttö sekä vammaisuus, 3) tutkimusten päällekkäisyys (sama tutkimusaineisto) ja 4) psykiatria, kliiniseen päihdehuoltoon, laitoshuoltoon (esim. vankila) ja kouluun liittyvä painotus

Lapsiperheen sosiaalisilla ongelmilla tarkoitettiin lapsen ongelmia (esim. ongelmat koulussa tai päiväkodissa, käytösongelmat, vanhempien tai viranomaistahojen huoli), vanhempien ongelmia (esim. kasvatustarvikkeet, sosiaaliset, taloudelliset, asumiseen liittyvät, alkoholi-, huume- tai mielenterveysongelmat) tai muita perheen ongelmia.

Vastuullinen tutkija kävi läpi jokaisesta tietokannasta löydetty viitteet (2 745 kpl) ja valitsi sen perusteella ensimmäisen vaiheen tutkimukset otsikon perusteella. Tämän jälkeen valinnat tehtiin tutkimusten abstraktien perusteella (64 kpl). Tähän valintaprosessiin osallistui koko tutkimusryhmä. Valinnat jaoteltiin niin, että jokaisen artikkelin valintaan osallistui kaksi tutkijaa. Mikäli valinnoista oli erimielisyyksiä, tehtiin valinta koko tutkimusryhmän kesken. Kolmannessa vaiheessa vastuullinen tutkija kävi läpi tarkemman analyysin pohjalta valitut tutkimukset (44 kpl) artikkelien perusteella ja poisti niistä 13 kappaletta. Lisäksi yhtä sisäänottokriteerit täyttävää tutkimusta ei ehditty saada aikataulun puitteissa katsaukseen, joten lopullinen tutkimusten määrä oli 30 kappaletta. Sisäänottokriteerit täyttäviä kotimaisia julkaisuja tietokannoista ei löytynyt.

Tutkimusten analyysi tapahtui tiedonkeruulomakkeen avulla, joka suunniteltiin katsausta

varten. Tiedonkeruulomakkeen avulla pyrittiin etsimään jokaisesta tutkimukseen valitusta artikkelista samat tiedot asettamalla aineistolle kysymyksiä. Tietotarpeet määriteltiin etukäteen ja tutkimuksia tarkasteltiin kysymyksiä vasten. Olenaisinta oli intervention, tulosmuuttujien ja tulosmittareiden kuvaus sekä vaikuttavuuden esille saaminen.

Analyysin jälkeen tutkimusten tiedot purettiin taulukoihin. Analysointivaiheessa todettiin, että tutkimukset ovat niin heterogeenisiä, että niiden tarkastelu yhtenä ryhmänä ei ole mahdollista. Tätä tuki myös tutkimuksissa käytettyjen tulosmittareiden suuri määrä. Tutkimukset olivat hyvin erilaisia niin kohderyhmältään, interventioltaan, tulosmuuttujiltaan kuin tulosmittareiltaan. Kokonaisnäkömyksen saamiseksi ne päätettiin luokitella neljään luokkaan kohderyhmänsä perusteella: 1) lapsen ja nuoren käyttäytymisongelmiin, 2) nuoren vaikeaan riskitilanteeseen, 3) perheen riskitilanteeseen tai 4) perheen köyhyyteen.

Tutkimusten laadun arviointi

Tutkimusten laatua arvioitiin katsauksessa Jadad -skaalan satunnaistaminen ja poistuma-kysymyksen avulla. Kysymykset olivat: 1) kuvailtiinko tutkimus satunnaistetuksi (tämä sisältää sanojen 'randomly', 'random' ja 'randomization' käytön)? ja 2) kuvailtiinko tutkimuksessa osallistujien peruutukset ja keskeytykset? Jadad -skaalassa satunnaistaminen (Randomization) määritellään seuraavasti: Menetelmää, jolla pyritään yleistämään satunnaistamisen järjestys pidetään oikeana, jos se antaa jokaiselle tutkimuksen osallistujalle samanlaisen mahdollisuuden päästä jokaiseen interventioon. Lisäksi tutkijat eivät voi ennustaa mikä hoito on seuraavana. Jakomenetelmää syntyajan, sisäänpääsyajan, sairaalanumeron tai vuorottelun mukaan ei katsota oikeaksi. Tutkimusten poistuma tuli olla kuvattuna seuraavasti (Withdrawals and dropouts): Osallistujat, jotka otettiin mukaan tutkimukseen, mutta eivät olleet mukana havainnointiajan loppuun, tai jotka eivät sisällyneet analyysiin, on kuvattava tutkimuksessa. Artikkeleissa tulisi mainita jokaisen ryhmän peruutusten ja keskeytysten määrä sekä syyt. Jos tutkimuksessa ei ollut poistumia, pitää se mainita artikkelissa. Mikäli poistumaa ei ole mainittu, ei artikkeli

saa lainkaan pisteitä tältä osin.

Tutkimusten laatu arvioitiin antamalla 1 piste jokaisesta 'kyllä' ja 0 pistettä jokaisesta 'ei' -vastauksesta. Tutkimukselle annettiin yksi lisäpiste, jos koeasetelmassa oli käytetty satunnaisluku- taulukkoa tai tietokoneella luotua satunnaistamista. Mikäli satunnaistaminen oli kuvailtu ja se oli epäasianmukainen (osallistujat valittiin vuorot- tain, sairaalanumeron syntymäajan tai muun sel- laisen mukaan), vähennettiin tutkimukselta yksi piste.

Aineisto

Tutkimusaineisto muodostui 30 ulkomaisesta perheinterventioiden vaikuttavuustutkimuksesta, jotka oli raportoitu¹. Artikkelit oli julkaistu vuosina 1996 - 2005. Suurin osa tutkimuksista oli yhdys- valtalaisia (20 kpl). Amerikkalaisten tutkimusten suuri määrä voi johtua sekä käytetyistä tieto- kannoista ja hakukielestä että vaikuttavuustutki- muksen suuremmasta määrästä näissä maissa. Eniten on 7-12 kuukautta kestäneitä tutkimuksia, mutta yli 2 vuotta seurattuja interventioita on 9 kappaletta.

Tutkimusten yhteenlaskettu otoksen määrä on 13 350 henkilöä ja lopullisissa analyyseissa on ollut mukana yhteensä 12 427 henkilöä. Tutki- musten interventioyryhmissä oli mukana yhteensä 6 605 henkilöä (n=26) sekä koe 2 ja kontrolliryh- missä yhteensä 5 947 henkilöä (n=26). Kaikissa artikkeleissa ei ilmoitettu otoksen poistuman määrää, mutta saatavilla olevan tiedon mukaan katoprosenttien keskiarvo oli 18 % (n=22).

Tutkimusten tarkastelu osoitti, että perheille tehtyjen interventioiden kohderyhmät ja syyt ovat heterogeenisiä ja monipuolisia (Taulukko 1). Tut- kimusten interventiot olivat niin ikään monipuo-

lisia ja heterogeenisiä. Käsité perheinterventio sisältää monimuotoisia ja erilaisia palveluita tai hoitoja asiakkaille, koulutuksesta kotikäynteihin ja terapiaan. Interventiot luokiteltiin seitsemään eri luokkaan. Näitä ovat 1) koulutus ja tiedotta- minen, 2) ohjaus ja neuvonta, 3) tilannearviointi ja tiedonkeruu, 4) hoito, kuntoutus ja terapia, 5) ryhmätyö ja vertaistuki sekä 6) monimuotoinen palvelu.

Hoitoa, kuntoutusta tai terapiaa käytettiin inter- ventiona yli puolessa tutkimuksista (n=16). Loput interventiot olivat niin sanottuja kevyehköjä inter- ventioita. Eniten kevyehköissä interventioissa hyödynnettiin ohjausta ja neuvontaa (n=13). Yhteensä kymmenessä interventiossa oli käy- tetty useampia kuin yhtä luokiteltua interventiota. Suurin osa monimuotoisista interventioista oli kevyehköjä interventioita (n=7). Kevyehkön inter- vention tunnuspiirteenä oli, ettei niissä käytetty hoitoa, kuntoutusta tai terapiaa yhtenä interven- tiomuotona.

Perheinterventioiden vaikuttavuutta arvioitiin useasta eri lapsen ja vanhemman näkökulmasta. Tulosuuttujina oli useita päämuuttujia ja ala- muuttujia. Tulostittareiden määrä ja heterogee- nisuus oli vielä suurempi. Koska tutkimusten tulostittareina käytettiin Yhdysvalloissa validoi- tuja mittareita, jotka oletettiin tunnetuiksi eikä niitä selitetty tarkasti, tutkimusten analysoiminen ja luotettavuuden esille saaminen vaikeutui.

Suurin osa tutkimuksista perustui satunnaistet- tuun koeasetelmaan. Satunnaistamisesta kym- menen tutkimusta ei saanut lainkaan pisteitä, 18 tutkimusta sai yhden pisteen ja kaksi tutkimusta kaksi pistettä. Otoksen poistuman määrästä ja syistä alle puolet tutkimuksista (47 %) oli rapor- toitu hyvin ja loput niin, etteivät saaneet laa- tustettua poistuman tiedottamisesta lainkaan. Vain yksi artikkeleista sai täydet pisteet eli 2 pis-

Taulukko 1. Interventioiden kohderyhmät ja syyt

Kohderyhmä	N	Interventioiden syyt
Lapsen tai nuoren käyttäytymisongelmat	15	Lapsen tai nuoren aggressiivinen tai vastustava käyttäytyminen
Nuoren vaikea riskitilanne	7	Nuoren monimutkainen päihde-, rikollisuus tai muu vastaava ongelma
Perheen riskitilanne	4	Teinivanhemmuus
Perheen köyhyys	4	Vähävaraisuudesta kumpuavat ongelmat

tettä satunnaistamisesta ja yhden pisteen poistuman raportoisesta. 12 tutkimusta sai vähintään yhden pisteen kummastakin laatuksriteristä.

Tulokset

Puolet katsauksen artikkeleista liittyi lasten ja nuorten käyttäytymisongelmiin (n=15) (Taulukko 2). 12 tutkimusta käsitteli 2-12 -vuotiaiden lasten käyttäytymisongelmia ja kolme tutkimusta 6-17 -vuotiaiden lasten ja nuorten käyttäytymisongelmia. Yli puolet (60%) lasten ja nuorten käyttäytymisongelmiin liittyvistä tutkimuksista oli käyttänyt ainakin yhtenä interventiomuotona hoitoa, kuntoutusta tai terapiaa (n=9). Kuudessa tutkimuksessa oli käytetty monimuotoista interventiota. Näiden tietojen pohjalta näyttää siltä, että lasten ja nuorten käyttäytymisongelmien ratkaisemisessa uskotaan tarvittavan joko niin sanottua syvälliseen vuorovaikutukseen perustuvaa (hoito, kuntoutus ja terapia) tai monimuotoista hoitoa. Pelkästään yhtä kevyemmän tason interventiomuotoa (koulutus ja tiedottaminen, ohjaus ja neuvonta, tilannearvio ja tiedonkeruu tai ryhmätyö ja vertaistuki) käytettiin vain kahdessa tutkimuksessa.

Tutkimusten mukaan kaikki lasten ja nuorten käyttäytymisongelmien ratkaisemiseen käytetyt interventiot olivat vähintään kohtalaisen vaikuttavia. Yhdessäkään tutkimuksessa ei saatu huonoja vaikuttavuustuloksia. Tulosten pohjalta näyttää siltä, että perheinterventiot todella toimivat lasten ja nuorten käyttäytymisongelmien hoitomuotoina. Tarkasteltaessa vähintään neljän kuukauden seuranta-ajan jälkeen arvioituja interventioita, olivat tulokset edelleen hyviä. Interventiot olivat vaikuttavia myös pitkällä aikavälillä.

Seitsemässä artikkelissa oli tutkittu nuorten monimuotoiseen tai vaikeaan riskitilanteeseen sovellettujen interventioiden vaikuttavuutta (Taulukko 2). Käytettyjen interventioiden nimet olivat Specialist support teams for adolescents, Multisystemic therapy (MST), The family check-up (FCU), Multidimensional family therapy (MDFT) ja Brief family intervention: Iowa strengthening families program

Viidessä tutkimuksessa oli käytetty interventiona jotain hoitoa, terapiaa tai kuntoutusta. Toisin kuin lasten ja nuorten käyttäytymisongelmien hoidossa, nuorten riskitilanteen ratkaisemisessa ei käytetty lainkaan monimuotoisia interventioita.

Tulokset viittaavat siihen, että nuorten vaikeassa riskitilanteessa interventioksi valitaan useimmiten syvälliseen vuorovaikutukseen perustuva interventiomuoto. Multisysteemisen terapian vaikuttavuutta oli arvioitu yhteensä kolmessa artikkelissa. Näiden tutkimusten mukaan tulokset olivat melko vaatimattomia.

Näyttää siltä, että vaikeassa riskitilanteessa olevan nuoren ongelmiin ei voida vaikuttaa kovin helposti. Ongelmallisemmissä tilanteissa kuin lasten ja nuorten käyttäytymisongelmilla intervensiolla käytetään monimuotoisempia hoitoja. Näillä voidaan saada kohtalaisen hyviä tuloksia. Tuloksia tarkasteltaessa on syytä huomioida, että suurin osa tutkimusten kontrolliryhmistä sai saatavilla olevia muita palveluita tai jotain muuta hoitoa.

Katsauksen neljässä artikkelissa oli arvioitu perheinterventioita vaikuttavuutta koko perheen riskitilanteeseen (Taulukko 2). Kontrolliryhmät saivat tavallisia saatavilla olevia palveluita yhtä tutkimusta lukuun ottamatta. Kolmessa tutkimuksessa oli käytetty kevyehköjä interventioita ja yhdessä syvällisempää; hoitoa, kuntoutusta ja terapiaa. Monimuotoista interventiota käytettiin vain yhdessä tutkimuksessa.

Kaikki interventiot olivat kirjoittajien mukaan vähintään kohtalaisen vaikuttavia. Sekä kevyehköillä että syvällisemmällä interventiolla saatiin aikaan suhteellisen hyviä tuloksia. Näiden tulosten pohjalta näyttäisi siltä, että perheinterventioiden avulla voidaan saada hyviä tuloksia aikaan riskitilanteessa olevien perheiden auttamisessa. Kevyehköt interventiot riittävät varhaista tukea tarvitseville perheille.

Perheen köyhyyteen liittyviä tutkimuksia löytyi yhteensä neljä kappaletta (Taulukko 2). Kontrolliryhmät saivat tavanomaisia saatavilla olevia palveluita yhtä tutkimusta lukuun ottamatta. Vähävaraisten perheiden interventiot olivat monimuotoisia perheinterventioiden yhdistelmiä. Neljästä interventioryhmästä kolme sai aikaan hyviä tuloksia. Vaikka tutkimuksia on vähän, näyttäisi alustavasti siltä, että monimuotoisilla perheinterventioilla on positiivisia vaikutuksia vähävaraisten perheiden ongelmien vähentämisessä.

Katsauksen 30 tutkimuksen avulla voitaneen sanoa, että perheinterventioilla on positiivisia vaikutuksia. Näyttäisi siltä, että perheinterventioiden asiakkaat ovat usein tilanteessa, jossa hoitona ei käytetä vain yhtä kevyehköä interventiota (koulutus ja tiedottaminen, ohjaus ja neuvonta, tilan-

Taulukko 2. Osallistujamäärä, interventioiden määrä ja kesto, sekä seuranta

Kohderyhmä	Osallistujamäärä	Interventioiden määrä	Interventioiden kesto	Seuranta
Lapsen tai nuoren käyttäytymisongelmat N = 15	Tutkimuksissa yhteensä 1379 osallistujaa	Keskimäärin 13 kertaa (vaihteluväli 5-45 krt)	Eniten 7-12 kk kestäneitä tutkimuksia (12 kpl), keston keskiarvo 12 viikkoa (vaihteluväli 6-25 vko)	Yhdeksässä tutkimuksessa arvioitiin intervention vaikuttavuutta vähintään neljän kuukauden seuranta-ajan jälkeen
Nuoren vaikea riskitilanne N = 7	Tutkimuksissa yhteensä 1197 osallistujaa	Ei mainintaa	4 kuukautta - 4 vuotta	Neljässä tutkimuksessa oli raportoitu seuranta-ajaksi 6-20 kuukautta
Perheen riskitilanne N = 4	Tutkimuksissa yhteensä 1180 osallistujaa	Ei mainintaa	1-2 vuotta	1-2 vuotta
Perheen köyhyys N = 4	Tutkimuksissa yhteensä 8902 osallistujaa	Ei mainintaa	1-18 vuotta	Ei mainintaa

nearvio ja tiedonkeruu, ryhmätyö ja vertaistuki) vaan useimmiten palveluna tarjotaan joko edellisten yhdistelmiä tai syvällisempään vuorovaikutukseen perustuvaa hoitoa, kuntoutusta tai terapiaa. Perheiden riskitilanteiden ratkaisemisessa käytettiin paljon yksittäisiä kevyehköjä interventioita ja tulokset olivat suhteellisen hyviä. Tämä johtui varmasti siitä, että tilanteet eivät olleet niin vakavia ja monimutkaisia kuin muissa kohderyhmissä.

Lasten ja nuorten käyttäytymisongelmien hoidossa perheinterventioina näyttää toimivan monimuotoinen interventioita yhdistävä hoito tai pelkkä hoito, kuntoutus tai terapia. Vaikeassa riskitilanteessa olevien nuorten tilannetta pyrittiin hoitamaan useimmiten syvällisempää vuorovaikutusta edellyttävillä interventioilla. Niillä saatiin kohtalaisia tuloksia. Vähävaraisten perheiden ongelmia hoidettiin monimuotoisilla interventioilla ja tulokset olivat näissä suhteellisen hyviä. Mitä vaikeampi tilanne, sen monipuolisempia ja syvällisempiä interventioita tarvitaan näyttäisi pitävän paikkansa tämän tutkimusaineiston pohjalta.

Heterogeenisten interventioiden hajanainen vaikuttavuus

Perheinterventioilla pyritään vaikuttamaan perheiden olosuhteisiin ja estämään ongelmien

kärjistyminen huonompaan suuntaan. Vertailua tehtäessä on otettava huomioon, että tutkimuksiin osanottajat tulivat erittäin köyhiltä kaupunkialueilta, jollaisia Suomessa ei ole lainkaan. Perheinterventioiden vaikuttavuutta oli arvioitu useasta näkökulmasta. Tulosuuttujina oli useita päämuuttujia ja alamuuttujia. Tutkimusten tulostulosten määrä ja heterogeenisuus oli vielä suurempi. Mittareiden oletettiin lähtökohtaisesti olevan valideja, koska tulokset oli julkaistu tieteellisessä lehdessä. Koska tutkimusten tulostulostenä käytettiin Yhdysvalloissa validoituja mittareita ja niiden sisältöjä kuvattu tarkasti, vaikeutui tutkimusten analysoiminen ja luotettavuuden arvioiminen. Tulostulostenä angloamerikkalaisuus vähentää tutkimusten sovellettavuutta Suomen kontekstiin.

Tutkimusten, interventioiden ja kohderyhmien heterogeenisuudesta johtuen interventioiden tuloksia ja vaikuttavuutta ei voitu analysoida yhdenmukaisesti koko aineiston osalta. Tämän vuoksi tutkimukset luokiteltiin kohderyhmittäin analysoitavaksi. Vaikka jaottelu selkiytti laajaa kokonaisuutta, ei katsauksen perusteella voida antaa lopullista vastausta interventioiden vaikuttavuudesta. Syynä ovat suuret eroavuudet kohderyhmien ongelmassa, käytetyissä interventioissa sekä tulosuuttujissa ja -mittareissa.

Tulokset kuitenkin selvästi viittaavat siihen, että perheinterventiot toimivat lasten ja nuorten

käyttäytymisongelmien hoidossa. Nuorten monimutkaiseen tai vaikeaan riskitilanteeseen sovellettiin useimmiten hoitoa, kuntoutusta tai terapiaa, jolloin tulokset olivat kohtalaisia. Monimuotoisilla perheinterventioilla oli positiivisia vaikutuksia vähävaraisten perheiden ongelmien vähentämisessä. Tulosten pohjalta näyttää siltä, että kevyehköjen perheinterventioiden avulla voidaan saada hyviä tuloksia riskitilanteessa olevien perheiden auttamisessa. Mutta lasten ja nuorten käyttäytymisongelmien ratkaisemisessa tarvitaan syvällisempään vuorovaikutukseen perustuvaa (hoito, kuntoutus ja terapia) tai monimuotoista interventiota.

Tarkasteltaessa koko aineistosta (30 tutkimusta) saatuja tuloksia, ei voida selkeästi päätellä, riittääkö pelkkä kevyehkö interventio. Tutkimukset viittaavat siihen, että tarvitaan monimuotoisia tai syvälliseen vuorovaikutukseen perustuvia interventioita. Yksittäisillä kevyehköillä interventioilla ei voida vaikuttaa pitkäaikaisiin ja syviin ongelmiin. Kevyehkö interventio toimii tulosten perusteella kuitenkin varhaisessa tukemisessa ja monimuotoisempien ongelmien ennaltaehkäisemisessä.

EHKÄISEVIEN KOTIKÄYNTIEN VAIKUTUKSET IÄKKÄIDEN TOIMINTAKYKYYN

Tavoitteet

Toljamon ym. (2005) katsauksen tavoitteena oli kerätä ja arvioida saatavilla oleva tieteellinen näyttö ehkäisevien kotikäyntien vaikuttavuudesta 65-vuotiaiden ja sitä vanhempien ihmisten toimintakykyyn, pitkäaikaiseen tai pysyvään laitoshoitoon joutumiseen, kuolleisuuteen ja hoidon kustannuksiin. Lisäksi tavoitteena oli arvioida näytön sovellettavuutta suomalaiseseen sosiaali- ja terveyspalvelujärjestelmään.

Ehkäisevällä kotikäynnillä tarkoitetaan katsauksessa sosiaali- tai terveydenhuollon henkilöstön tekemää käyntiä ikääntyneen kotiin. Käynnin tarkoituksena on arvioida ikääntyneen fyysistä, psyykkistä, kognitiivista ja sosiaalista toimintakykyä. Arvioinnin pohjalta etsitään yhdessä ikääntyneen kanssa keinoja, joilla hänen toimintakykynsä, terveyttään ja hyvinvointiaan voidaan pitää yllä ja parantaa. Käynnin yhteydessä arvioidaan myös asumisolosuhteita ja tiedotetaan olemassa olevista palveluista, joilla ikääntyneen

hyvinvointiin ja terveyteen voidaan vaikuttaa.

Katsauksen kohderyhmänä olivat kotona asuvat 65-vuotiaat ja sitä vanhemmat. Tämä ikäraja valittiin, koska sitä käytetään useimmissa kansainvälisissä ikääntymistä koskevissa tutkimuksissa. Alhaisen ikärajan ja siitä aiheutuneen suuren ikähajonnan vuoksi tuloksia tarkasteltiin myös eri ikäryhmissä (nuoret ja vanhat vanhukset).

Tiedonhaku

Aineisto kerättiin Medline-tietokannasta ja Cochrane-kirjastosta sekä kysymällä kuudelta suomalaiselta vanhustyön asiantuntijalta ja tutkijalta heidän tiedossaan olevia julkaisuja. Ensimmäinen kirjallisuushaku tehtiin Stakesin ehkäisevien kotikäyntien suunnittelua, sisältöä ja arviointikäytäntöjä selvittäneen hankkeen yhteydessä. Medline-tietokannasta etsittiin tuolloin vapaasanahauulla vuodesta 1966 alkaen ne artikkelit, joissa 'preventive home visit' oli tekstisana ja jotka oli luokiteltu meta-analyseiksi. Meta-analyysillä tarkoitetaan menetelmää, jossa kvantitatiivisesti yhdistellään järjestelmällisen kirjallisuuskatsauksen menetelmin tavoitettujen alkuperäistutkimusten tuloksia päätelmien tekoa varten. Haun tuloksista rajattiin mukaan ne, joissa kohderyhmänä olivat yli 65-vuotiaat. Haussa löytyi kaksi verrattain tuoretta systemaattista katsausta tekijöinä Elkan ym. (2001) ja Stuck ym. (2002). Näiden arvioimme kattavan luotettavasti ehkäisevistä kotikäynneistä ja niiden vaikuttavuudesta vuosina 1966-2001 tehdyt kansainväliset tutkimukset. Lisäksi niissä oli arvioitu alkuperäistutkimusten menetelmällinen laatu. Näiden kahden systemaattisen katsauksen alkuperäistutkimukset olivat katsauksen aineistona. Alkuperäistutkimuksia kartoittavia uusia tietokantahakuja ei tehty hankkeelle asetetun aikataulun takia.

Elkanin ym. (2001) katsauksen päätulosmuuttujina olivat kuolleisuus, sairaalaan joutuminen, pitkäaikaiseen laitoshoitoon joutuminen, toimintakyky ja terveydentila. Tutkijat etsivät aiempia tutkimuksia sähköisistä tietokannoista (Medline 1966-1997, Cinahl 1982-1997, Embase 1980-1997, Cochrane-kirjasto), Internetistä, käsihaulla (Health Visitor 1982-1997), katsausartikkeleiden lähdeviiteistä, julkaisemattomista lähteistä, lehtimainoksista ja asiantuntija-kontaktien kautta. Vain englanninkieliset julkaisut otet-

tiin mukaan. Haun tuloksena tutkijat saivat 1215 tiivistelmää tai viitettä, joista laatukriteerit täytti 15 satunnaistettua vertailututkimusta.

Stuckin ym. (2002) systemaattisessa katsauksessa tutkittiin ehkäisevien kotikäyntien vaikuttavuutta toimintakykyyn, vanhainkotiin joutumiseen ja kuolleisuuteen. Tutkijat käyttivät tiedonhaussa sähköisiä tietokantoja (Medline 1985-2001, Psycinfo 1985-2001, Embase 1985-2001, Cochrane Cont-rolled Trials Register), katsausartikkelien ja kirjojen lähdeviitteitä, konferenssijulkaisuja ja erikoislehtiä sekä asiantuntijakontakteja. Kielirajauksina olivat englanti, ranska, saksa, italia ja espanja. Haulla löytyi 1349 abstraktia, joista kriteerit täytti 17 satunnaistettua vertailututkimusta.

Nyt tarkasteltavassa katsauksessa etsittiin Cochrane-kirjastosta tuoreita systemaattisia katsauksia hakusanoilla 'preventive home visit' OR 'in-home visit' AND 'elder-ly'. Haun tuloksena ei löytynyt edellä mainittujen kahden lisäksi muita systemaattisia kirjallisuuskatsauksia ikääntyneiden luo tehdyistä ehkäisevistä kotikäynneistä. Hakutermein 'home visit' AND 'prevention' AND 'elderly' löytyi vuonna 2000 tehty systemaattinen katsaus (van Haastregt ym. 2000). Siinä ei ollut tehty meta-analyysia kuten tämän tarkastelun päälähteissä. Haastregtin ym. katsaus sisälsi samat alkuperäislähteet kuin tämän tarkastelun päälähteet

Katsausta laadittaessa otettiin yhteyttä tutkija Ruth Elkanin (Nottingham, UK) sähköpostitse ja tietoja kirjallisuushausta ja alaryhmäanalyysista tarkennettiin. Vastaus tiedusteluun saatiin tammikuussa 2004 ja toinen viesti elokuussa 2004. Anders Stuckiin ei otettu yhteyttä.

Suomalaisten valmiiden tai meneillään olevien vertailututkimusten kartoittamiseksi otettiin yhteyttä kuuteen alan asiantuntijaan tai tutkijaan sähköpostitse ja näiltä pyydettiin tietoja ehkäiseviin kotikäynteihin liittyvistä satunnaistetuista vertailututkimuksista. Vastausten perusteella Suomessa ei ole tehty eikä juuri nyt ole meneillään satunnaistettuja vertailututkimuksia ehkäisevien kotikäyntien vaikuttavuudesta.

Alkuperäisartikkeleista etsittiin tiedot potilaista, interventiosta ja keskeisistä tuloksista ja ne taulukoitiin mahdollisimman vertailukelpoisessa muodossa. Keskeisten tulosten (toimintakyky, laitoshoitoon joutuminen ja kuolleisuus) osalta laskettiin uudelleen vaikutuksen suuruus julkaisuissa ilmoitettujen potilasmäärien avulla.

Aineisto

Tämän katsauksen aineistona olivat Elkanin ym. (2001) ja Stuckin ym. (2002) katsausten satunnaistetusta vertailututkimuksista ne, joissa kotikäynnin toteutti terveydenhuoltoalan ammattilainen. Tarkasteluun jäi lopulta 21 kriteerit täyttävää tutkimusta².

Tarkasteltavat satunnaistetut vertailututkimukset oli julkaistu vuosina 1982-2001. Valtaosa koski 75-vuotiaita tai sitä vanhempia. Eniten tutkimuksia oli tehty Iso-Britanniassa, Yhdysvalloissa ja Tanskassa. Toimintaympäristönä oli useimmin kaupunki. Yhdessä julkaisussa mainittiin, että tutkimuksen toinen osa oli tehty maaseudulla. Kahdessa tanskalaistutkimuksessa kohdealueena oli koko kunta. Intervention toteutti yleisimmin sairaanhoitaja tai terveydenhoitaja, mutta mukana saattoi lisäksi olla toinen ammattilainen. Seuranta-ajat vaihtelivat viidestä kuukaudesta viiteen vuoteen.

Tässä tarkasteltavassa katsauksessa käytetään käsitettä 'ehkäisevä kotikäynti', jonka käyttö on viime vuosina lisääntynyt ja vakiintunut sekä kansallisessa että kansainvälisessä tutkimuksessa. Ehkäisevällä kotikäynnillä tarkoitetaan tässä sosiaali- tai terveydenhuollon ammattilaisen tekemää kotikäyntiä, jonka tarkoituksena on arvioida vanhuksen toimintakykyä ja etsiä keinoja sen ylläpitoon. Interventiona pidetään ehkäisevää kotikäyntiä tai sellaisten sarjaa sekä käyntien perusteella esitetyjä toimenpiteitä, joilla ratkaistaan esiin tulleita ongelmia tai vähennetään itseenäisen kotona selviämisen esteitä.

Tutkimuksissa tehtyjen kotikäyntien määrä vaihteli yhdestä kahteentoista. Myös sisällöltään kotikäynnit vaihtelivat kovasti. Useimmiten toimenpiteet olivat ongelmatilanteiden tunnistamista, ohjausta ja neuvontaa sekä tarvittaessa palveluihin ohjaamista. Muutamissa tutkimuksissa kotikäynteihin sisältyi erilaisia mittauksia ja testauksia. Kuudessa tutkimuksessa tehtiin kokonaisvaltainen geriatrinen arviointi seurantaikänteineen. Kokonaisvaltainen geriatrinen arviointi sisältää lääketieteellisen, psykososiaalisen ja toimintakyvyn kokonaisarvion seurantaikänteineen. Kolmessa tutkimuksessa sairaanhoitajan käynti tai käyntien sarja käsitti terveydentilaan liittyvien tekijöiden tarkistamista tai seulontaa (screening, health check). Näissä ohjelma oli kaksivaiheinen siten, että väestöstä tunnistettiin ensin riskiryhmä esimerkiksi postikyselyn avulla,

ja riskiryhmän luo tehtiin sitten ehkäiseviä kotikäyntejä.

Kotikäynnit ryhmiteltiin kohdennettuihin ja kohdentamattomiin. Tietyille riskiryhmälle kohdennettuja kotikäyntejä tehtiin kymmenessä tutkimuksessa. Kotikäynnit oli kohdennettu äskettäin sairaalasta kotiutuneille, kaatumisalttiille, heikkokuntoisille, veteraaneille tai ikääntyneille, joilla oli tunnistettuja erityisiä ongelmia. Kohdentamattomia kotikäyntejä oli tehty 11 tutkimuksessa. Kohdentamaton kotikäynti tarkoitti yleensä ikääntyvän normaaliikäisestä luo tehtyä ennaltaehkäisevää käyntiä.

Tutkimusten ryhmäkoot vaihtelivat siten, että kolmen tutkimuksen interventio- ja vertailuryhmässä oli kummassakin alle sata henkeä, kun laajimman tanskalaistutkimuksen ryhmäkoko oli yli 2000. Seitsemässä tutkimuksessa ryhmäkoko oli 100-200 henkeä ja kymmenessä 201-400 henkeä.

Ehkäisevien kotikäyntien vaikuttavuutta arviointiin useimmin kuolleisuuden vähenemisen ja vanhainkotiin tai pitkäaikaiseen laitoshoitoon joutumisen avulla. Yleisesti mitattiin myös muutosta fyysisessä toimintakyvyssä. Joissakin tutkimuksissa oli käytetty useampaa kuin yhtä päämuuttujaa. Kolmessatoista tutkimuksessa oli käytetty tunnettuja, validoituja mittareita, kuten Barthelin ADL -mittaria tai Beckin depressioasteikkoa. Tässä tarkasteltavassa katsauksessa keskeiset tulostulokset ovat toimintakyky, pitkäaikaiseen tai pysyvään laitoshoitoon joutuminen, kuolleisuus ja kustannukset.

Elkan ym. (2001) arvioivat tutkimusten laatua Reischin asteikolla, joka kattaa tutkimuksen tarkoituksen, asetelman, otoskoon, menetelmäkuvauksen ja sen sopivuuden tutkimukseen, osallistujien kohtelun sekä hoitojärjestelyjen arvioinnin. Pistemäärä vaihtelee nolasta yhteen siten, että korkeat pisteet tarkoittavat tutkimusten parempaa suhteellista laatua. Selvää kynnyksarvoa heikkojen ja hyvien tutkimusten välille ei voida asettaa. Elkanin ym. katsauksesta otettujen tutkimusten laatuasteet vaihtelivat 0.21 - 0.65 välillä. Kuusi tutkimusta sai 0.5 pistettä tai enemmän. Alle 0.5 laatuasteetta sai kuusi tutkimusta.

Stuck ym. (2002) arvioivat tutkimusten laatua tarkastelemalla satunnaistamistapaa, osallistujien määrää sekä tutkimus- ja hoitohenkilökunnan sokkouttamista tavoiteltavien tulosten suhteen. Tarkemmin arvioitavat kohteet olivat keski-ikä

ja kuolleisuus vertailuryhmässä, intervention kesto, kotikäyntien määrä ja geriatrisen arvioinnin kokonaisvaltaisuus. Samoin arvioitiin tutkimuspaikan maantieteellinen sijainti ja tutkimusryhmän kokoonpano. Arvioinnin mukaan viidessä tutkimuksessa oli kuvattu asetelman satunnaistaminen ja seitsemässä myös sokkouttaminen. Stuckin ym. (2001) tuloksiin eivät vaikuttaneet maantieteellinen sijainti tai tutkimusryhmä. Kuuden tutkimuksen laatua nosti niissä tehty kokonaisvaltainen geriatrisen arviointi seurantaikänteineen.

Katsausten tekijöiden arvion perusteella laadutaan parhaimmaksi ja siten mahdollisesti luotetavampaa näyttöä antaviksi nousi 12 tutkimusta. Muiden tutkimusten antamaa näyttöä tulee tämän perusteella harkita erityisen kriittisesti. Käytettyjen kriteerien perusteella laadullisesti heikommat tutkimukset ovat julkaisuvuodeltaan vanhempia.

Tulokset

Ehkäisevät kotikäynnit sisälsivät hyvin erilaisia interventioita ja interventioon kuuluneiden kotikäyntien määrä vaihteli yhdestä yli kymmeneen. Niiden sisältökin vaihteli yhdestä tai muutamasta toimenpiteestä (esim. kehoitus ottaa rokotuksia) monipuoliseen, usean ammattiryhmän toteuttamaan geriatriseen arviointiin. Niinpä myös tuloksia mitattiin hyvin eri tavoin. Kaikissa tutkimuksissa ei ollut käytetty validoituja mittareita, vaan muotoiltu omia, tutkimuskohtaisia mittauskeinoja. Kaikissa tutkimuksissa ei ollut asetettu yhtä selkeää päätulosta. Aina ei myöskään esitetty lähtömittauksien tuloksia, vaan pelkästään muutoksen suuruus.

Kun tutkimukseen osallistuneiden joukko oli varsin heterogeeninen ja niin päämuuttujat kuin niiden mittaustavat vaihtelevat, ei tulosten yhdistäminen laskennallisesti (meta-analyysi) ollut perusteltua. Ongelmaa korostaa se, että eri terveydenhuoltojärjestelmissä iäkkäiden, kotona asuvien ihmisten saaman hoidon ja hoivan perustaso voi eri ympäristöissä vaihdella paljon. Tätä taustavaihtelua ei tutkimuksista yleensä ole mahdollista saada esiin.

Keskeisten tulosten (toimintakyky, laitoshoitoon joutuminen ja kuolleisuus) osalta laskettiin uudelleen vaikutuksen suuruus julkaisuissa ilmoitettujen potilasmäärien avulla. Tarkasteltavaan katsaukseen hyväksytyjen satunnaistettujen tut-

kimusten tulokset kiteytetään seuraavassa.

Ehkäisevillä kotikäynneillä on mahdollista tukea etenkin IADL-mittarilla arvioitua sosiaalisen omaisuuden ylläpitämistä, jos ikääntyneen toimintakyky ei ole lähtötilanteessa huono. Monissa tutkimuksissa johtopäätöksenä esitetty vaikutus fyysisen toimintakyvyn ylläpitämiseen tai parantamiseen sen sijaan ei saa tukea varsinaisista mittauksista. Ehkäisevien kotikäyntien vaikutusten arviointia toimintakykyyn heikentää erityisesti se, että tutkimuksissa oli harvoin mitattu lähtötilanteen toimintakykyä. Niinpä toimintakyvyssä tapahtuvien tavoiteltavien muutosten arviointi jää niukaksi. Kaikki kuusi tutkimusta, joissa vaikutus oli positiivinen, oli Elkanin ym. systemaattisessa katsauksessa arvioitu laadultaan hyväksi.

Myönteisiä vaikutuksia toimintakykyyn havaitaan tavallisimmin, kun kotikäynnit ja niihin liittyvät interventiot on suunnattu riskiryhmille (ei kuitenkaan sairaalasta kotiutuville), kun tehdään kokonaisvaltainen geriatrinen arviointi ja kun käyntejä on enemmän kuin yksi. Tutkimuksen seuranta-ajalla ei näytä olevan vaikutusta tulokseen.

Tulosten perusteella ehkäisevät kotikäynnit voivat vähentää tai siirtää myöhempään pitkäaikaiseen tai pysyvään laitoshiitoon joutumista. Kolmessa tutkimuksessa koeryhmä joutui laitoshiitoon tai vanhainkotiin tilastollisesti merkittävästi harvemmin kuin vertailuryhmä; absoluuttinen ero vaihteli 7.7 %:sta 12.4 %:iin. Yhdeksässä tutkimuksessa ryhmien välillä ei todettu tilastollisesti merkittäviä eroja; tulokset vaihtelivat intervention paremmuudesta (3.1 % tai vähemmän) verrokkien pienempään laitostumiseen (4.4 % tai vähemmän). Todellinen vaikuttavuus lienee siten melko vaatimaton, alle kymmenen prosentin tasolla. Kohderyhmän ikä, käyntien tiheys, intervention sisältö tai seuranta-ajan pituus eivät näyttäisi liittyvän vaikuttavuuteen.

Tulosten mukaan ehkäisevät kotikäynnit saattavat jonkin verran vähentää kuolleisuutta. Kotikäyntiryhmän kuolleisuus oli kolmessa tutkimuksessa matalampi kuin verrokkien (absoluuttinen ero 4.5-6.5 %). Seuranta-aika niissä oli kahdesta viiteen vuoteen. Yhdessä tutkimuksessa kuolleisuusero ilmeni kaupunkilaisvanhuksilla, vaikka maalla asuvaan ryhmään kotikäynneillä ei ollut vaikutusta.

Muissa 16 tutkimuksessa, joissa kuolleisuus oli tiedossa, ei havaittu tilastollisesti merkittäviä

eroja. Valtaosa näistä tutkimuksista oli metodisesti hyvinlaatuisia, mutta kaikissa kuolleisuus ei ollut päämuuttujana vaan sivuhavaintona. Absoluuttinen ero vaihteli koeryhmän 9.8 prosentin edusta verrokkien 9.0 prosentin etuun. Tutkimukset, joissa kuolleisuuden havaittiin vähenevän, oli tehty hieman nuoremmille iäkkäille kuin ne, joissa vaikutusta ei havaittu. Interventioiden sisällöllä, kohderyhmällä tai seuranta-ajan pituudella ei näyttänyt olevan yhteyttä kuolleisuuteen.

Tarkastelluista alkuperäistutkimuksista viidessä oli tarkasteltu kotikäyntien kustannusvaikutuksia. Näistä neljässä tutkimuksessa kustannukset per henkilö olivat interventioyryhmässä korkeammat kuin kontrolliryhmässä 1-3 vuoden seuranta-aikana. Näyttäisi siltä, että tällä aikavälillä ehkäisevät kotikäynnit interventioineen eivät johtaisi kustannussäästöihin, vaan pikemminkin lisäkustannuksiin.

Jossain määrin näyttöä kotikäyntien vaikuttavuudesta

Katsauksen aineistona olivat Elkanin ym. (2001) ja Stuckin ym. (2002) systemaattisten katsausten 21 alkuperäistutkimusta, jotka analysoitiin uudelleen. Katsausten kysymyksenasettelu sekä mukaan otettavien satunnaistettujen vertailututkimuksen sisäänottokriteerit olivat yhteneviä tässä tarkasteltavan katsauksen tavoitteen kanssa. Katsausten kirjallisuushaku oli tehty varsin kattavasti ja alkuperäistutkimusten menettämällinen laatu oli arvioitu yleisesti hyväksytyjä periaatteita noudattaen. Aineiston vahvuutena voidaan pitää kattavuutta ja alkuperäistutkimuksissa käytettyä hakustrategiaa. Käytössä oli todennäköisesti ehkäisevien kotikäyntien vaikuttavuudesta kaikki julkaistut satunnaistetut vertailututkimukset.

Aineiston suurimpana heikkoutena voidaan pitää alkuperäistutkimusten hajautumista lähes kahdenkymmenen vuoden jaksolle 1982-2001, sillä sekä ikääntyneiden toimintakyky että terveydenhuollon toiminta ovat muuttuneet tänä aikana länsimaissa huomattavasti. Aineiston laatua heikentää validoimattomien mittareiden suuri osuus ja se, että mittareiden asteikkoa ei ilmoitettu. Puutteellisin mittarein saatu tulos voi olla tilastollisesti merkittävä, vaikkakaan ei välttämättä kliinisesti tärkeä. Sovellettavuuden kannalta ongelmiana on myös se, että alkuperäistutkimuk-



set oli tehty muualla kuin Suomessa eikä edes Pohjoismaissa. Eri maiden vanhustenhuolto kotikäynnteinen voi olla hyvinkin erilaista. Tarkastelua vaikeutti se, että kotikäyntien yksityiskohtaista toteuttamista oli alkuperäistutkimuksissa usein kuvattu hyvin niukasti jos lainkaan. Monista puuttui myös vertailuintervention kuvaus. Se oli usein ohitettu maininnalla 'tavanomainen hoito', jonka informaatioarvo on vähäinen.

Suomalaisessa vanhustenhuollossa vallitsee asiakaslähtöisen räätälöidyn palvelun paradigma. Tavoitteena on palvella yksilöllisesti kutakin asiakasta tämän tarpeiden pohjalta. Siten myös 'ehkäisevät kotikäynnit' voivat sisältää erilaisia asiakkaiden kotona annettavia palveluja tai palvelukokonaisuuksia saman nimikkeen alla. Meillä strukturoitujen tai puolistrukturoitujen interventioiden perinne on varsin lyhyt. Jotta interventioiden vaikuttavuutta voidaan kokeellisesti mitata, on ne kuvattava työmenetelmän selkeästi ja eroteltava muista samanaikaisista toimista.

Alkuperäistutkimusten toteuttamistavat vaihtelivat hyvin paljon mm. kotikäynnin pituuden, käyntimäärien ja kohderyhmien suhteen. Tutkimuksessa saattoi olla mukana sekä riskiryhmiin kuuluvia että valikoimaton joukko ikääntyviä ihmisiä. Tutkimuksissa ei myöskään näyttänyt olevan yhtenäisiä näkemyksiä siitä, mitä ehkäisevällä kotikäynnillä tarkoitetaan. Tämä heikentää näytön sovellettavuutta suomalaisen palvelujärjestelmään.

Tieteellinen tutkimusnäyttö ehkäisevien kotikäyntien vaikutuksista ikääntyneiden toimintakykyyn, pitkäaikaiseen tai pysyvään laitoshoitoon joutumiseen, kuolleisuuteen ja kustannuksiin jäi tässä tarkastelussa heikoksi tehtyjen ulkomaisen tutkimusten perusteella. Tutkimukset antavat jonkin verran näyttöä siitä, että kohdennetut ehkäisevät kotikäynnit voivat vähentää kuolleisuutta ja vanhainkotiin joutumista eräissä ikääntyvien ryhmissä, lähinnä ns. nuorten ikääntyvien eli alle 75-vuotiaiden joukossa.

Tutkimusten mukaan ei ole näyttöä siitä, että ehkäisevät kotikäynnit vähentävät ikääntyneiden toimintakykyyn heikkenemistä yleisesti. Kuitenkin kotikäyntien vaikutukset toimintakykyyn voivat olla suotuisia niillä ikääntyneillä, joilla on alhainen kuolleisuusriski ja niissä kotikäyntiohjelmissa, joissa tehdään kokonaisvaltainen geriatrinen arviointi ja seuranta.

Tässäkin katsauksessa tulosten tulkintaa vaikeutti se, että analysoiduissa ulkomaisissa

tutkimuksissa intervention, tässä ehkäisevän kotikäynnin toimintaympäristö ja palvelujärjestelmä, itse kotikäynti interventiona, intervention tekijä ja hänen koulutuksensa, seuranta-aika ja kohderyhmä vaihtelivat varsin paljon.

KESKUSTELU

Olemme kuvanneet tässä artikkelissa rehellisesti omat systemaattisten tutkimuskatsausten tekoon liittyvät kokemuksemme, kohtaamamme teoreettiset ja vertailuun liittyvät ongelmat ja ratkaisumme, jotta seuraavia katsauksia tekevät voivat välttää samat ongelmat. Kokemuksemme opettivat meille nöyryyttä sekä johtopäätösten teossa, tulosten siirtämisen kysymyksissä että itse katsausten laadintaprosessissa.

Pulmia riitti. Tutkimusten etsiminen tietokannoista osoittautui varsin ongelmalliseksi. Pulmia syntyi tietokantojen ominaisuuksien, hakusanojen ja hakufraasin vuoksi. Esimerkiksi perheinterventioiden vaikuttavuutta koskevan katsauksen edetessä huomattiin, että perhetyö ei käsitteenä toiminut kansainvälisissä tietokannoissa ja tutkimusten asiasanoituksissa. Koska kohteeksi valittiin sosiaalialan perhetyön interventiot, pulmaksi tuli sosiaalialan erityisten perheinterventioiden löytäminen, kun joukkoon tuli myös alaltaan epäselviä tutkimuksia. Myös perheinterventioiden rajapinnat näyttivät epäselviltä, joten kaikki löydetty artikkelit pidettiin mukana. Tiedonhaun käsitteellisistä ongelmista johtuen lähdimme siitä, että katsauksen aineisto edustaa vain osaa tehdystä perheinterventioiden vaikuttavuustutkimuksista.

Katsausten tuottamisprosessi opetti sen, että tiedonhakuprosessi on tarpeellista aloittaa ensin kartoittavalla haulla, jonka tulosten pohjalta ratkaistaan lopulliset hakusanat. Tutkimukset on syytä rajata selkeään erilliseen tarkkaan mieluiten standardoituun interventioon tai kohderyhmään. Perheinterventio osoittautui liian yleiseksi käsitteeksi kansainvälisissä tietokannoissa. Kun satunnaistetut tutkimusasetelmat ovat tutkimusten valintakriteereinä, katsauksiin tulee pääasiassa angloamerikkalaisia tutkimuksia. Tämä vähentää interventioiden ja tulosten sovellettavuutta meillä.

Tulosmittareiden vaihtelu, moninaisuus ja niiden validius on lisähaaste. Vertailtavuuden ja tiedon siirrettävyyden ongelmana on sekä kulttuurien että palvelujärjestelmien sekä työtapojen



ja niiden nimikkeiden vaihtelu kansainvälisesti.

Aloimme myös pohtia kirjallisuuskatsausten kustannusten ja saadun tuotoksen välistä suhdetta.

Edellä eritellyistä pulmista huolimatta on syytä kehittää menetelmiä, joiden avulla sosiaalialan interventioiden ja toimenpiteiden vaikuttavuustietoa voidaan koota ja systematisoida. Jos sitä tehdään vaikuttavuusnäytön pohjalta, se edellyttää työmenetelmien ja interventioiden kuvaamista, jopa standardointia, jotta voidaan vakuuttua siitä, että verrataan samoja interventioita.

Jatkossa on syytä pohtia, onko sosiaalialan tutkimuskatsauksissa kyse siitä, että 1) on olemassa jonkinlainen objektiivinen vaikuttavuustiedon hierarkia, jossa luotettavinta ja objektiivisinta tietoa tuottaa satunnaistettu koeasetelma ja sen tuottaman tiedon meta-analyysi, kuten lääketieteessä ja terveydenhuollossa ajatellaan. Vai onko kyse enemmänkin siitä, että 2) tiivistetään ja systematisoidaan eri tutkimusasetelmin tuotettua tietoa sekä käyttäjien ja ammattilaisten tuottamaa tietoa, kuten hiljaista tietoa.

Tässä on otettava huomioon se, että tutkimusasetelmat voivat olla keskenään yhteen sopimattomia ontologisilta ja tietoteoreettisilta sitoumuksiltaan (ks. luku Vaikuttavuustutkimuksen päälinjat sosiaalialalla), mutta niiden anti voi olla eri näkökulmista teemojen avartaminen. Kun kutakin tiedon lajia tarkastellaan sen omista lähtökohdista ja laatukriteereistä käsin, voidaan tuottaa moninäkökulmaista tietoa. On alettukin keskustella siitä, miten eri tutkimusasetelmin tuotettua tietoa voitaisiin syntetisoida sekä siitä, miten yhteen soviitta määrällisen ja laadullisen tutkimuksen tuloksia, kuten EPPI-keskus ja SCIE jo tekevät (ks. Oliver ym. 2005).

Tehdyn kansainvälisen tutkimuskatsaustyön perusteella voidaan sanoa, että satunnaistetun koeasetelman perusteella tehtyä tutkimusta systematisoivat katsaukset tuottavat hyvin ylimalkaista tietoa toimenpiteiden vaikuttavuudesta. Kahden tarkastelemamme katsauksen tuottama tieto on todennäköisesti niin triviaalia, että sosiaalialan ammattilaiset tietävät ne muutenkin. Näyttää myös siltä, että mitä pidempi ja moninaisempi interventio on, sitä epävarmempaa tietoa satunnaistettu koeasetelma tuottaa. Satunnaistetuista kokeista tehtyä meta-analyysiäkin voidaan tarkastella kriittisesti. Mitä eri puolilla maapalloa tehtyjen ja todennäköisesti ainakin jossain määrin erilaisten, vaikkakin näennäisesti samojen inter-

ventioiden tilastollinen meta-analyysi tarkoittaa. Tullaanko verranneeksi eri asioita; omenoita ja päärynöitä?

Koska pohjoismaista puuttuu koe- ja kontrolliasetelmiin (randomized controlled trials) perustuva sosiaalialan interventioiden tutkimus, tulee satunnaistetun koeasetelman vaatimuksilla tehtyjen katsausten aineistoksi lähinnä angloamerikkalaista tutkimusta. Toimintaympäristö, palvelujärjestelmät, interventiot ja niiden tekijöiden koulutustaso voivat vaihdella kovinkin paljon eri maissa. Tämä tuottaa omat varauksensa tulosten siirrettävyyteen ja sovellettavuuteen. Tarkastelemiemme systemaattisten tutkimuskatsausten pohjalta tarvitaan vielä keskustelua systemoitujen tutkimuskatsausten hyödyistä, tulosten siirrettävyydestä ja varauksista erityisesti sosiaalialalla. Tutkimuskatsausten tekeminen on hyvin vaativaa ja työntensiivistä. Niiden tuottaminen edellyttää tutkijaryhmää, jossa voidaan ratkoa esiin tulleita ongelmia yhdessä.

Jotta vaikutusten syntymekanismeja voidaan ymmärtää, tarvitaan prosessin tutkimusta. Toimenpiteiden ja interventioiden kohteena olevat ihmiset eivät elä tyhjiössä, vaan vaikuttavia tekijöitä on monenlaisia ja niitä on kaiken aikaa ympärillä. Realistinen arviointi on askel suuntaan, jossa yhdistyvät prosessi- ja tulosarviointi. Pawson (2002) on kehittänyt tapaa, jolla tällaisia realistiseen arviointiin perustuvia tutkimuksia voidaan syntetisoida. Kuitenkin, realistinenkaan arviointi ei ylitä kahtiajakoa intervention menetelmä vs. konteksti (ks. Kuvio 1). Saattaa olla, että jäsennellympi kuva vaikuttavuudesta saadaan sellaisten lähestymistapojen avulla, jossa menetelmää ja kontekstia ei eroteta kaksijakoisesti toisistaan. Tällainen avaus on esimerkiksi relationaalinen arviointi³ (ks. Koivisto 2006).

Systemaattisia tutkimuskatsauksia voidaan tehdä vain sellaisista tutkimuksista ja tiedosta, mitä on olemassa. Toistaiseksi sosiaalialalla on kaikkiaan vähän vaikuttavuustutkimusta ja olematon määrä sellaisia tutkimuksia, joissa vaikuttavuutta tutkitaan prosessi- ja tulosarvioinnin yhdistävällä tavalla, tavalla jossa ei puhuta erikseen prosessi- ja tulosarvioinnista. Tällaisessa tilanteessa on tarpeellista valita moninäkökulmainen ote katsausten teossa ja koota ja systematisoida erilaisten lähestymistapojen, koeasetelmien, realistisen arvioinnin jne., tuottamaa tietoa. Kun systemaattiset tutkimuskatsaukset tiedon koostamisen ja tiivistämisen muotona ovat

yleistymässä, tarvitaan hyvää tutkimusten ja projektien raportointia. Hyvin toteutettu tutkimus, jonka raportointi on vaikutusten tarkastelun kannalta puutteellista, jää katsausten aineiston ulkopuolelle eikä hyödytä sosiaalialan tietoperustan kehittämistä. Vaikutusten arvioinnin kannalta tutkimuksen ja/tai projektin vaiheet ja valinnat tulisi raportoida mahdollisimman läpinäkyvästi ja eksplisiittisesti, jotta lukija voi tehdä päätelmiä siitä, mihin tutkimuksen vaikuttavuusnäyttö perustuu. Vaikuttavuustietoa eittämättä tarvitaan sosiaalialan tietopohjan kehittämiseksi, mutta on tärkeää harkita, mitä me tarkoitamme, kun puhumme luotettavasta tiedosta.

VIITTEET

¹ Tutkimusaineisto on listattu katsauksessa (Westman ym. 2005), listan saa tarvittaessa myös artikkeliin kirjoittajilta.

² Tutkimusaineisto on listattu katsauksessa (Toljamo ym. 2005), listan saa tarvittaessa myös artikkeliin kirjoittajilta.

³ Relationaalinen arviointi tuottaa tietoa siitä, millaisten toimijoiden, toiminnan ja resurssien avulla tavoiteltu muutos saavutetaan. Arviointisuuntauksesa tarkastellaan esimerkiksi asiakkaan elämäntilannetta, sosiomateriaalisina verkostoina. Verkostot koostuvat ihmisistä ja muista elementeistä, esimerkiksi päihdeasiakkaan kohdalla ”ryppyporukasta”, päihteistä, fyysisistä tiloista (metsä, asuntola), rahasta ja tavoitteista. Verkostot ovat inhimillisen toiminnan ja vuorovaikutuksen tuloksia. Lähestymistavalla tutkitaan, miten verkostot muuttuvat asiakkaan, sosiaalialan työntekijän ja muiden toimijoiden vuorovaikutuksen tuloksena. Arvioinnin tuloskriteerit määritellään tavoiteltavan muutoksen pohjalta. Muutoksen tutkiminen edellyttää asiakkaan verkoston seuraamista ennen interventiota, sen aikana ja jälkeen..

LÄHTEET

Boaz, Annette, Ashby, Deborah & Young, Ken: Systematic reviews: what have they got to offer evidence based policy and practice? ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London, Jan 2002, (Working Paper 2) 2002.

Campbell, Donald T. & Stanley, Julian C: Experimental and Quasi Experimental Designs for Research. Rand McNally, Chicago 1963.

Dahler-Larsen, Peter: Vaikuttavuuden arviointi. Hyvät käytännöt. menetelmä-käsikirja. Helsinki: Stakes, FinSoc 2005.

Elkan, Ruth, Kendrick, Denise, Dewey, Michael, Hewitt, Michael, Robinson, Jane, & Blair, Mitch: Effectiveness of home based support for older people: systematic review and meta-analysis. *BMJ* 323 (2001), 7315, s. 719-725.

Guba, Egon G. & Lincoln, Yvonne S.: Fourth Generation Evaluation. Sage, Newbury Park, London, New Delhi 1989.

Koivisto, Juha: Tiedon kerääminen ja systematisointi sosiaalialalla. Katsaus viiden ulkomaisen arviointiinstituutin strategioihin. Stakes, FinSoc arviointiraportteja 1/2005, Helsinki 2005.

Koivisto, Juha: Sosiaalialan näyttökeskustelu: miten sosiaalisten interventioiden vaikuttavuus osoitetaan? *Janus* 14(2006) 1, s. 53-60.

Latour, Bruno: One more turn after social turn. Teoksessa Ernan McMullin (toim.) *The Social Dimensions of Science*. University of Notre Dame Press, Indiana 1992, s. 272-294.

Leiman, Mikael: Vaikuttavuustutkimuksen pulmallisuus psykoterapiassa. Teoksessa Katja

Hartikainen & Jukka Aaltonen (toim.) *Näyttöön perustuva psykoterapia*. Jyväskylän yliopisto, Jyväskylä 2004.

Meklin, Pentti: Tavoitteiden saavuttamisen arviointi kuntataloudessa. Teoksessa Myllymäki A, Vakkuri J (toim.) *Tulos, normi, tilivelvollisuus*. Tampere: University press, Tampere 2001, s. 91-117.

Mäntysaari, Mikko: Arviointitutkimuksen taustaoletukset. Teoksessa Risto Eräsaari & Tuija Lindqvist & Mikko Mäntysaari & Marketta Rajavaara (toim.) *Arviointi ja asiantuntijuus*. Oy Yliopistokustannus University Press Finland Ltd, Tampere, 1999, s. 54-68.

Oliver, Sandy, Harden, Angela & Rees, Rebecca: An Emerging Framework for Including Different Types of Evidence in Systematic Reviews for Public Policy. *Evaluation* 11(2005), 4, s. 428-446.

Patton, Michael Quinn: Utilization Focused Evaluation. *The New Century Text*. Sage, Thousand Oaks, London and New Delhi 1997.

Pawson, Ray: Evidence Based Policy: I. In Search of Method. ESRC UK Centre for Evidence Based Policy and Practice, Queen Mary, University of London. ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 3. <http://www.evidencenetwork.org/Documents/wp3.pdf>. 2001.

Pawson, Ray: Evidence-based policy: the promise of 'realist synthesis'. *Evaluation* 8(2002). 3, s. 340-358.

Pawson, Ray, Boaz, Annette, Grayson, Lesley, Long, Andrew & Barnes, Colin: Types and quality of knowledge in social care. *Knowledge Review* 3. Social Care Institute for Excellence. <http://www.scie.org.uk/publications/list.asp#kr>. 2003.

Pawson, Ray & Tilley, Nick: *Realistic Evaluation*. Sage, London 1997.

Rajavaara, Marketta: Arviointitutkimuksen hyödynnettävyys. Teoksessa Risto Eräsaari & Tuija Lindqvist &

- Mikko Mäntysaari & Marketta Rajavaara (toim.) Arviointi ja asiantuntijuus. Oy Yliopistokustannus University Press Finland Ltd, Tampere 1999, s. 31-53.
- Rajavaara Marketta: Yhteiskuntaan vaikuttava Kela. Katsaus vaikuttavuuden käsitteisiin ja arviointiin, Sosiaali- ja terveysturvan katsauksia 69, Kela: Vammala 2006.
- Robson, Colin: Käytännön arvioinnin perusteet. Opas evaluaation tekijöille ja tilaajille. Tammi, Helsinki 2001.
- Rossi, Peter H & Freeman, Howard, E: Evaluation: A systematic approach. Sage, Newbury Park, London and New Delhi 1993.
- Stuck, Andreas, Egger, Matthias, Hammer, Andreas, Minder, Christoph, & Beck, John: Home visits to prevent nursing home admission and functional decline in elderly people: systematic review and meta-regression analysis. JAMA 287(2002), 8, s. 1022-1028.
- Toljamo, Maisa, Haverinen, Riitta, Finne-Soveri, Harriet, Malmivaara, Antti, Sintonen, Harri, Voutilainen, Päivi, Mäkelä, Marjukka (2005) Ehkäisevien kotikäyntien vaikutukset iäkkäiden toimintakykyyn. Systemaattisiin kirjallisuuskatsauksiin perustuva menetelmäarvio
- Stakes, FinSoc arviointiraportti 4/2005.
- Vedung, Evert: Arviointiaalto ja sen liikkeellepanevat voimat. Helsinki: Stakes, FinSoc, Työpapereita 2, 2003.
- Vedung, Evert: Nätverksevaluering, Helsinki: Stakes, FinSoc, Working Papers 2006.
- Westman, Riikka, Haverinen, Riitta, Ristikartano, Veera, Koivisto, Juha & Malmivaara, Antti (2005) Perheinterventioiden vaikuttavuus. Järjestelmällinen tutkimuskatsaus. STAKES, FinSoc, Arviointiraportteja, Helsinki 2005.