

12

Tuomarista maailmanparantajaksi: Arvioinnin muuttuva rooli kompleksisessa maailmassa

Petri Uusikylä, Kirsi Hyytinen ja Mari Räcköläinen

Arvioinnin rooli ja merkitys ovat kiinteästi sidoksissa politiikan ja hallinnon toimintatapoihin sekä arvioiteja tilaavien instituutioiden tiedon tarpeisiin ja käyttötarkoituksiin. Tässä artikkelissa tarkastellaan arvioinnin muuttunutta roolia matkalla positivistisesta arviointiperinteestä osallistavaan, valmentavaan ja tuloksia hyödyntävään toimintakulttuuriin. Kompleksinen toimintaympäristö ja arvioinnin kohteena olevien politiikkojen monimutkaistuminen ja yhteenkietoutuminen ovat viime vuosina asettaneet arvioinnille aivan uudenlaisia haasteita. Artikkelissa hahmotetaan miten arviointitoimintaa tulisi kehittää, jotta sen kautta voitaisiin löytää ratkaisuja yhteiskunnan pirullisiin ongelmiin. Samalla myös arvioitsijoiden osaaminen ja kyky hahmottaa nopeasti muuttuvia ja keskinäisriippuvaisia ilmiöitä joutuvat koetukselle. Artikkelin lopussa esitetään, että arvioitsijoiden tulee rohkeasti osallistua globaaliin keskusteluun, haastaa mekanistisia tulkintoja, tarjota uusia näkökulmia ja viitekehyksiä sekä toimia asiantuntevina välittäjinä ja vuoropuhelun synnyttäjinä kansalaisten, päätöksentekijöiden ja muiden globaalia muutosta suuntaavien toimijoiden kesken.

Avainsanat: arviointi, oppiminen, systeeminen näkökulma, kompleksisuus

Johdanto

Arvioinnilla on keskeinen merkitys organisaation oppimisen tukemisessa. Perinteisessä hyvän hallinnon arvioinnissa on ymmärretty, että kokonaisvaltainen ymmärrys organisaation oppimisen tukemisessa merkitsee yksisilmukaisen oppimisen (toiminta) ja kaksisilmukaisen (rakenteet) oppimisen lisäksi myös kolmatta silmukkaa eli metaoppimisen kykyä ja tiedostamista. Ilmiöiden keskinäisriippuvuus ja moninaisuus haastavat tapaamme hahmottaa jatkuvasti muuttuvaa maailmaa ja erityisesti sitä, minkälaiset järjestelmät, rakenteet ja tavat tehdä yhteistyötä toimivat kompleksisessa maailmassa. Resilienssin merkitys korostuu tällaisessa tilanteessa. Koronapandemia on osoittanut konkreettisesti sen, että organisaatioissa resilienssi ei ole vain sitä, että etukäteen varaudutaan tulevaan vaan ennen kaikkea se on järjestelmän kykyä toimia mukautuvasti ja ketterästi yllättävissä tilanteissa. Arviointi kehittyy ajassa ja on yhteydessä muuhun yhteiskunnan kehitykseen sekä oppimis- ja tietokäsityksissä tapahtuviin muutoksiin. Arvioinnit eroavat toisistaan siinä, mikä on niiden tarkoitus, miksi ne tehdään ja mikä on niiden tavoite. Arvioinnin määritelmässä korostetaan, ettei arviointi voi olla spontaania päättelyä kokemusten pohjalta, vaan sillä on aina oltava jokin tarkoitus. Usein määrittelyissä painotetaan, että arviointi on aina myös jollakin tavoin hyödyllistä, käytännöllistä, eettistä sekä tarkoituksenmukaista arviointitiedon käyttäjälle, jolloin arviointia tehdään jonkin toiminnan edistämiseksi.

Pari vuosikymmentä sitten tehokkuuden ja tuottavuuden vaatimus sekä niitä tukeva uusi julkishallinnon johtamistapa nostivat taloudelliset vaikutukset keskeisiksi arvioinnin kriteereiksi myös julkisella sektorilla. Arviointitoiminta oli kehittynyt pitkälti osana tilivelvollisuusraportointia ja siihen liittyvää mittaamista. Sitä oli alettu pitää toimijoiden ja toimenpiteiden olemassaolon oikeuttamisen eli legitimoinnin välineenä. Keskeisenä tavoitteena oli vakuuttaa, että julkisia varoja käytetään tarkoituksenmukaisesti ja että kansalaiset saavat vastinetta maksamilleen veroeuroille. Julkisten varojen käytön seuraaminen on erittäin tärkeää, mutta vanhaan maailmaan sovitettujen arviointikäytännöt eivät riitä yhä monimutkaistuvassa maailmassa. Kapea tilivelvollisuusarviointi tai tulospainotteinen arviointi onkin viime vuosina korvautunut systeemillä arviointinäkökulmalla. Sen tarkoituksena on tuoda arviointiin uudenlaisia tulkintakehikoita ja arviointikriteereitä, joiden kautta olisi mahdollista ylittää mekanistinen ja usein todellisuutta liiaksi yksinkertaistava lineaarinen kausaalijattelu. Viimeistään COVID-19-pandemia laajoine yhteiskunnallisine vaikutuksineen on havahduttanut meidät kompleksisen maailman systeemiseen tarkasteluun. Tästä seuraakin kysymys: miltä osin tämä näkyy arviointiparadigman muutoksena?

Tarkastelemme artikkelissamme arvioinnin murroskohtia ja sitä, miten nämä murrosvaiheet ovat haastaneet arvioinnin menetelmien ja lähestymistapojen kehittämistä. Ensimmäisessä murrosvaiheessa (1980- ja 1990-lukujen vaihteessa) perintei-

nen positivistinen arviointiparadigma alkoi saada haastajia vahvasti osallistavuutta korostavilta konstruktivisteilta. Ajan myötä liikaa yhteisen tulkinnan ja osallisuuden syövereihin uponnut konstruktivismi sai kuitenkin luovuttaa tilaa hyödynnettävyyttä, tuloksellisuutta ja vaikuttavuutta korostavalle suuntaukselle. Tämän tulossuuntautuneen arviointiparadigman takana olivat vahvasti kansainväliset organisaatiot kuten OECD, Maailmanpankki sekä YK:n alaiset järjestöt. Voimakkaasti osallisuuden pohjautuvaa kehittävää arviointia, joka tähtää muutokseen, voidaan puolestaan pitää siirtymänä seuraavaan murrokseen. Ruvettiin korostamaan oppimista arvioinneista ja arviointien aikana.

Kolmannessa murrosvaiheessa systeemisyyden ja ilmiölähtöisyys nostivat polttopisteeseen monikriteeriarvioinnin, joka integroitiin erilaisiin systeemiin lähestymistapoihin. Uusimmassa, käsillä olevassa murrosvaiheessa myös systeemistä lähestymistapaa aletaan haastaa liian staattisena, luonnontieteistä omaksuttuna mallina, ja tilalle tarjotaan uudenlaista kompleksisuusarvioinnin käsitteistöä ja viitekehystä. Arvioitsijoille kasataan myös paineita rakentaa uudenlaisia kompleksisen järjestelmän arviointiin soveltuvia malleja ja arviointiasetelmia. Heitä vaaditaan osallistumaan planetaarisiin missioihin (vrt. Pattonin *Blue Marble Evaluation*) sekä toimimaan globaaleissa kehittäjäverkostoissa kriittisinä ääminä.

Saattaa olla, että olemme siirtyneet tai siirtymässä arviointiajattelussa adaptiiviseen ajatteluun, jossa arvioinnin merkitys näyttäytyy nimenomaan osana organisaatioiden kykyä vahvistaa ennakoivasti resilienssiä, jotta niiden toiminta kompleksisissa ja epälineaarisissa konteksteissa mahdollistuisi.

Tilivelvollisuuden arvioinnista oppimisen ja muutoksen tukemiseen

Arvioinnin perinteisenä funktiona on pidetty tilivelvollisuuden varmistamista eli tulosvastuun toteutumisen arviointia ja seurantaa. Tällaisen arvioinnin ensisijaisena tavoitteena on ollut näyttää asioiden tila ja osoittaa, miten ennalta asetetut tavoitteet on saavutettu. Arvioinnin tarkoituksena voi olla myös tiedon lisääminen ja tietoperustan vahvistaminen, jolloin arvioinnilla pyritään lisäämään ymmärrystä arvioitavasta ilmiöstä, sen taustalla vaikuttavista tekijöistä ja siitä, miten nämä tekijät, suunnitellut toimenpiteet ja ohjelmat liittyvät yhteen (esim. Chelimsky 1997; Chelimsky & Shadish 1997; Stern 2005). Arviointi voidaan nähdä myös kiinteänä osana toiminnan kehittämistä. Tällöin arvioinnin tarkoituksena on tukea instituutiota ja organisaatioita tai edistää ohjelmien ja projektien toimeenpanoa sekä vahvistaa arvioinnin kohteiden osallisuutta oman elämänsä parantamiseen. (Esim. Husen 1999; Patton 1994, 2002; Stern 2005.)

Arviointitraditioiden arviointiasetelmissa ilmenee arvioinnin tavoite sekä se, millaisin menetelmin arviointitietoa hankitaan, millainen on arvioijan tehtävä ja arvioijan ja arvioinnin kohteen välinen suhde ja miten arvioinnin tuloksia käytetään. Vallinnat arviointiasetelmissa ilmentävät muun muassa sitä, miten vahvasti arvioinnin kohde on osallisena arvioinnissa ja onko arviointi lähtökohdiltaan kontrolloivaa vai luottamukseen pohjautuvaa. Arvioinnin tarkoituksen lisäksi juuri arviointisuhteen luonne erottaa eri arviointitraditiot toisistaan (Rökköläinen 2011; Rökköläinen & Räisänen 2013).

Arviointiparadigma käsittää ne kriteerit, joiden perusteella arvioinnin kysymyksiä määritetään ja valitaan ja miten niitä lähestytään arviointiteoreettisesti ja metodologisesti. Paradigma on aikansa ja yhteisönsä tuottama, ja sitä voidaan tarkastella esimerkiksi arvioinnin sukupolvina (esim. Guba 1990; Guba & Lincoln 1989). Arviointinäkemys kuvaa muun muassa arvioijien ja arviointiyhteisöjen käsitystä arviointitiedon luonteesta, käsitystä siitä, mitä on mahdollista tietää, miten tietoa tuotetaan ja miten sitä voidaan käyttää. Arviointinäkemykset kehittyvät, kun syntyy kritiikkiä aikaisempaa arviointiajattelua kohtaan. Tällöin arvioinnin perusteissa alkaa syntyä jännitteitä, jotka liittyvät erityisesti arviointitarpeen määrittelyyn, arvioinnin tarkoitukseen sekä arvioijan ja arvioinnin kohteen rooleihin ja niiden keskinäisiin suhteisiin. (Rökköläinen 2011.)

Mittaamiseen, tulostavoitteisiin ja ennalta määriteltyihin standardeihin nojaviin ja objektiivisuutta painottaviin arviointien jälkeen ryhdyttiin korostamaan, että arviointitieto syntyy arvioijan ja arvioitavan välisessä vuorovaikutuksessa, jossa se konstruoidaan aina uudelleen. Arvioijakin on silloin toimiva subjekti eikä ulkopuolinen, riippumaton tarkkailija. Arvioijan tehtävänä on toimia oppimisprosessin virittäjänä ja keskustelun johtajana, jolloin tavoitteena on eri osapuolten konsensus ja uuden oppiminen. Tämän lähestymistavan taustalla on vahva konstruktivistinen paradigma, jonka edustajat kritisoiivat edellisiä sukupolvia arvioitavan kohteen yksipuolistamisesta ja passivoinnista. Konstruktivistisessä näkemyksessä painotettiin, ettei todellisuus ole objektiivinen eikä yhtenäinen, vaan siihen vaikuttavat monet inhimilliset, sosiaaliset ja kulttuuriset seikat. (Ks. esim. Guba & Lincoln 1989; Rökköläinen 2011.)

Myös käsitys arvioijan roolista sekä arvioijan ja arvioinnin kohteen välisestä suhteesta alkoi muuttua, jolloin ryhdyttiin painottamaan arvioinnin kohteen aktiivista roolia arviointien toteuttamisessa ja voimaantumista omassa asiassa. Osallistumisesta edistävissä ja voimaannuttavissa arvioinneissa keskeiseksi vaikuttavaksi tekijäksi on nähty luottamus arvioijan ja arvioinnin kohteen välillä, koska arviointiin osallistuvien suhde ei perustu ulkopuoliseen kontrolliin, vaan yhteistoiminnallisuuteen, vuorovaikutukseen ja tasavertaisuuteen. Arvioijan tehtävänä on silloin toimia oppimisprosessin virittäjänä. Samaan aikaan ryhdyttiin painottamaan myös kontekstuaalisen analyysin ja autenttisen arvioinnin merkitystä erityisesti työssä oppimisessa ja oppivissa organisaatioissa. (Ks. Fetterman 2001; House 1993; 1997; Rossi, Freeman

& Lipsey 1999; Patton 1994; 1997; 2016; Poikela 2004; Poikela & Rökköläinen 2006; Rökköläinen 2011.)

Myös perinteiset käsitykset arviointitiedon yleistettävyydestä ja arviointitiedon käytöstä ovat muuttuneet. Tulosten yleistettävyyden ja ”objektiivisen” totuuden sijaan uusimmissa kontekstisidonnaisuutta ja osallistumista painottavissa arvioinneissa korostuvat erityisesti arvioinnin merkitys arvioinnin kohteelle ja arvioitavalle asialle sekä arviointien hyödyllisyys ja arviointien seuraamukset (ks. Patton 1997 *utilization focused evaluation* ja Fettermann 2001 *empowerment evaluation*). Osallistavassa ja kehittävässä arvioinnissa korostuvat erityisesti arvioinnin seuraamukset, kun kiinnostus kohdistuu siihen, mikä merkitys arvioinnilla on arvioitavalle asialle, esimerkiksi oppimiselle. Osittain kritiikkinä sellaista käsitystä vastaan, että tieto syntyy vain yksilöllisen ja yhteisöllisen konstruoinnin tuloksena, arvioinnissa vahvistui realistinen lähestymistapa. Realistisessa arvioinnissa nähtiin, että objektiivinen ja tiedostettavissa oleva maailma on olemassa, jolloin on mahdollista selvittää myös asioiden todellinen tila ja jopa kausaalisuhteita. Arvioinnin tarkoituksena on tällöin muutokseen vaikuttavien mekanismien ymmärtäminen. (Esim. Anttila 2005; Saari 2002; ks. myös Kushner 2005.) Tätä lähestymistapaa on puolestaan kritisoitu merkityksen tarkastelun puutteesta, arvioitavan kohteen esineellistämistä, vallitsevien olojen liiallisesta vakiinnuttamisesta ja siitä, etteivät arvioinnin tulokset siirry käytännön ratkaisuiksi (ks. esim. Rajavaara 2007; Vedung 1997).

Niin sanotut teoriaperustaiset arvioinnit (*theory-based evaluation*, ks. esim. Stame 2004ab) syntyivät kritiikkinä liian menetelmäpainotteisia arviointeja kohtaan. Teorian korostaminen liittyy myös tavoitteeseen vahvistaa arviointien yleistettävyyttä ja arvioinneista oppimista. Ajatuksena on, että arvioinneista on mahdollista oppia, kun vahvistetaan yleistettävyyttä tuottamalla teoreettisia jäsennyksiä arvioitavasta kohteesta, kuten kehittämisohjelmasta tai uudistuksesta. Samalla voidaan vastata kehittämishankkeisiin usein liittyvään suureen tiedon tarpeeseen. Näin voidaan myös kehittää teoreettisesti perusteltu rakenne, viitekehys tai prosessi arvioinnin toteuttamiseksi. Liikkumalla teorian ja empirian välillä ja muodostamalla hypoteeseja voidaan ylittää arviointien kertaluonteisuus ja vahvistaa tulosten kestävyttä. Luovissa kehittämishankkeissa arviointi voi olla tutkivaa toimintaa ja se voidaan nähdä ongelmanratkaisuprosessina, joka tukee käytännön kehittämistyötä tuottamalla tietoa siitä, miten hankkeet toimivat, miksi ja missä olosuhteissa (ks. myös Anttila 2005).

Painotettaessa osallisuutta arvioinnissa tavoitteena on tukea ihmisiä ja organisaatioita käyttämään arviointia hyödyllisenä, käytäntöihin vaikuttavana työvälineenä. Osallisuuden ja arviointitiedon käytön välinen yhteys on kuitenkin monitahoinen ja kysymys hyödyllisyydestä jännitteinen. Keitä ovat ”osallistujat” tai ”osalliset”? Ovatko he arvioinnin tilaajia ja maksajia vai heitä, joiden toimintaa arvioidaan tai jotka ovat arvioitavien palveluiden kohderyhmää ja ”loppukäyttäjää”? Painotettaessa arvioinnin hyödyllisyyttä ja käyttökelpoisuutta voidaan esimerkiksi korostaa, että arviointi on

toteutettava läheisessä yhteistyössä arvioinnin tuloksia käyttävien, kuten arvioinnin tilaajan tai rahoittajan kanssa, jotta arvioinnin tuloksia ylipäättänsä voidaan ja halutaan käyttää (esim. Patton 1997 *utilization-focused evaluation*). Toisaalta osallistumista voidaan tarkastella arvioijan ja eri intressiryhmien/asianosaiesten (*stakeholders*) välisenä suhteena, jolloin arvioijan tulee olla tietoinen eri tahoista ja sisällyttää arviointisuunnitelmaan toimenpiteitä, joilla varmistetaan keskeisten yhteistyötahojen osallistuminen arvioinnin toteutuksessa. Siten arviointeihin tulisi osallistua niiden henkilöiden, joiden toimintaan arviointi kohdistuu ja joihin uudistusten vaikutukset ulottuvat (esim. Rossi, Freeman & Lipsey 1999). Syvästi osallistava ja yhteistoiminnallinen arviointi voi lisätä yksilöiden ja yhteisöjen voimaantumista, vapautta, luovuutta ja vaikutusmahdollisuuksia. Tällöin arviointia ei toteuteta vain hyödyllisen arviointitiedon saamiseksi, vaan tavoitteena on auttaa arviointiin osallistuvia toteuttamaan arviointia itsenäisesti, käyttämään arvioinnin tuloksia oman asiansa ajamiseen ja muutoksen toteuttamiseen, ja näin edistää yksilöiden ja yhteisöjen vaikutusmahdollisuuksia omaa elämää koskevassa asiassa. Voimaannuttavaan arviointiin liittyy olennaisesti tulevaisuuden tarkastelu ja tavoitteiden asettaminen kehittämistyötä varten muutosten aikaansaamiseksi (esim. Fetterman 2001; Saunders 2006).

Osallisuuden, arviointikohteen aktiivisuuden ja voimaantumisen vahvistuminen arviointitraditioissa merkitsee, että arviointiasetelmissa ja -järjestelmissä siirtyään kontrollista luottamukseen. Voidaan puhua jopa luottamukseen perustuvasta arviointikulttuurista. Osallistumisen syvyyden, luottamuksen ja oppimisen välillä on yhteys. Osallisuus arviointiprosessin kaikissa vaiheissa arviointitarpeen määrittämisestä ja arviointikysymysten muotoilusta aina arviointitulosten tulkintaan ja käyttöön asti vahvistaa luottamusta arviointisuhteessa, mikä puolestaan edistää oppimista ja arviointitulosten käyttöä muutosvoimana. Luottamus, osallisuus ja osallistava laadunvarmennus vaikuttavat myönteisesti tulosten hyödyntämiseen ja oppimiseen. Arviointitiedon osuvuus, validiteetti, on merkityksellisempää oppimiselle kuin kaiken kattava läpinäkyvyys ja yleistettävyyys. Myös liiallinen standardisointi ja tulosten kontrolli heikentävät oppimisen tuloksia. (Rökköläinen 2011; Räisänen & Rökköläinen 2014; Rökköläinen & Räisänen 2013; ks. myös Gardner 2006.)

Osallisuutta ja voimaantumista pidetään kehittävän arvioinnin keskeisinä tavoitteina ja jopa laatuvaatimuksina ja välttämättöminä ehtoina. Kehittävän arvioinnin kriteerinä on, ettei arviointia tehdä vain hyödyllisen arviointitiedon saamiseksi, vaan arviointiprosessin tarkoituksena on erityisesti tukea osallistujia oman toimintansa kehittämisessä ja lisätä heidän vaikutusmahdollisuuksiaan omassa asiassaan. Voimaannuttava arviointi painottaa syvän ja laajan osallistumisen lisäksi käytännön elämään vaikuttamista ja konkreettisten muutosten aikaansaamista. Kehittävässä arviointitraditiossa olennaista on, että arviointi on muutosvoima ja oppimista tapahtuu jo arviointiprosessin aikana. Jopa tilivelvollisuus voi olla viisasta (*"intelligent accountability"*), mikäli se vahvistaa luottamusta, edistää osallisten identiteettiä ja

auttaa asettamaan pitkän tähtäimen tavoitteita ja katsomaan tulevaisuuteen. Tämä muuttaa perinteistä tilivelvollisuuden arviointisuhdetta niin, että myös osalliset kantavat vastuuta arvioinnista ja jakavat sen riskit arvioitsijoiden kanssa. Voimaannuttavissa, oppimista edistävissä arvioinneissa käytetään tiedon tuottamiseksi usein itsearviointia ja reflektiota. (Crooks 2002; O’Neil 2002; Rökköläinen 2011; Rökköläinen & Räisänen 2013; Poikela & Rökköläinen 2006; ks. myös Patton 1997; 2002.)

Luottamus ilmenee erityisesti arvioinnin kohteen ja arvioijan suhteen luonteessa. Luottamusta edistävässä arvioinnissa arvioijan rooli muuttuu ohjaavaksi eikä täydellinen kontrolli eikä objektiivisuus (arvioijan ulkopuolisuus) ole mahdollista, vaan myös arvioija on osallisena arviointiprosessissa. Arvioinnin hyödyllisyys ja käyttökelpoisuus edellyttävät, että arviointi toteutetaan läheisessä yhteistyössä arviointitiedon tuottajien, arvioinnin kohteen ja tiedon käyttäjien kanssa ja arviointi kehittää samalla arvioitavaa toimintaa. Arvioijan tehtävänä on varmistaa sekä arvioinnin kohteen osallisuus arviointikysymysten asettamisessa että arviointitulosten oikeudenmukainen käyttö. Palaute on olennainen osa kehittävää ja luottamusta edistävää arviointia, ja ihmisillä tulee olla mahdollisuus itse tarkistaa ja arvioida tiedon luotettavuus. Arvioinnin kohteena olevat ovat oman toimintansa aktiivisia reflektioijia, he ovat osallisina arviointiprosessissa ja heillä on siinä vastuuta ja vaikutusvaltaa (ks. esim. O’Neil 2002). Luottamusta edistävässä arviointikulttuurissa painotetaan organisaatioiden ja yhteisöjen omaa laadukasta hallinnointia eikä niinkään kontrollia ulkoapäin. Tällöin arviointitarve nousee ja arviointikysymykset asetetaan toiminnan sisältä. Osallistava arviointi perustuu yhdessä neuvoteltuun arvoperustaan. (Ks. esim. Rökköläinen 2011.)

Osallistavista ja voimaannuttavista arvioinneista on siirrytty uudelleen painottamaan arviointien tulosten hyödyntämistä ja arvioinnin tuloksista oppimista muutosvoiman lähteenä. Näkemykset siitä, että uutta tietoa antava arviointiprosessi on hyödyllinen väline päätöksentekijöiden työssä ja sovellettaessa kehittämistyön tuloksia uusissa kehittämiskohteissa, ovat voimistuneet organisaatioiden vahvistaessa tiedolla johtamista (*knowledge management*) ja oppimista tuloksista (esim. Mayne 2008; 2011; Vähämäki 2017; Vähämäki, Schmid & Molander 2011; Vähämäki & Verger 2019). Tiedon tuottamisen tapana on entistä useammin itsereflektio ja aktiivinen dokumentointi eli arvioinnin kohde itse tuottaa arviointitietoa aktiivisesti ja systemaattisesti ja on vahvasti osallisena arviointiprosessissa. Tällöin arvioinnin käyttökelpoisuus (*utility*) tarkoittaa, että arviointi palvelee käyttäjän tiedontarpeita ja arviointitulokset ovat luotettavia, jotta niiden pohjalta on mahdollista tehdä muutoksia. Arvioinnin toteuttamiskelpoisuus viittaa arvioinnin käytännön toteuttamisen järkevyyteen ja realismiin (vrt. myös Patton 2002; 2016; Rossi ym. 2005).

Arviointia osana tiedolla johtamista ja tuloksista oppimista on ryhdytty painottamaan esimerkiksi kehityspolitiikkaan ja kehitysyhteistyöhön kohdistuvissa kehitysevaluoinneissa (OECD 2016; 2019; USAID 2017; Ministry for Foreign Affairs of

Finland 2019). Sen lisäksi, että arviointia pidetään välttämättömänä läpinäkyvyyden ja hyvän hallinnoin varmistamiseksi, sillä nähdään olevan tärkeä merkitys demokraattisten prosessien edistämiseksi. Usein arvioinnin kohteena on avun, rahoituskanavien, hankkeiden, ohjelmien ja aloitteiden tehokkuus ja vaikuttavuus. Painotus on ollut tilivelvollisuudessa sekä politiikkatavoitteiden, strategisen suunnittelun ja päätöksenteon varmistamisessa. Myös näkemys tulosperustaisuudesta ja selkeiden arviointikriteerien tarpeesta vahvistui vaikuttavan kehitysyhteistyön ja sen arvioinnin perustana. Kehitysevaluoinnit toteutetaan usein kuitenkin hyvin kompleksisissa tilanteissa, läpileikkaavien teemojen ja (poliittisten) vaatimusten ristipaineissa ja ongelmat ovat hyvinkin viheliäisiä (*wicked problems*), eikä muutosta ole yksiseitteistä todentaa. Viime aikoina onkin ryhdytty entistä painokkaammin edellyttämään jälleen arvioinnin hyödyllisyyttä, kehittävyttä ja oppimisen mahdollistamista. Tällöin evaluoinnin lähtökohtana on kontekstuaalisuus eli tarvelähtöisyys ja tulosten soveltaminen paikallisiin olosuhteisiin. Käytäntöyhteisöt (*communities of practice*) ovat nimitäin merkittäviä tiedon jalostamisessa, tiedon muuntamisessa ymmärrykseksi (esim. Bester 2016; 2008). Myös oppimisen ja tulosohtauksen (ns. *results-based management*) yhteyttä on ryhdytty painottamaan kehitysyhteistyön toimeenpanossa vaikuttavien tulosten välttämättömänä ehtona (esim. OECD 2014; 2016; 2019).

On osoittautunut, että evaluointien yhteydessä oppimisesta puhutaan kuitenkin hyvin yleisesti eikä useinkaan erotella sitä, kenen oppimisesta on kysymys (yksilön, yhteisön vai organisaation vai kokonaisten systeemien). Usein painottuu myös evaluointien tulosten hyödyntäminen eikä niinkään oppiminen arviointiprosessissa. Oppiminen määritellään helposti kapeasti tulosten levittämisenä ja hyvien käytäntöjen jakamisena. Lisäksi tuloksista oppiminen on harvoin mielletty tekemällä oppimisena omassa työssä. (Ks. esim. Vähämäki 2017; Vähämäki & Verger 2019.) Tämän vuoksi monimutkaistuvissa ympäristöissä muutosta tukevaa systeemistä ymmärrystä on ollut evaluoinneissa vaikea tavoittaa. Ulkoministeriö toteutti laaja-alaisen evaluoinnin tietojohtamisesta, tuloksista oppimisesta ja oppien hyödyntämisestä Suomen kehityspolitiikassa ja yhteistyössä (Ministry for Foreign Affairs of Finland 2019). Evaluoinnin mukaan tarvitaan sekä dokumentoitua että hiljaista tietoa tuloksista (mitä?) ja tulosten saavuttamiseen vaikuttavista tekijöistä (miksi? ja miten?). Tulostieto ei siten ole vain indikaattoreita ja dataa saavutuksista, vaan se käsittää myös ymmärryksen siitä, mistä tulokset muodostuvat, miten ne syntyvät, keitä on mukana ja miten niistä opitaan. Oppiminen organisaatiossa puolestaan edellyttää motivaatiota, kannustimia, osaamista sekä mahdollisuuksia ja institutionaalista kysyntää tuloksista oppimiselle. Evaluointi oli teoriaperusteinen, ja se vahvisti ymmärrystä tulosohtaukseen vaikuttavista rakenne- ja prosessitekijöistä ja auttoi selkiyttämään siihen liittyvää käsitteistöä.

Kompleksisten systeemien haastavuus on ohjannut systeemisen johtamisen malleja kohti kestävää kehitystä. Niissä yksi keskeinen elementti on oppiminen, mutta muutoksen johtamista ei nähdä kronologisena prosessina, vaan sitouttaminen, moti-

vointi, vakuuttaminen sekä läpinäkyvyyden vahvistaminen ja tilivelvollisuusvaateisiin vastaaminen toistuvat siinä usein sykleinä ja eri vaiheet voivat olla samanaikaisia (esim. Dreier, Nabarro & Nelson 2019; ks. myös OECD 2019).

Seuraavaksi tarkastelemme lähemmin arvioinnin murrosta, jossa siirrytään kohti systeemistä ymmärrystä arvioitavista ilmiöistä.

Muutoksen tukemisesta sen systeemisen ymmärrykseen

Jo ensimmäisten arviointisukupolvien aikana eroteltiin käsitteellisesti summatiivinen, kokoava arviointi ja formatiivinen, kehittävä arviointi. Summatiivinen arviointi tarkoittaa vaikutusten ja tulosten arviointia ja formatiivinen puolestaan prosessiarviointia. Summatiiviset arvoinnit ovat lähtökohtaisesti useimmiten ulkopuolisen arvioijan tekemiä, kun taas formatiiviset arvoinnit itse- ja prosessiarviointeja. (Ks. esim. Scriven 1967; Cronbach 1982.) Voidaan ajatella, että kehittävä arviointi – sellaisena kuin se alun perin nousi esiin (Patton 1994; 1997; 2011) – muotoutui eräänlaiseksi kolmanneksi arviointityypiksi summatiivisen ja formatiivisen arvoinnin rinnalle.

Sittemmin kehittävä arviointi on edelleen kehittynyt arviointiajatteluksi, joka pyrkii erityisesti vastaamaan arvoinnin haasteisiin monimutkaistuvassa yhteiskunnassa, jota luonnehtivat yhteiskunnan monimutkaiset vuorovaikutussuhteet sekä jatkuvat ja nopeat muutokset (ks. Todd & Wolpin 2010; Weijermars & Wesemann 2013). Kehittävän arvoinnin pääasiallisena tavoitteena on yhtäältä tuoda kokonaisvaltaista ymmärrystä monimutkaisen maailman ilmiöistä ja siihen liittyvien eri tekijöiden ja toimijoiden välisistä monimutkaisista ja dynaamisista vuorovaikutussuhteista. Toisaalta se pyrkii tuottamaan reaaliaikaisesti tietoa ja ymmärrystä yhteiskunnan muutos- ja sopeutumiskykyä vahvistamaan. Se nojaa menetelmällisesti sekä ennakointiin, reflektointiin, sopeuttamiseen että moninäkökulmaiseen arviointiin. Nämä lähestymistavat pyrkivät varmistamaan, että arviointi on lähtökohtaisesti moninäkökulmaista, ottaa huomioon useat erilaiset vaihtoehtoiset tulevaisuudet ja vaikutukset sekä on keskeinen osa strategista päätöksentekoa ja organisaation oppimista (Patton 2011; Hyytinen 2017).

Arviointitoiminta on kehittynyt edellisessä luvussa kuvatulla tavalla osana tilivelvollisuusraportointia ja siihen liittyvää mittaamista. Siksi siitä kehittyi vuosien saatossa toimijoiden ja toimenpiteiden olemassaolon oikeuttamisen väline. Arvioinneilla pyrittiin pitkään ennen kaikkea vakuuttamaan, että julkisia varoja käytetään tarkoituksenmukaisesti ja kansalaiset saavat vastinetta maksamilleen veroeuroille. Usein tätä kuvaavat mittarit ovat nojanneet perinteisiin tuottavuutta ja tehokkuutta kuvaaviin mittareihin. Tilivelvollisuus oli pitkään julkisen hallinnon pääasiallinen arvoinnin lähestymistapa.

Julkisten varojen käytön seuraaminen on erittäin tärkeää, mutta perinteisten, mitareihin nojaavien arviointikäytäntöjen on todettu olevan riittämättömiä systeemisiä haasteita tarkasteltaessa ja niiden vaikutuksia hahmotettaessa. Ne perustuvat suoraviivaiseen panos-tuotos-vaikutus-tyyppiseen ajatusketjuun, joka sopii hyvin rajattuihin ilmiöihin, joiden syy-seuraussuhteet tunnetaan. Ne eivät kuitenkaan tuota riittävän reaaliaikaista eivätkä kokonaisvaltaista tietoa systeemisessä kontekstissa.

Perinteisesti arviointi kohdistuu arvojen ja onnistumisten tarkasteluun kapea-alaisesti: tyypillisesti arvioinneissa tarkastelun kohteena on muutos teknologian kehityksessä ja tuottavuudessa. Sen sijaan laaja-alaiset vaikutukset laajempien yhteiskunnallisten arvojen tai kansalaisiin kohdistuvien hyötyjen näkökulmasta, joista esimerkkinä voi mainita yksilön hyvinvointiin ja elämän laatuun liittyvät vaikutukset, ovat jääneet vähäiselle huomiolle. Arvioinnin kohdentuminen vaikutusten teknis-taloudellisiin ulottuvuuksiin johtuu siitä, että niiden muutokset ovat mitattavissa kvantitatiivisesti (Dahlberg 2018). Teknisiin ja taloudellisiin vaikutuksiin kohdentuva lähestymistapa on kuitenkin liian kapea-alainen laajojen systeemisten ja dynaamisten muutosten ja vaikutusten arvioimiseksi (Djellal & Gallouj 2013a; Hyytinen 2017). Yhteiskunnan isot systeemiset haasteet ja niiden vaatimat laaja-alaiset ratkaisut ovatkin vauhdittaneet tarvetta kehittää myös uudenlaisia arvioinnin lähestymistapoja, joilla vaikutusten systeeminen ja dynaaminen luonne saadaan esiin.

Kehittävän arvioinnin konkretisoiminen ja saattaminen käytäntöön arviointityössä edellyttää konkreettisten menetelmien ja lähestymistapojen kehittämistä. Muutostarve on luonut pohjan monikriteeriarvioinnin kehittämiseksi, jossa systeemiin ilmiöihin liittyvää muutosta sekä eri toimijoille syntyvää arvoa ja vaikuttavuutta tarkastellaan monien eri arvojen näkökulmista (Djellal & Gallouj 2013a). Perinteisten teknisten ja taloudellisten arvojen rinnalla monikriteeriarvioinnissa huomio kiinnitetään myös ratkaisujen sosiaaliseen ja systeemiseen luonteeseen sekä niiden synnyttämään laatuun ja arvoon laajemmin yhteiskunnallisesta tai kansalaisten saaman hyödyn näkökulmasta. (Gadrey 2010; Djellal & Gallouj 2010; 2013a; Rubalcaba, Michel, Sundbo, Brown & Reynoso 2012.) Tavoitteena on tehdä näkyväksi sellaiset vaikutukset, jotka eivät ole materiaalisia tai helposti numeroilla mitattavissa monimutkaisessa ja systeemisessä toimintaympäristössä.

Perustelu uudenlaisen ja laajemman arvioinnin lähestymistavan tarpeelle nojaa tukevasti ”laaja-alaista innovaatiota” koskevaan keskusteluun (Dosi, Freeman, Nelson, Silverberg & Soete 1988; Kline & Rosenberg 1986; Lundvall 1992; Nelson & Winter 1982). Keskustelu nostaa polttopisteeseen ratkaisujen systeemisyiden: yhteiskunnan isoihin haasteisiin vastaavien ratkaisujen kehittäminen ja käyttöönottoaminen ovat luonteeltaan monimutkaisia ja vuorovaikutteisia prosesseja, jotka sisältävät paljon epävarmuustekijöitä. Systeemisyys ei liity yksinomaan kehittämisprosessin vaan myös ratkaisujen luonteeseen. On itsestään selvää, että isojen haasteiden ratkaiseminen edellyttää useiden eri teknologioiden ja niihin kiinnittyvien palvelujen

yhdistelmiä (Harrison, Klein & Browne 2010). Sen lisäksi niiden käyttöönotto ja levittäminen edellyttää laajasti muutoksia järjestelmän eri tasoilla aina päätöksenteosta kansalaisten arjen valintoihin asti (Geels 2002, 2004; Geels & Kemp 2007).

Palveluinnovaatioita käsittelevässä kirjallisuudessa (Djellal & Gallouj 2013b; Rubalcaba ym. 2012) monikriteeriarvioinnin keskeisiksi vaikuttavuuden ulottuvuuksiksi on tunnistettu kuusi näkökulmaa, joiden kannalta hyötyjä ja vaikutuksia tarkastellaan. Näkökulmat ovat (1) teknologinen, (2) talouden sekä markkinoiden kehitys, (3) uudet innovaatiot, (4) vuorovaikutukseen liittyvät tekijät, (5) kansalaisten saamat ja (6) maineeseen liittyvät hyödyt. Monikriteeriarviointia on sovellettu ja kehitetty edelleen erilaisissa systeemisten ilmiöiden konteksteissa vaikuttavuuden varmistamiseksi ja osoittamiseksi. Esimerkiksi yhteiskunnallisten kokeilujen yhteydessä mallia on sovellettu integroimalla monikriteeriarviointia kehittävää arviointia ja oppimista tukevaan vuorovaikutteiseen oppimisprosessiin (Hyytinen, Saari & Elg 2019). Moninäkökulmaisuus arvioinnissa tekee näkyväksi kokeilujen kokonaisvaltaiset ja laaja-alaiset vaikutukset. Se kiinnittää huomiota tasapainoisesti kokeilujen sosiaalisiin (ihmiseen ja yhteisöön liittyviin) ja teknistaloudellisiin piirteisiin. Kokeilun arvo on punnittava myös sen kautta, millaisia kokemuksia se on saanut aikaan käyttäjissä ja lähiyhteisössä. Kokeilujen laaja-alaisia vaikutuksia tarkastellaan mallissa seuraavien ulottuvuuksien avulla: vaikutukset kansalaisiin, työntekijöihin ja väestöön sekä brändiin, palvelujen ja teknologioiden integroitumiseen sekä talouteen. Mallin vuorovaikutteisuus (Patton 2011; Saari & Kallio 2011) tuo yhteisen pöydän ääreen kokeilun toteuttajat, niiden käyttäjät ja kokeilutoimintaa resursoivat tahot. Yhdessä he punnitsevat kokeilujen onnistumista ja tunnistavat keinoja niiden hyötyjen ja vaikutusten varmistamiseksi toteutuksen eri vaiheissa. Parhaimmillaan aktiivinen vuorovaikutus toimijoiden kesken toimii kehittäjäjoukon yhteisenä oppimisalustana sekä tukee jatkuvaa oppimista ja toiminnan suuntaamista läpi kokeilun.

Systeemisyyden tarkastelemiseksi ja ymmärtämiseksi on kehitetty useita erilaisia menetelmiä ja lähestymistapoja, joita arviointitoiminnassa on enenevästi viime vuosina hyödynnetty. Agenttipohjainen mallinnus, verkostanalyysi ja systeemidynamiikka mallinnus ovat esimerkkejä lähestymistavoista, joilla systeemin osatekijöiden välistä vuorovaikutusta pyritään ymmärtämään (Hanneman & Riddle 2011; Helbing 2012; Holz ym. 2015; Williams & Hummenbrunner 2011). Eri lähestymistavoilla on omat perustelunsa ja kannattajansa; lähestymistavan valinta myös perustuu tarkastelun käyttökohteeseen (Helbing 2012; Williams & Hummenbrunner 2011).

Yleisesti tarkasteltuna systeemiajattelu tarjoaa vaihtoehdoisen lähestymistavan ymmärtää vaikutusten syntymistä ja vaikutusketjuja verrattuna perinteiseen input-output-analyysiin. Systeemiset lähestymistavat tekevät näkyväksi muutoksen mekanismit, joiden välityksellä interventiot ja panostukset muuttuvat tuloksiksi ja vaikutuksiksi (Chen 2005, 231). Toisin sanoen ne pureutuvat niin sanottuun arvioinnin mustaan laatikkoon. Perinteinen loogiseen malliin perustuva arviointi ei auta ym-

märtämään niitä mekanismeja ja osatekijöitä, joilla on merkitystä vaikuttavuuden aikaansaamisessa (Dyehouse, Bennett, Harbor, Childress & Dark 2009, 187), kun taas systeemimallit kiinnittävät hyvinkin yksityiskohtaisesti huomion eri tekijöiden ja toimijoiden välisiin monimutkaisiin kytkentöihin ja dynaamisiin vuorovaikutussuhteisiin, jotka puolestaan auttavat ymmärtämään tekijöitä jotka yhtäältä estävät tai toisaalta vauhdittavat vaikuttavuuden syntymistä (Merrill ym. 2013).

Vaikka viime vuosina systeemiseen arviointitapaan on kohdistunut paljon kiinnostusta (Cabrera ym. 2008; Dyehouse ym. 2009; Mayne 2012; Nieminen, Hyytinen, Salminen & Ruutu 2020; Nieminen & Hyytinen 2015; Patton 2011; Williams & Hummelbrunner 2011), se ei vielä ole valtavirtaa arvioinnissa ja siksi on saattanut jäädä vieraaksi monille arvioinnin ammattilaisille. Tätä on selitetty kirjallisuudessa yhtäältä lähestymistapojen moninaisuudella (Hargreaves & Podems 2012) ja toisaalta perinteisten tieteellisten raja-aitojen olemassaololla – lähestymistapojen epistemologisten perusteiden eroista erilaisiin tutkimusstrategioihin. Siksi erilaisia lähestymistapoja ei ole lähdetty yhdistämään ja hyödyntämään erityisesti yhteiskuntatieteelliseen tradition nojaavien arvioitsijoiden keskuudessa (Hargreaves & Podems 2012). Systeemiset lähestymistavat perustuvat tyypillisesti luonnontieteelliseen tutkimusperinteeseen ja pyrkivät luomaan kuvan vaikutusten syntymisestä selkeiden vuorovaikutus- ja riippuvuussuhteiden kautta. Huolimatta siitä, että ylläkuvatut systeemiset arviointitavat antavat dynaamisen kuvan muutoksesta ja eri tekijöiden ja toimijoiden välisestä vuorovaikutuksesta, saattavat ne yksinkertaistaa liiaksi koko ajan monimutkaistuvia ilmiöitä. Haasteista huolimatta on tärkeää, että arvioinnin ammattilaiset tutustuvat systeemijatteluun ja hyödyntävät arvioinnin kohteeseen sopivia systeemisiä menetelmiä ja lähestymistapoja (Cabrera ym. 2008; Williams & Imam 2007).

Erilaisten systeemisten lähestymistapojen hyödyntäminen tukee yhteiskunnan toimintojen kokonaisvaltaista kehittämistä. Tällöin ymmärretään, että kokonaisuus on enemmän kuin osiensä summa. Esimerkiksi suomalaisen innovaatiojärjestelmän kehittäminen ei voi rajautua vain TKI-toimintojen tai rahoituksen kehittämiseen, vaan se on samaan aikaan sidoksissa myös koulutusjärjestelmän toimivuuteen, kaupunkisuunnitteluun, liikenteen ja logistiikan ratkaisuihin, kokonaisturvallisuuteen tai sosiaaliseen tasa-arvoon. Ilmiölähtöisessä tarkastelussa (ks. Raisio, Jalonen & Uusikylä 2018; Lähteenmäki-Smith & Virtanen 2020) siirrytään väistämättä rajatusta systeemisestä arvioinnista laajempaan kompleksisuuden arviointiin.

Arviointia kaaoksen reunalla – kuinka kompleksisuus haastaa arvioinnin?

Edellä on kuvattu sekä osallisuuden ja oppimisen merkitystä arvioinnissa että monitahoarvioinnin ja systeemisen arvioinnin erityispiirteitä suhteessa tavoitteiltaan perinteisempään tilivelvollisuuden arviointiin. Tarkastelussa todettiin, että systeemisessä arvioinnissa kiinnitetään huomiota ilmiöiden keskinäisriippuvuuksiin ja ymmärretään, että yhteen kytkeytyneessä maailmassa kokonaisuus on enemmän kuin osiensa summa. Systeemisen arviointiajattelun taustalla on usein luonnontieteellisestä (ekologisesta) ajattelusta kumpuava käsitys siitä, että häiriötilojen jälkeen systeemi pyrkii aina takaisin tasapainotilaan. Tässä osiossa haluamme haastaa tämän käsityksen ajatuksella, että systeeminen epätasapainotila ei ole mikään hetkittäinen poikkeama, vaan jatkuvan emergenssin ja kompleksisuuden aikaansaama normaali tila. Ei ole tasapainotilaa uutena normaalina, johon päädytään sen jälkeen, kun häiriötilasta on selvitty. Kaaoksen reuna on eräänlainen transiitotila järjestyksen ja epäjärjestyksen välillä. Tämä siirtymävyöhyke on rajatun epävakan alue, joka saa aikaan jatkuvan dynaamisen vuorovaikutuksen järjestyksen ja häiriön välillä. Tämä nostaa esiin kysymyksiä: Millainen arviointi soveltuu kompleksiseen, keskinäisriippuvaiseen ja jatkuvassa muutoksessa olevaan maailmaan? Miten näkemys haastaa perinteisiä arviointitiedon hyödyntämiseen ja oppimiseen liittyviä konventioita?

Kompleksisuusajattelua on viimeisen parinkymmenen vuoden aikana sovellettu mitä erilaisimpien ilmiöiden yhteydessä (Mitleton-Kelly 2003; Byrne & Callaghan 2014; Bale, Vargara & Foxon 2015; Stroh 2015; Ison & Straw 2020). Sen avulla on jäsennetty muun muassa poliittishallinnollisten instituutioiden haasteita, organisatoristen muutosprosessien läpivientiä, innovaatioiden syntymistä ja omaksumista, liiketoimintaverkostojen toimintaa sekä uusien teknologioiden leviämistä. Kompleksisuuslinssit auttavat näkemään ongelmat ja haasteet uudessa valossa, mikä on välttämätöntä vaikuttavien uusien ratkaisujen löytämiselle. Väitämme, että kompleksisuuslinssin avulla voidaan rakentaa kielellinen silta eri alan tutkijoiden, arvioijien, päättäjien ja kansalaisten välille tavalla, joka auttaa tunnistamaan ratkaisevia käännteitä ja tekee epävarmuudessa ja tietämättömyydessä tarpomisesta paitsi siedettävää, myös hyödyllistä. Kompleksisuusajattelussa maailma näyttää monitulkintaisena kokonaisuutena. Siinä on lukematon määrä toisistaan riippuvaisia alkuarvoja. Se toimii oman historiansa kokemuksista, tulevaisuutensa oletuksista ja nykyhetken tulkinnoista käsin. Asian ydin on tavoittaa, mitä nämä eri tulkinnot ja näkökulmat synnyttävät vuorovaikutuksessa keskenään ja miten ne muokkaavat laajempaa kokonaisuutta.

Kompleksisuuden käsite korostaa, että järjestelmän osatekijät voivat itseorganisoi-
tua ja synnyttää näennäisesti koordinoitua ja adaptiivista käyttäytymistä jopa ilman keskushallintoa tai keskitettyä johtoa (Klijn 2008) ja näin vahvistaa järjestelmän resi-

lienssiä (Duit & Galaz 2008; Duit, Galaz, Eckerberg & Ebbeson 2010). Organisaatioissa resilienssi ei ole vain sitä, että etukäteen varaudutaan tulevaan, vaan ennen kaikkea se on järjestelmän kyky toimia mukautuvasti ja ketterästi yllättävissä tilanteissa. Ei riitä, että johto varautuu tulevaan luomalla vaihtoehtoisia tulevaisuuden skenaarioita. Järjestelmän rakenteiden ja yhteistyön muotojen tulisi lähtökohtaisesti olla resiliентtejä (Hyytiälä 2020). Arvioinnin näkökulmasta ollaan perustavaa laatua olevien kysymysten äärellä. Siirryttäessä arvioimaan systeemisii ominaisuuksia, kuten emergenssiä, epälineaarisuutta ja adaptaatiota, joudutaan väistämättä tarkistamaan monia arvioinneissa usein esiintyviä pinttyneitä oletuksia tai käytettyjä menetelmiä: tulosperustaisuutta, kustannus-hyöty-ajattelua, lineaarista tai syklistä oppimisenäkökulmaa sekä muutosteoriaa ja niin edelleen. (Uusikylä 2019.) Arvioitsijoiden tai arviointitukijoiden kiinnostus kompleksisuutta kohtaan näkyy aihealueen julkaisumäärissä. Google Scholar -haku sanoilla [evaluation&complexity&governance] vuosina 2016–2020 antaa yhteensä 159 000 osumaa. Vaikka mukana onkin tämän artikkelin kannalta vähemmän relevantteja osumia, kertoo se kuitenkin yleisestä kiinnostuksesta aiheeseen arviointitukijoiden keskuudessa. Seuraavassa pyritään lyhyesti hahmottamaan, mistä kompleksisuusteoriassa on kyse ja mitä kompleksisuusajattelun peruskäsitteet voisivat tarkoittaa arvioinnin näkökulmasta.

Kompleksisen vaikuttavuusarvioinnin lähtökohtana on muutoksen dynamiikan hahmottaminen keskinäisriippuvaisten ilmiöiden ja toimijoiden vuorovaikutuspintojen kautta. Tällöin muutos ei ole lineaarista, vaan palautemekanismit limittyvät ilmiöiden rajapintoihin. Samoin muutoksen dynamiikka ei noudata volyyymiin pohjautuvaa vaikutusajattelua vaan spiraalista kehitymispolkua, jossa pienillä toimintakäytäntöihin, mentaalimalleihin ja käyttäytymiseen vaikuttavilla muutoksilla saattaa olla systeemisellä tasolla hyvinkin merkittävä vaikutus. Mallin lähtökohtana on holistinen ajatus siitä, että kokonaisuutta ei voida johtaa osien summasta. Useimmat seuranta- ja indikaattorijärjestelmät (esim. kilpailukykyindeksit, Pisa-mittaukset jne.) kuitenkin perustuvat juuri vanhakantaiseen reduktionistiseen ajatteluun. Ongelmaksi muodostuu kuitenkin se, että panoksia ja tuotoksia/vaikutuksia analysoimalla tai panos-tuotos- tai panos-vaikutus-suhteita tarkastelemalla emme vielä ymmärrä, millaisten mekanismien kautta muunto tapahtuu. Toiminnan kehittämisen ja oppimisen näkökulmasta tämä on kuitenkin välttämätöntä. Mikäli haluamme vapautua vaikuttavuusharhasta ja siitä mahdollisesti seuraavista virhepäätelmistä, meidän on avattava makrotason kausaalinen suhde meso- ja mikrotason muuntomekanismien kautta (Uusikylä ym. 2018; Uusikylä 2019).

Jotta vaikuttavuustavoitteen (esim. yritysten kilpailukyvyyn kasvu, maahanmuuttajien kotoutumisen toteutuminen, kehitysmaiden naisten yhteiskunnallisen aseman parantuminen) alla toteutetuilla toimenpiteillä (lainsäädäntöuudistukset, kokeilut, pilotit, ohjelmat, keskustelunavaukset jne.) voitaisiin saavuttaa haluttuja vaikutuksia tai vaikuttavuutta, tulee toimintaan osallistuneiden organisaatioiden tai yksilöiden si-

säistää ne ja ottaa mallit tai uudet ajatukset osaksi toimintaa. Vaikutusketju voi edetä organisaation kautta (esim. strategiat tai prosessit) yksilötasolle tai ilman organisaation välitystä henkilöltä toiselle. Sisäistetyt innovaatiot tai uudenlainen ajattelutapa eivät kuitenkaan välttämättä vaikuta konkreettisesti, elleivät ne johda ole käyttäytymisen muutokseen. Vasta kun yksilöt alkavat toimia toisin, voi systeeminen muutos levitä. Perinteisesti organisaation toiminnan on katsottu muuttuvan yksilöiden toiminnan muutoksen summana. Tuoreessa organisaatiokirjallisuudessa (Centola 2018) on korostettu, että käyttäytyminen leviää yksilöltä toiselle tehokkaimmin ”tartunnan kautta”, esimerkiksi voimakkaiden mielipidevaikuttajien tai muutosagenttien vaikutuksesta. Tällainen tartunta voi lähteä hyvinkin nopeaan, jopa eksponentiaaliseen kasvuun tietyn kynnyksarvon ylittyessä. Tällöin tapahtuu niin sanottu *tipping point* -ilmiö. Tietyn ajallisen viiveen jälkeen yksilö- ja organisaatiotason toiminnallinen muutos näkyy järjestelmätasoisena muutoksena.

Seuraavassa tarkastelemme keskeisten kompleksisuusteoreettisten käsitteiden yhteyttä arviointien käsitemaailmaan ja arvioinneissa käytettyihin menetelmiin. Tarkastelu on tiivistetty taulukkoon 12.1.

Taulukko 12.1. Kompleksisuusteoreettiset käsitteet arvioinnissa.

Käsite	Määritelmä	Merkitys arvioinnin kannalta
Emergenssi	Emergenssi (ilmaantuminen) viittaa uusien, odottamattomien rakenteiden, mallien tai prosessien syntyymiseen kompleksisessa järjestelmässä. Syntyvää kokonaisuutta ei voida ennustaa sen perusteella, mitä sen osista tiedetään. Näennäisesti asiat vain tapahtuvat ilman mitään erityistä ja tunnistettavissa olevaa syytä. Syntyvä kokonaisuus ei koostu vain osatekijöistä. Syntyvät yksiköt voivat olla vuorovaikutuksessa niiden osien kanssa, joista ne ovat syntyneet.	Emergenssiä voi lähestyä parhaiten formatiivisen arvioinnin lähtökohdista. Siinä arviointi on tiiviissä yhteydessä toiminnan kautta muotoutuvaan kokonaisuuteen. Monet reaaliaikaisen arvioinnin asetelmat pyrkivät myös samaan. Ongelmaksi saattaa muodostua puutteellinen ymmärrys emergenssin tekijöistä ja näiden välisistä interaktioista (vrt. alaspäin suuntautuva kausaatio)
Itseohjautuvuus	Itseohjautuvuus on enemmän tai vähemmän spontaani prosessi ilman ulkoista pakottamista tai hallintaa. Itseohjautuvuutta tapahtuu, kun järjestelmät vaihtavat tietoja, ryhtyvät toimiin ja sopeutuvat jatkuvasti toisten toimiin sen sijaan, että jokin keskitetty taho asettaisi sitovaa suunnitelmaa. Vaikka suunnittelua tai hallintaa ei ole keskitetysti, järjestäytyminen tapahtuu jatkuvasti mukautuen toimintaympäristön parhaimman sopivuuden löytämiseksi.	Perinteinen tavoitteisiin sidottu arviointi ei kykene ottamaan huomioon itseohjautuvuutta, ellei se nimenomaisesti ole ennakkoon määritelty toiminnan tavoite tai päämäärä. Itseohjautuvuuden tunnistaminen edellyttää arvioinnilta adaptiivista lähestymistapaa. Tähän on yhdistettävä myös arvioinnin oma itsereflektio, so. arviointijoiden tulee kaiken aikaa kriittisesti tarkastella oman ymmärryksensä muodostumista ja arvioinnin sekä kohteen välisen dialogin toimivuutta.

Kytkeytyneisyys ja keskinäisriippuvuus	Suurin osa järjestelmistä on sisäkkäin muissa järjestelmissä ja monet järjestelmät ovat pienempien järjestelmien osia. Tavat, joilla järjestelmän toimijat muodostavat yhteyden ja liittyvät toisiinsa, ovat kriittisiä järjestelmän selviytymisen kannalta, koska näistä yhteyksistä muodostuvat toimintamallit ja kaavat sekä takaisinkytkennät.	Arvioinnissa toimijoiden väliset suhteet ja suhderakenteen ovat joskus tärkeämpiä kuin toimijat itse. Kaikki järjestelmät ovat olemassa omassa ympäristössään, ja ne ovat myös osa tätä ympäristöä. Arvioinnin kannalta voi olla vaikeaa erottaa, mikä on ympäristön vaikutus toimintaan ja toiminnan vaikutus ympäristöön koska ne muokkaavat toisiaan samanaikaisesti.
Epälineaarisuus ja palautemekanismit	Epälineaarisuus tarkoittaa, että järjestelmän käyttäytyminen ei välttämättä riipu alkutilojen arvoista. Epälineaarisuus voi aiheuttaa epäsuhtaa syiden ja seurausten välillä – ts. vähäpätöiset syyt tai toimenpiteet voivat saada aikaan merkittäviä vaikutuksia ja päinvastoin. Palaute vaikuttaa järjestelmän käyttäytymiseen. Positiivinen palaute parantaa järjestelmän kykyä (stimulointi), kun taas negatiivisen palautteen vaikutukset ovat päinvastaiset (tasapainottaminen). Positiivinen ei liity ”hyvään” eikä negatiivinen ”huonoon” vaan ne ovat systeemisiiä vaikutuksia.	Tämä haastaa vaikuttavuusarviointien kausaalimallit, muutoksenteoriat ja loogiset viitekehukset. Näissä ns. <i>logic</i> -malleissa pyritään <i>a priori</i> -tiedon pohjalta kuvaamaan vaikutusketjuja. Kompleksisessa arvioinnissa näiden tulisi löytyä aineiston ja sen tulkinnan kautta.
Dissipatiiviset rakenteet ja kaaoksen reuna	Joskus järjestelmässä voi tapahtua merkittävä muutos uuteen vaiheeseen, jota hallitsevat erilaiset attraktorit. Dissipatiiviset rakenteet viittaavat vaihesiirtoon, symmetrian rikkoutumiseen ja useiden vaihtoehtojen avautumiseen. Kompleksinen järjestelmä ajautuu toistuvasti kohti kaaoksen reunaa tai kauemmaksi tasapainotilasta.	Arvioinneissa attraktoreiden ja dissipatiivisten rakenteiden tunnistaminen vaatii järjestelmän herkkää ymmärrystä ja äärimmäisen hienovaraisia menetelmiä, joiden kautta systeemidynaamiset siirtymät voidaan havaita. Useimmiten näiden perustaksi ei ole löydettävissä perinteisessä mielessä vahvaa empiiristä evidenssiperustaa.

Taulukossa 12.1 on kuvattu eräiden kompleksisuusteoreettisten käsitteiden sisältöä ja niiden suhdetta arviointiin. Kompleksisessa arvioinnissa (ks. Byrne 2005; 2013; Patton 2011; Uusikylä 2019) arvioitsija ei voi tukeutua perinteisen, reduktionistisen ohjelma-arvioinnin kausaalimalleihin tai kokeellisiin asetelmiin. Byrne (2013) toteaa, että tutkijoiden tai arvioijien tulisi keskittyä pidemmän ajan muutosanalyysiin ja paikantaa näiden kautta muutoksen dynamiikka ja sen taustalla olevat käyttäytymis- tai toimintakaavat. Donella Meadows (2008, 170) ilmaisee asian vielä tyylikkäämmin:

Before you disturb the system in any way, watch how it behaves. If it's a piece of music or a fluctuation in a commodity price, study its beat. If it's a social system, watch it work. Learn its history.

Byrne ja Callaghan (2014) toteavat, että kompleksisten järjestelmien arvioinnissa kausaalisten vaikutusketjujen sijaan tulisi tarkastella systeemin muutosta ohjaavia sääntöjä ja prosesseja. Taulukon 12.1. mukaisesti kompleksisen arvioinnin tulisi pyrkiä

ottamaan systeemisen muutoksen dynamiikka haltuun kokonaisvaltaisen, pitkäjänteisen ja systeemin murroskohtiin keskittyvän tarkastelun kautta.

Kompleksisten systeemien, prosessien tai rakenteiden arviointi ei hylkää aiemmin sovellettuja arvioinnin kriteereitä, kuten laatu, tehokkuus, tuottavuus tai vaikuttavuus, vaan tuo lisätulkintaa niiden määrittelyyn. Esimerkiksi valtion virasto A voi toteuttaa oman palvelutuotantonsa hyvin laadukkaasti, tehokkaasti tai kustannus-
tehokkaasti, mutta jos se tuottaa palvelunsa irrallisena palveluekosysteemin muusta toiminnasta, se voi omalta osaltaan jopa vaikeuttaa tai heikentää palvelun kokonaisvaikuttavuutta. Tämä on tyypillinen osaoptimointiongelma. Kompleksisessa arvioinnissa korostuu holistinen näkökulma, jonka mukaan kokonaisuus on enemmän kuin osiensa summa. Näin myös arvioinnin tulisi kyetä tarkastelemaan laajempia ilmiökokonaisuuksia ja ymmärtää niiden systeemistä dynamiikkaa, verkostovaikutuksia ja nopeasti muuttuvia toimintatilanteita. Erityisesti nopeasti muuttuvat toimintatilanteet korostavat adaptiivisuutta, jota tarkastellaan seuraavaksi.

Synteesi ja kohti adaptiivista arviointia

Olemme edellä hahmotelleet arviointiajattelun (paradigman) muutoksia kolmen eri vaiheen kautta: 1) siirtymä tilivelvollisuudesta oppimiseen, osallisuuteen ja kehittämiseen, 2) siirtymä osallisuuden ja yhteisen kehittämisen näkökulmasta tulosten hyödyntämiseen ja 3) siirtymä systeemiseen ja kompleksiseen arviointinäkökulmaan. Nämä kaudet eivät luonnollisesti esiinny selvärajaisina vaan jatkuvasti myös päällekkäisinä ja liukuvina. Tilivelvollisuuden muodot liukuvat tiukasta tavoitesidonnaisuudesta kehittävän arvioinnin muutosorientaatioon, osallisuus jalostuu yhteisen merkityksen muodostamisesta yhteiskehittämiseen ja kokeilukulttuuriin, systeemisyys ja kompleksisuus menevät limittäin ja niin edelleen. Tästä huolimatta on mielekästä tarkastella näitä erillisinä näkökulmina arvioinnin kehittämiseen. Taulukossa 12.2 olemme eritelleet kunkin murroksen syitä, ratkaisuja ja uuden näkökulman siirtymistä osaksi arviointikäytäntöjä.

Taulukossa on esitetty arviointinäkökulmien keskeiset siirtymät 1980-luvun lopulta vuoteen 2020. Siirtymien saumakohtia olisi myös mahdollista määritellä uusien mallien, viitekehysten tai näkökulmien lanseerauksen kautta. Tällaisia olisivat esimerkiksi konstruktivistinen arviointinäkökulma, realistinen arviointi, käyttäjälähtöinen arviointi, luottamukseen perustuva arviointi, systeeminen arviointi ja kompleksisten järjestelmien arviointi. Näihin kaikkiin voidaan kiinnittää taulukossa esitettyjä avaintekstejä tai -julkaisuja. Tässä artikkelissa emme kuitenkaan halunneet lähteä näin yksityiskohtaiseen tarkasteluun, vaan olemme nähneet näkökulmien muutoksen pikemminkin trendinomaisena kuin vaiheittaisena kehittymisenä. Näemme

arviointitraditioiden muutoksen merkitsevän eri painotusten ja lähestymistapojen uudelleen muotoutumista, adaptaatiota arviointitarpeen muuttuessa. Siten edellisen murrosvaiheen traditiot eivät katoa, vaan pikemminkin ne sulautuvat toisiinsa ja saavat uusia sisältöjä arviointiympäristöjen muuttuessa. Esimerkiksi kysymykset arviointisuhteen luonteesta tai arviointitarpeen ja tarkoituksen määrittämisestä ovat läsnä arviointiajattelun muuntuessa murroksesta toiseen, mutta ne ilmaistaan siirtymissä eri tavoin. Siirtymät kertovat myös arviointiyhteisöjen itseohjautuvuudesta sekä kyvystä arvioida omaa toimintaansa ja muuntua. Arvioinnin on löydettävä paikansa osana ketteriä organisaatioita, jotta se voisi edistää organisaatioiden ja koko yhteiskunnan kyvykkyyttä muuntautua tulevaisuutta kohden. Haluamme tällä synteesillä luoda pohjaa adaptiiviselle arviointiajattelulle ja siitä käytävälle keskustelulle.

Taulukossa 12.2 on esitetty perusteluja muutoksen tarpeellisuudesta sekä kunkin uuden näkökulman tarjoamia vastauksia tai ratkaisukeinoja havaittuihin ongelmiin. Näkökulmien välille muodostuu mielenkiintoisia jännitteitä, jotka näkyvät myös seuraavissa siirtymissä. Näin jännitteiden kautta syntyy eräänlainen polkuriippuvuus, jonka ymmärtäminen on tärkeää uuden näkökulman käsitteellisessä haltuunotossa. Lisäksi on muistettava, että jokainen vaihe siirtyy näkökulmana myös seuraavaan ja sitä seuraavaan. Esimerkkinä voidaan nostaa osallisuuden ja vaikuttavuuden välinen jännite. Konstruktivistinen arviointimalli lähti kritisoimaan vahvaan kausaaliselittämiseen perustuvia positivistisia arviointiasetelmia ja tilastollisten aineiston tuottamaa ulkopuolista analytiikkaa. Osallisuutta ja voimaannuttamista korostunut ajattelu puolestaan synnytti vahvaa ymmärrystä toiminnan laadullisista prosesseista ja kehittämisen perusteista, mutta ei pystynyt tarjoamaan selkeää näyttöä toiminnan tuloksista ja vaikutuksista. Tämän tilanteen korjaamiseksi vahvistui arvioinnin hyödyntämistä korostava näkökulma ja realistinen arviointimalli. Tätä taustaa vastaan on varsin todennäköistä, että systeeminen ja kompleksinen arviointiajattelu synnyttää jälleen uuden vastavoiman, joka voi olla luonteeltaan yksinkertainen arviointi. Adaptiivisen arvioinnin funktiona voi silloin olla selkiyttäminen, osittaminen ja pilkkominen ymmärryksen lisäämiseksi. Toisaalta resilienssin vahvistamiseksi adaptiivinen arviointi kohdistuu kehittämisprosessien lisäksi ratkaisuihin ja niiden arviointiin.

Taulukko 12.2. Siirtymät arviointiajattelussa.

Arviointi	Miksi?	Miten? (menetelmät ja lähestymistavat)	Tiedon tarve ja hyödyntäminen	Mikä muuttui?	Keskeiset lähteet
I Murros: Tiilivollisuudesta osallisuuteen ja oppimiseen (1980-luvun lopusta 2000-luvun alkuun)	Positiivinen (laajempi ajatus objektiivisesta mittauksesta taustalla) arviointiparadigma murtui samaan aikaan keskitettyjen suunnittelujärjestelmien kaatumisen kanssa. Arvioinnista haluttiin tehdä innostava, osallistava sekä oppimista ja toimintaa kehittävä väline.	– Osallisuuden lisääminen – Yhteinen tulkinta arvioinnin kohteesta – Arvioitsijan rooli oli fasilitoida muutosta – Pyrkimys voimaantumiseen	– Jaettu ymmärrys ja yhteiset merkitykset – Jatkuvan oppimisen päämäärä – Toiminnan kehittäminen – Asianajo ja vaikuttaminen – Muutoksen aikaansaaminen	+ Ulkoisen arvioinnin ohelle syntyi vahva itsearvioinnin kulttuuri + Arviointi ja kehittäminen saatettiin yhteen – Ongelmana tulosten hyödynnettävyys (relativismi) – Arviointien näyttöperusta jäi epämääräiseksi – Arviointisuhde kontrollista enemmän luottamukseen pohjautuvaksi	Guba & Lincoln 1989); Fetterman 2001; House 1993; 1997; Rossi, Lipsey & Freeman 2004; Patton 1997; Patton, McKegg & Wehipeihana 2016; Räkköläinen 2011 Patton, 1994.
II Murros: Osallistavasta ja voimaannuttavasta arvioinnista tuloksia hyödyntävään arviointiin (2000-luvun alusta 2010-luvun puoliväliin)	Arvioinnin hyödynnettävyys nousi keskiöön. Arviointien yleistyessä ja institutionalisoituessa peräänkuulutettiin niiden hyötyä ja lisäarvoa. Etenkin OECD, Euroopan komissio ja Maailmanpankki ajoivat vahvaa tulosperustaisuutta ja selkeitä arviointikriteereitä.	– Perustettiin arviointiyksiköitä ja julkaistiin arviointimateriaaleja – Käynnistyi keskustelu arviointiosaamisesta, standardeista ja arvioinnin laadusta – Logic-mallit ja muutoksen teorian oppiminen	– Arviointi kytkettiin tulossuunnitelmaan ja budjetoitiin – Arviointi kytkettiin laajempaan keskusteluun päätöksenteon tietoperustasta – Hyöty ja käytettävyys korostuvat	+ Arviointien vertailtavuus parani (yhteiset kriteerit) + Arviointi-osaaminen lisääntyi – Mekanistisuus ja ritualismi kasvoivat	Patton 1994 EVALSED 2008; OECD 2002 Pawson & Tilley 1997; Pawson 2006)
III Murros: Osallisuudesta systeemisyyteen ja kompleksisuuteen (2010-luvun puolivälistä eteenpäin)	Taustalla oli kasvava ymmärrys siitä, että arvioinnin tulee ymmärtää kohteensa laajemmassa kontekstissa (vrt. realistinen arviointi) ja osana muita systeemejä. Pyrkimys irtautua kapeasta ohjelma-, sektori- ja instituutioperusteisesta tarkastelusta.	– Systeemisen dynamiikan, vuorovaikutusrakenteiden ja polku- ja keskinäisriippuvuuksien arviointi – Irtautuminen lineaarisista syy-seuraus-suhteista – Palautemekanismit ja takaisinkytkennät. – Monitahaisen arviointinäkökulman korostaminen	– Dynaaminen ymmärrys arvioinnin kohteesta – Kehittämisen suunnittaminen oikeisiin vipukohtiin. – Adaptiivisuus	+ Holistinen käsitys arvioinnin kohteesta (ilmiölähtöisyys) + Systeemitieteiden hyödyntäminen mallinnuksessa ja arvioinnissa – Tulosten yleistettävyyden ja kommunikoinnin ongelma – Systeemisen ymmärryksen saavuttamisen vaikeus	Cabrera ym, 2008; Dyehouse ym. 2009; Mayne, 2012; Funnel & Rogers 2011; Patton, 2011; Williams & Hummelbrunner 2011; Hyytinen 2017
	Pyrkimys jatkaa systeemisen arvioinnin kehitystyötä ja soveltaa opittua yhteiskunnan pirullisiin ongelmiin ja kompleksisiin ilmiökokonaisuuksiin. Painottaa voimakkaasti emergenssin ja keskinäisriippuvuuksien merkitystä.	– Kompleksisen systeemin perusominaisuuksien ymmärtäminen – Emergoituvien prosessien tunnistaminen ja arviointi – Sensemaking ja tilannekuvan aikaansaaminen	– Yhteisten tulkintakehikoiden hahmottaminen – Dialogisuus ja yhteinen tilanteen määrittely – Vuorovaikutus-rakenteiden analysointi – Analyysistä synteeseihin	+ Arviointien suunnittaminen pirullisiin ongelmiin ja kompleksisiin ilmiöihin + Uudenlaisen käsitteistön lanseeraaminen – Vaikeaselkoisuus ja kommunikaatio-ongelmat – Vaikeus tarjota vahvaa näyttöä	Pawson 2013; Byrne 2013; Patton 2011; Uusikylä 2019; Ramalingam 2013; Frej & Ramalingam 2011; Forss, K. & Schwartz, R. 2011

Arviointi on viime kädessä aina sidoksissa arvioinnin kohteeseen. Esimerkiksi ohjelma-arvioinnissa ohjelman rakenne-, logiikka ja ”ideologia” asettavat voimakkaita reunaehdoja arvioinnille. Adaptiiviset ohjaus- ja ohjelmointiperiaatteet pyrkivät hyö-

dyntämään kompleksisuusteoriaa ja systeemiajattelua. Ne pyrkivät ymmärtämään, miten eri elementtien väliset yhteydet vaikuttavat toisiinsa ja miten ne tuottavat muutosta (Bain ym. 2016, 23). Toisin kuin perinteisessä, lineaarisessa ohjelma-ajattelussa, adaptiivinen ohjelmointi tarkoittaa ohjelmien suunnittelua ja toteuttamista tekemällä oppimisen kautta, jatkuvasti tutkimalla, kokeilemalla ja testaamalla ja mukauttamalla erilaisia lähestymistapoja koko ohjelmaprosessin ajan, jolloin paikalliset kumppanit ja sidosryhmät voivat tunnistaa ongelmat ja työskennellä yhdessä ratkaisujen kehittämiseksi (Derbyshire & Donovan 2016, 29). Adaptiivinen ohjelmointi ja arviointi korostavat jatkuvaa oppimista, kokeilukulttuuria ja emergenssiä. Näin sekä ohjelmatyötä että arviointia tehdään kentällä, joka on usein monimutkainen, arvaamaton ja nopeasti muuttuva (Ramalingam 2013; Frej & Ramalingam 2011).

Adaptiivinen arviointi on suhteellisen uusi käsite. Eoyang ja Oakden (2016) ovat korostaneet, että arvioitsijoiden tulee herkästi tunnistaa monimutkaisten ohjelmien mallit ja ymmärtää niiden dynamiikka ja interventiologiikka sidosryhmille merkityksellisillä tavoilla. Tällaisessa työotteessa synergia arvioinnin ja ohjelmatyön välillä syntyy kompleksisuusteorian ja adaptiivisen ohjelmaymmärryksen välisen vuoropuhelun kautta. Wallenius, Uusikylä ja Tran Minh (2019) sovelsivat adaptiivisen ohjelmatyön ajatusta samoin kuin sen tueksi kehitettyä systeemistä arviointimallia Suomen ja Vietnamin välisen innovaatiokumppanuusohjelman (IPP) loppuarvioinnissa. Tätä työtä varten he joutuivat laatimaan ”kaksoisarviointiaselman” eli suunnittelemaan ja toteuttamaan sekä perinteisen ulkoministeriön arviointikriteereitä vastaavan arvioinnin että sen rinnalla toteutettavan systeemisen arvioinnin. Arvioinnin yhtenä tuloksena oli, että IPP-ohjelman myönteiset vaikutukset perustuivat suurelta osin ohjelman adaptiiviseen toteuttamiseen, mikä oli perusteltua uutta luovassa hankkeessa, jossa pyrittiin luomaan uudenlaista start-up-kulttuuria Vietnamiin. Samalla myös suuri osa ohjelman tuloksista oli ei-aiottuja positiivisia vaikutuksia.

Arviointien suunnittelussa keskeistä on arvioinnin kohteen ja sen erityispiirteiden ymmärtäminen. Yksinkertaiseen tai vain vähän monimutkaiseen kohteeseen soveltuvat erilaiset arviointiaselmat ja -menetelmät kuin kompleksiseen tai kaoottiseen. Raisio ym. (2018) sovelsivat Sitralle tehdyssä ”Kesy, sotkuinen vai pirullinen” -raportissaan niin sanottua kompleksisuuskuutiota analysoidessaan Sipilän hallituksen hallitusohjelma kärkihankkeita. Tutkijat havaitsivat, että riippumatta siitä, ovatko hankkeet kesyjä vai pirullisia, niitä valmistellaan, toteutetaan, seurataan ja arvioidaan samalla tavalla.

Arviointi maailmaa pelastamassa – mitä seuraavaksi?

Viimeaikaiset globaalit kriisit ja uhat (esim. Syyrian sota, terrorismi), maailmantaloutta horjuttaneet pandemiat (COVID-19) ja maapallon tila sekä siihen liittyvät YK:n kestävän kehityksen tavoitteet (ns. SDG-tavoitteet) ovat vaikuttaneet myös arvioinnista käytävään keskusteluun. Osaltaan systeemi- ja kompleksisuusnäkökulmien korostuminen ovat seurausta juuri tästä. Arvioitsijoiden odotetaan kykenevän paremmin selvittämään kompleksisten systeemien vaikutuksia ja esittävän arvioita siitä, millaisten politiikka- tai sääntelytoimien kautta näitä voitaisiin hallita.

Michael Q. Pattonin uusiin teos ”*Blue Marble Evaluation*” (Patton 2020) haastaa kuitenkin vielä voimakkaammin arviointia ja arviointiyhteisöjä suuntautumaan ihmiskunnan kohtalon ja tulevaisuuden kannalta suuriin kysymyksiin. Pattonin mukaan arvioijat tarvitsevat uudenlaisia näkökulmia ja parempaa osaamista voidakseen osallistua globaalien ongelmien ratkaisemiseen. Hänen mukaansa ”sinisen planeetan evaluaattoreiden” tulisi arvioida, seurata, parantaa ja tukea globaaleja muutospyrkimyksiä ja lopulta kyetä arvioimaan näiden globaalien muutosponnistelujen vaikuttavuutta, tehokkuutta, merkityksellisyyttä ja kestävyyttä. Tehtävään sisältyy paljon enemmän kuin tulosindikaattoreiden ja kestävän kehityksen tavoitteiden seuraaminen. Pattonin (2020) mukaan tämä tarkoittaa koko arviointiajattelun, sen työkalujen, menetelmien ja prosessien tuomista globaalille tasolle – ja uusien lähestymistapojen kehittämistä maailmanlaajuisten järjestelmien arvioinnin tarpeisiin.

Vaikka hahmottaisi hieman vaatimattomampiakin tavoitteita tulevaisuuden arvioitsijoille on selvää, että nykyisellä arviointiajattelulla käytössä olevilla menetelmillä ja aineistoilla ei kyetä arvioimaan kompleksisen, vahvasti keskinäisriippuvaisen ja äärimmäisen nopeasti muuttuvan maailman vaikeita ongelmia. Tarvitaan jännitteiden kautta kehittyvää uutta arviointiajattelua. Arvioitsijoiden tulee rohkeasti osallistua globaaliin keskusteluun, haastaa mekanistisia tulkintoja ja reduktionistisia arviointimalleja, tarjota uusia näkökulmia ja viitekehyksiä sekä toimia asiantuntevina välittäjinä ja vuoropuhelun synnyttäjinä kansalaisten, päätöksentekijöiden ja muiden globaalia muutosta suuntaavien toimijoiden kesken.

Lähteet

- Anttila, P. 2005. *Ilmaisu, teos, tekeminen ja tutkiva toiminta*. Hamina: Akatiimi oy.
- Bale, C. S. E., Varga, L. & Foxon, T. J. 2015. Energy and complexity: New ways forward. *Applied Energy* 138, 150–159.
- Bester, A. 2016. Results Based Management in the United Nations Development System. A report prepared for United Nations Department of Economic and Social Affairs for the 2016 Quadrennial Comprehensive Policy Review. <https://www.un.org/>
- Byrne, D. 2005. Complexity, Configurations and Cases. *Theory, Culture and Society* 22(5), 95–111.
- Byrne, D. 2013. Evaluating complex social interventions in a complex world. *Evaluation* 19(3), 217–228.
- Byrne, D. & Callaghan, G. 2014. *Complexity Theory and Social Sciences. The State of the Art*. Routledge, New York.
- Cabrera, D., Colosi, L. & Lobdell, C. 2008. "Systems thinking". *Evaluation and Program Planning* 31(2), 299–310.
- Centola, D. 2018. *How Behavior Spreads. The Science of Complex Contagions*. Princeton University Press, Princeton.
- Chelimsky, E. 1997. The Coming Transformation in Evaluation. Teoksessa Chelimsky, E. & Shadish, W. (toim.) *Evaluation for 21st century. A handbook*. Thousand Oaks: Sage, 1–26.
- Chelimsky, E. & Shadish, W. R. 1997. *Evaluation for the 21st century: A handbook*. Thousand Oaks: Sage.
- Chen, H. T. 2005. Theory driven evaluation. Teoksessa Mathison, S. (toim.) *Encyclopedia of evaluation*. Thousand Oaks: Sage, 415–419.
- Crooks, T. J. 2002. Educational Assessment in New Zealand Schools. Profiles of Educational Assessment Systems World-Wide. *Assessment in Education* 9(2), 237–253.
- Derbyshire, H. & Donovan, E. 2016. Adaptive programming in practice: shared lessons from the DFID-funded LASER and SAVI. <https://beamexchange.org/resources/798/programmes>
- Djellal, F. & Gallouj, F. 2010. The Innovation gap and the performance gap in the service economies: a problem for public policy. Teoksessa Gallouj, F. & Djellal, F. (toim.) *The Handbook of Innovation in Services. A Multi-disciplinary Perspective*. Cheltenham, UK: Edward Elgar, 653–673.
- Djellal, F. & Gallouj, F. 2013a. The Productivity in services: measurement and strategic perspectives. *The Service Industries Journal* 33(3–4), 282–299.
- Djellal, F. & Gallouj, F. 2013b. How public-private innovation networks in services (ServPPINs) differ from other innovation networks: what lesson for theory? Teoksessa Gallouj, F., Rubalcaba, L. & Windrum, P. (toim.) *Public-private innovation networks in services*. Cheltenham, UK: Edward Elgar.
- Dosi, G., Freeman C., Nelson R., Silverberg, G. & Soete, L. 1988. *Technical change and economic theory*. London: Pinter Publishers.
- Duit, A. & Galaz, V. 2008. Governance and Complexity: Emerging Issues for Governance Theory. *Governance* 21(3), 311–335.
- Duit, A., Galaz, V., Eckerberg, K. & Ebbesson, J. 2010. Governance, Complexity, and Resilience. *Global Environmental Change* 20(3), 363–8.
- Dyehouse, M., Bennett, D., Harbor, J., Childress, A. & Dark, M. 2009. A comparison of linear and systems thinking approaches for program evaluation illustrated using Indiana Interdisciplinary GK-12. *Evaluation and Program Planning* 32, 187–196.

- Eoyang, G. & Oakden, J. 2016. Adaptive Evaluation: A synergy between complexity theory and evaluation practice. *Emergence: Complexity and Organization* 18(3-4). doi: 10.emerg/10.17357.e5389f5715a734817dfbeaf25ab335e5.
- EVALSED 2008. *The Evaluation of Socio-economic Development*. The Guide. Brussels: European Commission.
- Fetterman, D. 2001. *Foundations of empowerment evaluation*. London: Sage.
- Frej, W. & Ramalingam, B. 2011. Foreign policy and complex adaptive systems: Exploring new paradigms for analysis and action, SFI Working Paper, Sante Fe: SFI. <http://www.santafe.edu/media/workingpapers/11-06-022.pdf>
- Forss, K. & Schwartz, R. 2011. Introduction. Teoksessa Forss, K., Marra, M. & Schwartz, R. (toim.) *Evaluating the Complex. Attribution, Contribution, and Beyond* Volume 18. London: Transaction Publishers.
- Gadrey, J. 2010. The environmental crisis and the economics of services: the need for revolution. Teoksessa Gallouj, F. & Djellal, F. (toim.) *The Handbook of Innovation in Services. A Multi-disciplinary Perspective*. Cheltenham, UK: Edward Elgar, 653–673.
- Gardner, J. 2006. Assessment for Learning: A Compelling Conceptualization. Teoksessa Gardner, J. (toim.) *Assessment and Learning*. London: Sage, 197–204.
- Geels, F. W. 2002. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy* 31, 1257–1274.
- Geels, F. W. 2004. From sectoral systems of innovation to socio-technical systems. Insights about dynamics and change from sociology and institutional theory. *Research Policy* 33, 897–920.
- Guba, G. E. (toim.) 1990. *The Paradigm Dialog*. London: Sage.
- Guba, G. E. & Lincoln, Y.S. 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage.
- Harrison, D., Klein, J.-L. & Browne, P. L. 2010. Social innovation, social enterprise and services. Teoksessa Gallouj, F. & Djellal, F. (toim.) *The Handbook of Innovation and Services*. Cheltenham and Northampton: Edward Elgar, 197–281.
- Hargreaves, M. B. & Podems, D. 2012. Advancing Systems Thinking in Evaluation: A Review of Four Publications. *American Journal of Evaluation* 33, 462–470.
- Husen, T. 1999. Research Paradigms in Education. Teoksessa Keeves, J.P. & Lakomski, G. (toim.) *Issues in Educational Research*. Amsterdam: Pergamon, 31–39.
- Hyytinen, K. 2017. Supporting service innovation via evaluation: a future oriented, systemic and multi-actor approach. Doctoral dissertation 14/2017, Aalto University publication series. VTT Science 146. <http://www.vtt.fi/inf/pdf/science/2017/S146.pdf>
- Hyytinen, K., Saari, E. & Elg, M. 2019. Human-Centered Co-Evaluation Method as a Means for Sustainable Service Innovations. Teoksessa Toivonen, M. & Saari, E. (toim.) *Human-Centered Digitalization and Services 2019*. <https://doi.org/10.1007/978-981-13-7725-9>
- Hyytiälä, H. 2020. Tehokkuuden vaalimisesta oppimiseen ja sopeutumiseen. Blogi 6.8.2020 Vaasan yliopiston Kompleksisuustutkimuksen ryhmän sivuilla. <https://blogs.univaasa.fi/kompleksisuus/2020/08/06/tehokkuuden-vaalimisesta-oppimiseen-ja-sopeutumiseen/>
- Ison, R. & Straw, E. 2020. *The Hidden Power of Systems Thinking: Governance in a Climate Emergency*. New York: Routledge.
- Klijin, E.-H. 2008. Complexity Theory and Public Administration: What's New? *Public Management Review* 10(3), 299–317.
- Kline, S. J. & Rosenberg, N. 1986. An Overview of Innovation. Teoksessa Landau, R. & Rosenberg, N. (toim.) *The Positive Sum Strategy – Harnessing Technology for Economic Growth*. Washington D.C.: National Academy Press.
- Kushner, S. 2005. The Limits of Constructivism in Evaluation. Teoksessa Stern, E. (toim.) *Evaluation Research Methods. Volume I. Sage Benchmarks in Social Research Methods*. London: Sage, 284–289.

- Lundvall, B.-Å. 1992. *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London and New York: Pinter Publishers.
- Lähteenmäki-Smith, K. & Virtanen, P. 2020. Societal Phenomena, Mission-oriented Public Policy and the New Evaluation Culture. Teoksessa Lehtimäki, H., Uusikylä, P. & Smedlund, A. (toim.) *Society as an Interaction Space. A Systemic Approach*. Translational Systems Sciences. Singapore: Springer, 69–88.
- Mayne, J. 2008. Results management: Can results evidence gain a foothold in the public sector? Teoksessa Reiper, O., Leeuw, F. & Ling, T. (toim.) *The Evidence Book*. New York: Transaction Publishers.
- Mayne, J. 2011. Addressing cause and effect in simple and complex settings through contribution analysis. Teoksessa Schwartz, R., Forss, K. & Marra, M. (toim.) *Evaluating the Complex*. New York: Transaction Publishers.
- Mayne, J. 2012. Contribution analysis: Coming of age? *Evaluation* 18, 270–280.
- Meadows, D. H. 2008. *Thinking in Systems: A Primer*. White River Junction, Vermont: Chelsea Green Publishing.
- Ministry for Foreign Affairs of Finland. 2019. "How do we Learn, Manage and make decisions in Finland's Development Policy and Cooperation?" Evaluation on Finland's Development Policy and Cooperation. Ministry for Foreign Affairs of Finland 2019/3. Helsinki: Ministry for Foreign Affairs of Finland.
- Nelson, S. & Winter, R. 1982. *An evolutionary theory of economic change*. Cambridge (Mass.): Belknap Press/Harvard University Press.
- Nieminen, M. & Hyytinen, K. 2015. Future-oriented impact assessment: Supporting strategic decision-making in complex socio-technical environment. *Evaluation* 21(4), 448–461.
- Nieminen, M. Hyytinen, K., Salminen, V. & Ruutu, S. 2020. Systemic Evaluation Approach to meet the Challenges of Complexity. Teoksessa Lehtimäki, H., Uusikylä P. & Smedlund A. (toim.) *Society as an Interaction Space, Translational Systems Sciences* 22. Singapore: Springer. https://doi.org/10.1007/978-981-15-0069-5_5
- OECD 2014. Measuring and managing results in development cooperation. A review of challenges and practices among DAC members and observers, OECD Publishing Paris. <https://www.oecd.org/dac/peer-reviews/Measuring-and-managing-results.pdf> (12.4.2021)
- OECD 2016. Providers' use of results information for accountability, communication, direction and learning. Survey results, OECD publishing Paris. https://www.oecd.org/dac/peer-reviews/Providers'_use_of_results_information_for_accountability_communication_direction_and_learning.pdf (12.4.2021)
- O'Neil, O. 2002. *A Question of Trust*. The BBC Reith Lectures 2002. Cambridge: Cambridge University Press.
- Patton, M. Q. 1994. Developmental Evaluation. *Evaluation Practice* 15(3), 311–319.
- Patton, M. Q. 1997. *Utilization-focused Evaluation*. Newbury Park, CA: Sage.
- Patton, M. Q. 2002. *Qualitative Research and Evaluation Methods*. 3. painos. London: Sage.
- Patton, M. Q. 2011. *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York, NY: Guilford.
- Patton, M. Q. 2020. *The Blue Marble Evaluation. Premises and Principles*. New York, NY: Guilford.
- Patton, M. Q., McKegg, K. & Wehipeihanan, N. (toim.) 2016. *Developmental Evaluation Exemplars. Principles in practice*. London: The Guilford Press.
- Pawson, R. 2006. *Evidence-based Policy. A Realistic Perspective*. London: Sage.
- Pawson, R. 2013. *The Science of Evaluation. A Realist Manifesto*. Los Angeles: Sage.
- Pawson, R. & Tilley, N. 1997. *Realistic evaluation*. Thousand Oaks: Sage.

- Poikela, E. 2004. Developing criteria for Knowing and Learning at Work: Towards Context-Based Assessment. *Journal of Workplace Learning* 16(5), 267–274.
- Poikela, E. & Rökköläinen, M. 2006. ”Intelligent Accountability” - kontekstiperusteisen arvioinnin lähtökohtia. *Ammattikasvatuksen aikakauskirja* 8(2), 6–18.
- Raisio, H., Jalonen, H. & Uusikylä, P. 2018. *Kesy, sotkuinen vai pirullinen ongelma. Tiedon käyttö yhteiskunnallisessa päätöksenteossa*. Sitran selvityksiä 139. Marraskuu 2018. Helsinki: Sitra.
- Rajavaara, M. 2007. *Vaikuttavuusyhteiskunta. Sosiaalisten olojen arvostelusta vaikutusten todentamiseen*. Sosiaali- ja terveystieteiden tutkimuksia 84. Helsinki: Kela.
- Rossi, P., Freeman, H. & Lipsey, M. 1999. *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage Publications.
- Rubalcaba, L., Michel, S., Sundbo, J., Brown, S.W. & Reynoso, J. 2012. Shaping, Organizing and Rethinking Service innovation: A multidimensional Framework. *Journal of Service Management* 23(5), 696–715.
- Räisänen, A. & Rökköläinen, M. 2014. Developmental Assessment of Learning Outcomes. Teoksessa Kalliola, S. (toim.) *Evaluation as a Tool for Research, Learning and Making Things Better*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Rökköläinen, M. 2011. *Mitä näytöt näyttävät? Luotettavuus ja luottamus ammatillisten perustutkintojen näyttöperusteisessa arviointiprosessissa*. Acta Universitatis Tamperensis 1636. Väitöskirja. Tampere: Tampere University Press.
- Rökköläinen, M. & Räisänen, A. 2013. Assessment of Learning Outcomes in Finnish vocational education and training. *Assessment in Education: Principles, Policy & Practice* 21(1), 109–124. DOI: 10.1080/0969594X.2013.838938.
- Saari, S. 2002. *Opettajankoulutuksen arviointi- ja kehittämisdiskurssi koulutuspoliittisessa kontekstissa*. Acta Universitatis Tamperensis 893. Tampere: Tampere University Press.
- Saunders, M. 2006. Do we Hear the Voices? The ”Presence” of Evaluation Theory and Practice in Social Development. Evaluation. *The International Journal of Theory, Research and Practice* 12(2), 251–264.
- Stame, N. 2004a. Theory-based Evaluation and Types of Complexity. Evaluation. *The International Journal of Theory, Research and Practice* 10 (1), 58–76.
- Stame, N. 2004b. Governance, Democracy and Evaluation. Evaluation. *The International Journal of Theory, Research and practice* 12(1), 7–16.
- Stern, E. (toim.) 2005. *Evaluation Research Methods. Volume I–IV*. Sage Benchmarks in Social Research Methods. London: Sage.
- Stroh, D. P. 2015. *Systems Thinking for Social Science*. Vermont: Chelsea Green Publishing.
- Todd, P. E. & Wolpin, K.I. 2010. Structural Estimation and Policy Evaluation in Developing Countries. *Annual Review of Economics* 2, 21–50.
- USAID. 2017. Evidence base for collaboration, learning and adapting – Summary of literature review. <https://usaidlearninglab.org>
- Uusikylä, P. 2019. *Endeavour to Find Evidence. The role evaluation in the complex systems of governance*. Academic dissertation. Helsinki: University of Helsinki.
- Vedung, E. 1997. *Public Policy and Program Evaluation*. New Brunswick, NJ: Transaction Publishers.
- Vähämäki, J. 2017. *Matrixing aid: the rise and fall of ”results initiatives” in Swedish development aid*. Doctoral dissertation. Stockholm Business School, Stockholm University.
- Vähämäki, J. & Verger, C. 2019. Learning from results-based management evaluations and reviews. OECD Development Co-operation working paper 53. March 2019. DOI: 10.1787/3fda0081-en
- Vähämäki, J., Schmidt, M. & Molander, J. 2011. *Results-based management in development cooperation*. Stockholm: Riksbankens Jubileumsfond.

- Wallenius, T., Uusikylä, P. & Tran Minh, A. T. 2019. Final Evaluation of the Innovation Partnership Programme (Finland-Vietnam), Phase II. Ministry for Foreign Affairs.
- Weijermars, W. & Wesemann, P. 2013. Road safety forecasting and ex-ante evaluation of policy in the Netherlands. *Transportation Research*, Part A, Policy and Practice 52, 64–72.