

Arvioinnin muutokset ja tulevaisuusnäkymät opetus- ja kasvatustieteen arvioijien puhumina



Päivi Atjonen

TIIVISTELMÄ

Artikkelissa tarkastellaan kokeneiden opetus- ja kasvatustieteen arvioijien näkemyksiä arvioinnin muutoksista ja tulevaisuudesta. Harkinnanvarainen näyte käsitti 15 arvioijaa, joita haasteltiin tila-, teema- ja ohjelma-arviointikokeuksista. Litteroidusta haastatteluaineistosta muodostettiin lähilukumenetelmällä muutos- ja tulevaisuuskuvausten aineisto (11 106 sanaa), josta tulokset tematisoitiin. Muutokset olivat: lisääntyneen arvioinnin lisääntyneet ongelmat, kehittämisen ja tulostuun rinnakkaisuus, kehittämisen ja prosessien korostaminen ja kontekstuaaliset muutokset (voimavarat, arviointiosaaminen, kilpailu). Tulevaisuusodotukset myötäilivät muutoksia korostaen arvioinnin kehittävää käyttöä sekä arvioinnin aiempaa merkittävämpää roolia yhteiskunnassa (mukaan lukien toimijoiden polarisoituminen ja pelko tulostuun tiukkenemisestä). Tulokset antavat opetus- ja kasvatustieteen arvioinnista yleismerkityksistä tietoa.

Avainsanat: arviointi, tila-, teema- ja ohjelma-arvioijat; muutos, tulevaisuus

ABSTRACT

Evaluation trends and its future perceived by educational evaluators

This article focuses on evaluation trends and its expected future perceived by experienced educational evaluators. Altogether 15 evaluators (discretionary sample) were interviewed from the context of situational, thematic and program evaluation. By means of close reading the transcriptions on changes and future were extracted from the total text data, ending up to 11 106 words. Main themes of trend statements were qualitatively analysed and thematised as follows: increased problems of increased evaluation, parallel trends of development and accountability, emphasis on development and processes, and perceived contextual chances (resources, evaluation capacity, competition). Expectations for the future resembled the main previous trends preferring developmental use of evaluation and evaluation's increased role in society (including polarisation of agencies and fear for tightened accountability). Results indicate meaningful knowledge on evaluation in the educational field.

Key words: evaluation; situational, thematic and program evaluators; changes, future

ALUKSI

Arviointi lisääntyy nyky-yhteiskunnassa voimakkaasti yhdistyen enenevästi mielikuvien tuottamiseen dynaamisista toimijoista, jotka kehittävät osaamistaan ajantasaisen tiedon perusteella. Arviointiaktiivisuuden osoittamisesta on tullut kireä kilpajuoksu etenkin kaupallisilla aloilla, mutta se yleistyy myös julkisella sektorilla opetus- ja koulutuskenttä mukaan lukien. Arvioinnista on modernisaation myötä kehkeytnyt oma toimialansa, jossa on spesifioituvia menetelmiä, strategioita, välineitä, toimijoita, tutkimusta ja rahavirtoja. Alalle on syntynyt, kuollut ja juurtunut monenlaisia koulukuntia eli käsityksiä siitä, millaista on hyvä arviointi. Yksinkertainen pohdiskelu ”hyvästä ja pahasta, oikeasta ja väärästä” on laventunut ajoittain kiiperiksi kiistoiksi. Arvioinnin historia ei siis ole ollut millään muotoa rauhallinen.

Opetus- ja kasvatustieteen arviointi on kaikkien muiden yhteiskuntasektoreiden tavoin sidoksissa aikaan ja kulttuuriin. Sata vuotta sitten hyvänä pidetyllä arvioinnilla ei olisi tänään mitään käyttöä, eikä suomalaisten arviointikäytänteiden vienti vaikkapa Aasiaan tai arabimaihin onnistu ainakaan lyhyessä ajassa. Arvioinnin pitää muuttua ajassa ollakseen toimiva yhdesäkään päätehtävistään eli kehittämisessä, tilivelvollisuuden täyttämässä tai tiedon tuottamisessa. Jotkut muutospaineet ovat nopeita ja vahvoja, jotkut ilmenevät hitaasti ja todellistuvat hapuillen. Joitakin arviointitapoja voi pitää jo tapahtunutta ”tuomaroivana”, palaute-sanan mukaisesti *feedback*-katsantoisena, vaikka kaikille aloille tarpeellisempaa olisi suunnata arviointia *feedforward*-hengessä tulevan kehityksen edistämiseen.

Tarkastelen artikkelissani haastatteluaineiston avulla arviointia työkseen tekevien henkilöiden tunnistamia arviointimuutoksia ja -tulevaisuutta. Haastateltavien työkokemus oli pääasiassa tila-, teema- ja ohjelma-arvioinneista, joten oppimistuloksien arviointia ei käsitellä lainkaan seuraavissa luvuissa. Suomessa ei juuri ole empiiristä tutkimusta opetus- ja kasvatustieteen arviointia työkseen tekevästä. Meillä ja osin myös kansainvälisessä kirjallisuudessa kuvataan arviointituloksia tai pohditaan käsitteellisesti sen edellytyksiä, mutta arvioitavien (esim. oppi-

laiden ja opiskelijoiden) kohteiden näkemyksiä tai arvioinnin tekijöiden ääntä on kuultu vähän.

ARVIOINNIN MUUTOSTRENDEJÄ

Arvioinnin historiallisissa tarkasteluissa on useimmin nojattu Guban ja Lincolnin (1989) teoksen sukupolvierittelyyn. Jäsenmän kehitystä nojautuen tuoreempaan lähteeseen eli ruotsalaisen Vedungin (2010) artikkeliin, joka avaa lähivuosisikymmeniin varsin toisenlaisen näkymän kuin teos ”Fourth generation evaluation” (ks. myös Basson 2009; Kushner 2011; Virtanen 2005). Vedung kuvaa Pohjoismaiden ja pohjois-atlanttisten valtioiden (Alankomaat, Britannia, Kanada, Suomi, Tanska, USA) arvioinnin muutoksia aalto-metaforan avulla. Vaikka Vedungin katsaus ei kohdennu spesifiin toimialaan, se on hyvin uskottava myös tämän artikkelin keskiössä olevan suomalaisen opetus- ja kasvatustieteen kannalta. Viimeksi mainitun arviointituntemuksessa voin nojata paitsi omaan pitkäaikaiseen työ- ja tutkimuskokemukseeni niin myös Simolan ja Rinteen tutkimusryhmien kriittisiin koulutusanalyysiin.

Tieteellinen aalto (science-oriented wave) hallitsi 1950–1960-lukuja ja oli erityisen kiinnostunut esimerkiksi ohjelma-arvioinnissa kvantitatiivisen kokeellisen asetelman avulla osoitettavasta vaikuttavuudesta. Kun yhteiskuntatieteellisen alan arvioinneissa toistuvasti todettiin vaikeaksi saada interventioiden vaikutuksia näkyviin tällaisin asetelmin, ja arvioitavat esineellistettiin arvioinnin kohteiksi, dialoginen arviointiaalto (dialogical wave) alkoi voimistua. Esimerkiksi hankkeiden suunnittelemattomia mutta hyödyllisiä sivuvaikutuksia alettiin arvostaa, ja ihmiskäsitys muuttui näkemään arvioitavien oman toimintansa asiantuntijoina. Laadullisten tiedonhankintatekniikoiden avulla pyrittiin saamaan sidos- ja marginaaliryhmien ääni monipuolisesti kuuluville ja ymmärrettiin arviointitiedon ”totuuden” olevan sosiaalisesti rakentunutta, ei absoluuttista.

Uusliberalistinen arviointiaalto (neoliberal wave) haastoi 1980-luvulta alkaen dialogisen arviointiaallon, joka kohtasi vaikeuksia todentaa vaikutuksia tai vaikuttavuutta ja jonka emansipatoriset tai arvioitavia osallistavat odotukset eivät toteutuneet toivotulla tavalla. Jo 1980-

luvulla alettiin puhua enenevästi tehokkuudesta, tuloksellisuudesta ja taloudellisuudesta, ja arvioinnin retoriikkaan pesiytyivät tulosvas- tuun, suoritteiden ja asiakastyytyväisyyden ter- mit. Nykyään eletään Vedungin mukaan tietoon perustuvan arvioinnin (evidence-based wave) aikaa, jota hän kutsuu eräänlaiseksi kokeellisuuden paluuksi ja joka perustuu ajatukseen ”what matters is what works and what works is called evidence”. Arvioitavien tulee kyetä osoittamaan toiminnastaan mieluiten numeerisia näyttö- jä, ja puhe vaikuttavuudesta (impact) – ei vain tuloksista, tuotoksista tai vaikutuksista (results, output, outcome) – on voimistunut (myös OECD 2013, 18–19; Santiago 2013, 41–43).

Vedung (2010, 265) muistuttaa, että jokainen aalto jättää ”valtioiden rantaan” jälkiä, joita seuraava aalto ei huuhto kokonaan pois. Sedimen- toitumisen vuoksi samanaikaisesti voi olla useita, jopa toisilleen vastakkaisia arviointi-ideolo- gioita ja -käytänteitä. Esimerkiksi Suomessa to- teutetaan nykyisin sekä osallistavaa, auditoivaa ja itsearviointiin rohkaisevaa arviointia että sa- manaikaisesti luotetaan PISA:n (Programme for International Student Assessment) kaltaisiin ylikansallisiin kvantitatiivisiin poikkileikkaus- arviointeihin, yhdenmukaistetaan koulutus- organisaatioiden laatujärjestelmiä ja haetaan vertailukelpoisia koulutusindikaattoreita. Ole- tettavasti haastattelemani arvioijat tunnistavat tämänkaltaisia päällekkäistyviä odotuksia.

Toinen hyvä jäsenitys löytyy Alkinilta (2012), joka on kuvannut eri aikakausien varioivia ar- viointipainotuksia ja niihin kiinnittyneitä auk- toriteetteja kolmen pääteeman avulla: metodit, arvot ja käyttö (yksityiskohtaisempi kuvaus ks. Atjonen 2015a, 212–219). Myös ne kuulunevat haastattelemieni arvioijien puheessa. Osa ar- viointitutkijoista (esim. Campbell, Cronbach, Rossi) on pitänyt tärkeänä pätevien *methodien* käyttöä arvioinnin laadun turvaajana. Sen pe- rinnäisin ydin on ollut tavoitteiden ja tulosten vertaamisessa: arvioinnin tulee kyetä vastaa- maan, onko esimerkiksi jossain koulutusohjel- massa saavutettu sitä, mihin pyrittiin. Vedungin kuvaamien aaltojen aikana myös kvantitatiivi- nen tai kvalitatiivinen metodinen ihanne ovat saaneet erilaisia painotuksia myötäillen tai ky- seenalaistaen yhteiskunta- tai kasvatustieteelli- sessä tutkimuksessa tapahtuneita menetelmä- ja

paradigmamuutoksia. Samalla on vauhditettu väittelyä tutkimuksen ja arvioinnin eroista ja yh- täläisyyksistä, jolloin huomio on saattanut lipua ohi pääasian eli että arviointikysymys ratkaisee lähestymistavan.

Monet tutkijat (esim. Scriven, Stake, House, Greene) ovat keskittyneet käsitteellistyksissään siihen, että arviointi on *arvon* antamista (engl. value, merit; *e-valua*-tion): pelkkä datan keruu ei ole arviointia. Vasta sen avulla tapahtuva ar- vottaminen, missä suhteessa tulos on hyvä tai huono ja kenen kannalta, muuttaa toiminnan arviointiksi. On tärkeää tunnistaa ja tehdä läpi- näkyväksi arvot, joista päätelmiä ja suosituksia johdetaan. Tämä puolestaan johtaa debattiin, pitääkö arvioijan aukikirjoittaa arvottamisensa suosituksiksi, vai ryhtyykö hän silloin epäasial- lisesti päätöksentekijäksi ja kaventaa arvioinnin toimeenpanijoiden autonomiaa.

Osa arviointiauktoriteeteista (esim. Stuffle- beam, Patton, Chelimsky, Cousins) on edistänyt määrätietoisesti arviointitulosten *käyttöä* ar- vioinnin kohteina olleiden ilmiöiden tai yhteisö- jen parantamisessa. Tällaisiin lukeutuu esimer- kiksi kehittävä arviointi (Patton 2011; Atjonen 2015a). Yksi arvioinnin kansainvälisesti suu- rimmista ongelmista on tulosten ja suositusten jääminen hyödyntämättä (Contandriopoulos & Brousselle 2010, 61–62; Scriven 2013, 25–28). Esimerkiksi OEDC-maat käyttävät kasvatustieteen arviointeihin paljon henkisiä ja taloudellisia voimavaroja (esim. Santiago 2013; Shewbridge 2013), mutta niiden käytännön seuraamukset saattavat jäädä vähäisiksi. Tosin joissakin maissa seuraukset saattavat nykyisin olla tuntuvia: stan- dardoiduin testein todennetut oppilaiden hei- kot oppimistulokset ovat johtaneet kokonaisten koulujen sulkemiseen esimerkiksi Britanniassa.

Arviointisuuntauksia käsitteellistetään tyypil- lisesti dikotomioilla, esimerkiksi tavoitesidon- nainen tai tavoitteista vapaa arviointi, tuotoksia tai prosesseja ensisijaistava arviointi, toteava ja kehittävä arviointi tai ulkoinen ja sisäinen ar- viointi. Alalle ovat ominaisia myös koulukun- taistuneet tai jopa ideologisoituneet näkemykset arvioinnin tehtävästä tai luonteesta; kirjallisuus- dessa esitellään vaikkapa valtattavaa, osallis- tavaa, kommunikatiivista, reaaliaikaista, res- ponsiivista ja realistista arviointia (tarkemmin Atjonen 2015a, 221–249). Voisi olettaa, että

haastattelemillani kokeneilla arvioijilla on erilaisia omakohtaisia havaintoja esimerkiksi edelle luetelluista suuntauksista.

Arviointi näyttäytyy eri tavoin erilaisissa kansallisissa koulutuskonteksteissa esimerkiksi sen hallinnon hajauttamisen asteen ja luottamuskulttuurin vuoksi. Suomessa on ainakin toistaiseksi vältytty arviointien lisääntymistä seuraavilta pahimmilta ylilyönneiltä (Simola 2015; Simola ym. 2013), joita OECD-maista on raportoitu. Jacob ym. (2015, 15) osoittavat 19 maan vertailussa yhdeksän indikaattorin avulla Suomen kuuluvan arviointikulttuuriltaan kolmen kehittyneimmän maan kärkiryhmään. Suomi ei ole ottanut käyttöön yleissivistävän koulun kansallisia kokeita, ja ammatillisen koulutuksen markkinaperusteinen eriytyminen (vrt. Lundahl ym. 2013) on vähäistä. Opettajastandardeja on välletty ja pidättydytty pitkälti kouluyksiköiden julkisista paremmuusarvioinneista (ns. ranking-listat). Silti kansainväliset yhdenmukaistavat paineet tuntuvat myös Suomessa. (Atjonen 2014; Ozga ym. 2011; Simola 2015; Simola ym. 2015.) Artikkelini tulososasta löytynee vastaus kysymykseen, ovatko paineet tuntuneet myös haastattelemini arvioijien arkityössä.

Patton (2010, 2013) on kiteyttänyt näkemyksiään arvioinnin trendeistä erilaisiksi luetteloiksi. Pidän hyödyllisenä tässä artikkelissa esiteltävän empiirisen tarkastelun kannalta erityisesti tätä kuusikohtaista ennakkointia (Patton 2010, 44):

1. Arviointi lisääntyy kansallisesti ja kansainvälisesti sekä monimuotoistuu ja globalisoituu.
2. Arvioinnista tulee hyväksytty ja arvostettu tieteitä yhdistävä osaamisala ja professio.
3. Poliittinen kiinnostus lisääntyy tilivelvollisuuteen, näyttöindikaattoreihin ja läpinäkyvyyteen.
4. Arviointikyvykkyyden ja -taitojen kehittäminen painottuu.
5. Väittely arvioinnin metodologisen pätevyyden ehdoista jatkuu.
6. Systeemisen ja kompleksisen arviointiviitekehityksen ymmärtäminen syvenee.

Edellä olevan muutoskatsauksen pohjalta muotoilin tässä artikkelissa raportoitavan tutkimuksen kysymyksiksi, *millaisia muutoksia arviointia*

ammattikseen tekevät ovat havainneet erityisesti opetus- ja kasvatusalaaan kohdentuneessa arviointityössään uransa aikana, ja millaisia oletuksia arvioinnin tulevaisuudesta arvioijilla on.

AINEISTON HANKINTA JA ANALYSOINTI

Artikkelini paikantuu pitkäaikaiseen ”Arviointi koulutuksen kentillä” -projektiin, johon on vuosien saatossa kerätty teoreettisen perustan ohjaamana monenlaisia empiirisiä aineistoja. Parin viime vuoden tutkimustyö on empiirisesti ammentanut keskeisimmin kokeneiden, arviointia työkseen tekevien henkilöiden haastatteluaineistosta. Sitä olen taivuttanut useiden tutkimuskysymyksiä palvelukseen ensisijaisesti laadullista tutkimusotetta hyödyntäen.

Haastateltavien valinta

Henkilövalinnan ensimmäisessä vaiheessa kartoitin laajasti tutkimustarkoituksiini mahdollisesti sopivia ihmisiä hyödyntäen suomenkielisiä arviointijulkaisuja, internet-sivustoja sekä keskusteluita kollegojen, arvioijien ja arviointiorganisaatioiden esimiesten kanssa. Käytin myös omaa pitkäaikaista opetus- ja kasvatusalan työkokemustani ja verkostojani.

Kartoituksen jälkeen minulla oli 25 henkilön nimiluettelo, josta seuraavaksi suljin pois kymmenen henkilöä. Tavoitteena oli haastateltavajoukko, joka *kokonaisuutena* kattaisi mahdollisimman hyvin seuraavat valintaehdot:

- opetus- ja kasvatusalalla arvioineita, joilla saisi olla kokemusta myös muilta sektoreilta (esim. nuorisotyö, sosiaali- ja terveystoimi, työvoimahallinto)
- tila-, teema- ja ohjelma-arviointia tehneitä, eli jätin pois pelkästään oppimistulosarviointeja tehneet
- kokeneita arvioijia, joille arviointi oli keskeinen työtehtävä
- ei ”puhtaasti” tutkimusta tehneitä (soveltaen erottelua arvioinnin, tutkimuksen ja arviointitutkimuksen kesken: Linnakylä & Atjonen 2008; Leimu 2010)
- joko julkisen (yliopistot, koulutuksen aluehallinto) tai yksityisen (konsulttiyritys, yksityisyrittäjä) sektorin palkkalistoilla olevia arvioijia

- sekä pieniä (järjestö)arviointeja että laajoja kansallisia tai kansainvälisiä politiikka-arviointeja tehneitä
- molempia sukupolia ja eri ikäryhmiä edustavia henkilöitä
- maantieteellisesti eri puolilla Suomea työskenteleviä arvioijia
- myös kansainväliseen arviointiin mahdollisesti osallistuneita sekä
- henkilöitä niin, että koko koulutusjärjestelmä perusasteesta toiseen ja kolmanteen asteeseen tulisi edustetuksi.

Kaikki valintahaaviini jääneet 15 henkilöä vastasivat myönteisesti haastattelupyynnöni, joten sain hyvin kaikkia edellä mainittuja kriteerejä edustavan joukon. Tutustuin näin 11 naiseen ja 4 mieheen, joilla oli arviointityökokemusta 4–30 vuotta. Osalla oli toimenkuvassaan jonkin verran muutakin työtä kuin arvioinnit. Itse toteuttamani haastattelut toteutuivat useilla paikkakunnilla, tyyppillisesti haastateltavan työpaikalla, ja ne kestivät 52–106 minuuttia.

Aineiston valinta ja analysointi

Haastattelurungon pääteemat perustuivat arviointia koskevaan lukeneisuuteeni, ja ne piti haastateltavien heterogeenisuus huomioiden muodostaa melko väljiksi. Teemat olivat seuraavat: yleiskuva arviointikokemuksista, arvioinnin lähestymis- ja tiedonhankintatavat, myönteiset kokemukset, kielteiset/ristiriitaiset kokemukset, arvioijan rooli, arviointietiikka, arviointia koskevat muutokset ja loppukointi. Lähetin muutamilla ranskalaisilla viivoilla avatun teemaluettelon haastateltaville etukäteen. Tämä artikkeli perustuu erityisesti teemaan ”arviointia koskevat muutokset”, mutta muutos- ja tulevaisuuspuhetta löytyi myös monen muun teeman yhteydestä.

Ensiksi tein koko 263-sivuisen (Times New Roman 12pt, 141 918 sanaa) haastattelulitteraation tekstihakuja sanoilla muutos, muutok*, muuttum* ja tuleva* ja kopioin niiden indikoimat tekstiotteet uudeksi tiedostoksi. Jatkoin etsintää alkuperäislitteraatiosta sanoilla inflaatio, ähky, trendi, kiire, aika*, resurs*, vaativ* ja vaatim*, ja lisäksi löydökset tähän uuteen tiedostoon. Kolmanneksi toteutin koko aineiston

huolellisen lähiluvun ja poimin lisää relevantteja tekstiotteita manuaalisesti.

Lopuksi poistin koostetiedostosta kaikki analyysiin kelpaamattomat tai sen kannalta tarpeettomat lausumat, joissa poikettiin juolahduksenomaisesti toiseen asiaan, ja osoitin poistetut kohdat merkinnällä (...). Varmistin haastatelluille satunnaisesti annettujen koodien H1–H15 näkymisen ja huolehdim, että tarvittaessa haastattelukysymys oli näkyvillä vastauksen kontekstin tunnistamiseksi. Liitin sitaattihin tarvittaessa maininnan, mihin konkreettiin arviointikohteeseen tai -esimerkkiin se liittyi, että opetus- ja kasvatusalassa pitäytyminen tulisi turvatuksi.

Näin sain analysoitavakseni 11 106 sanan (23 sivua) tekstitiedoston, jonka luin useaan kertaan muutosten ja tulevaisuus-odotusten hahmottamiseksi (Boyles 2013; Vaismoradi ym. 2013). Pelkistin ensimmäisessä vaiheessa yhteensä 113 pienteemaa pyrkien säilyttämään erilaiset näkemykset mahdollisimman kattavasti. Seuraavaksi integroin arviointia koskevan lukeneisuuteni perusteella yksittäisteemoja teemaryppäiksi, jotka kiteytin muutoslausumiksi (esim. ”Kilpailu ja valtakamppailut ovat lisääntyneet”). Muutoslausumat perustuivat tulevaisuusteemaa rikkaampaan aineistoon, koska sitä koskevia pienteemoja oli 82/113.

Aineiston laadun arviointia

Laajan maantieteellisen jakauman (ts. matkakustannusten hallinnan) vuoksi halusin saada haastatelluista henkilöistä mahdollisimman paljon tietoa. Se rajoitti alateemoihin syventymisen mahdollisuuksia, vaikka yksilöllinen keskusteluaika oli varsin pitkä. Siten esimerkiksi tässä artikkelissa käsiteltävä arvioinnin tulevaisuus jäi jälkikäteen katsoen vähäiselle huomiolle. Yhteydenotoissa korostin, että opetus- ja kasvatustalan arviointikokemus on ensisijainen keskustelujen kohde. Sen haastateltavat olivat muistaneet hyvin jo ennakkovalmistautumisessaan, ja he pitäytyivät hyvin kohdealan kysymyksissä ja esimerkeissä haastattelujen aikana.

Pitkään työkokemukseen siivilöityy tiedostamattakin monen arviointisubstanssin tuottamia elementtejä, joten en välttämättä kaikissa tarpeellisissa kohdissa osannut tarkentaa, oliko

jokin näkemys peräisin juuri opetus- ja kasvatustalouden arviointityöstä. Jos jokin esimerkki tai puheenvuoro liittyi selvästi vaikkapa teknologiaan tai työllistämiseen, en sisällyttänyt sitä tämän artikkelin aineistoon, jonka valintaa kuvasin edellä.

Haastateltavat puhuivat arvioinneista, jotka olivat kohdistuneet esimerkiksi ammatillisen koulutuksen kehittämiseen, perus- tai erityisopetuksen pedagogiseen vaikuttavuuteen (pl. oppimistulokset) tai nuorten koulutuksellisen osallistamisen ja ohjauksen parantamiseen. Monesti kyse oli jonkin toimenpideohjelman toimivuuden tai vaikuttavuuden arvioinnista oppilaitoksissa tai kokeiluhankkeissa. Osa oli tehnyt pienehköjä nuorisotoimen järjestöarviointeja tai arvioinut ison koulutusorganisaation opetushenkilöstön työolosuhteiden kohtamista tavoitteille interventioita. Mukaan mahtui muutamia niin sanottuja politiikka-arviointeja, koulutuspoliittisesti merkittävien ratkaisujen ”vaikuttavuuden” tarkasteluja. Arviointeja oli tehty joko nuorten tai aikuisten hankkeissa, eli pienten lasten tai varhaiskasvatuksen arvioinneista ei juurikaan keskusteltu.

Arvioinnin kohdeilmiöt olivat siis vaihdelleet paljon, joten niistä ääneen pohditut muutokset olivat heterogeenisiä. Se on tutkitun ilmiön monimuotoisuuden säilyttämistavoitteen kannalta selvä vahvuus, mutta tekee tulkintojeni sanoittamisesta vaativaa ja voi synnyttää lukijalle toiveen *konkreettisemmasta* esitystavasta. Pidän poikkeuksetta kiinni lupauksestani vaieta sellaisista esimerkeistä, joita haastateltavat eivät asiakasluottamussyistä halunneet raportoitaviksi, vaikka ne olisivat edistäneet lukijan samastumiskykyä.

Aineisto ei oikeuta opetus- ja kasvatustalouden laajempiin päätelmiin, mutta mahdollistaa laadullisen tutkimuksen perusidean mukaisesti lukijalle periaatteellisia samastumiskokemuksia. En pyri kohdealallekaan yleistämiseen, mikä on lähtökohtaisesti toteuttamalleni laadulliselle tutkimukselle täysin vierasta. Sen sijaan tavoitelen *yleismerkityksellistä* tietoa, joka reflektoi aiemman tutkimustiedon ja teoreettisen kontekstoinnin avulla.

Opetus- ja kasvatustalouden on yksi esimerkkikonkreettisuus yhteiskunnan julkisista palveluista, joihin kuuluu sen kanssa sekä samankaltaisia että erilaisia toimialoja. Inhimilliseen kasvuun

ja kehitykseen toiminnallaan tähtäävät sektorit tai hankkeet voivat jakaa tässä raportoitavan tapaisia arviointikokemuksia, mutta vaikkapa tekniset tai työllistämiseen liittyvät palvelut tai yritystoiminta tuottanevat toisenlaisia muutoshuomioita. Viimeksi mainittu on kuitenkin kyseenalaistettava olettama siksi, että tutkimuskirjallisuuden mukaan arviointimuutokset myös ylittävät alaspesifejä piirteitä.

TULOKSET

Arvioinnissa tapahtuneet muutokset

Löysin muutospuheesta kymmenen teemaa, joista eniten käsiteltiin arvioinnin kehittävyttä, siihen lisääntyneitä odotuksia, soveltuvia menetelmiä ja kansallista koordinoimattomuutta. Kiteytin tulokset seuraaviksi muutoslausumiksi:

1. Arviointi on muuttunut ulkoisesta tarkastelusta kehittäväksi.
2. Arvioinnista on tullut osa organisaation tavanomaista toimintaa.
3. Arviointiin kohdistuvat odotukset ovat lisääntyneet.
4. Arviointiote on muuttunut kvantitatiivisesta kvalitatiiviseksi.
5. Arviointi on muuttunut tuotoksista prosessia korostavaksi.
6. Arviointi on muuttunut vaikuttavuutta ja tehokkuutta tavoittelevaksi.
7. Kilpailu ja valtakamppailut ovat lisääntyneet arvioinnissa.
8. Arviointia tehdään tiukentuneissa aika- ja resurssikehyksissä.
9. Arviointiosaaminen on heterogeenistunut.
10. Arviointi on lisääntynyt, ja se on koordinoimattomasti toteutettua.

Muutoslausumien luettelo alkaa myönteisiksi tulkitsemistani, useimmin puhutuista kehityssuunnista ja päättyy määrällisesti lukuisimmin esille otettuihin huolenaiheisiin eli osaamisen heterogeenistumiseen ja puutteelliseen koordinointiin; luettelon keskivaiheen kilpailua ja valtaa käsiteltiin vähiten. Muodostaakseni merkityksellisiä teemakokonaisuuksia työstän seuraavissa luvuissa aineistositaattien avulla muutoslausumia hieman eri järjestyksessä kuin ne esiintyvät edellä olevassa luettelossa.

Lisääntyneen arvioinnin lisääntyneet ongelmat

Haastateltavat vahvistivat monien tutkimusten (mm. Jacob ym. 2015; Patton 2010; Santiago 2013) esiin tuoman trendin, että *arviointi ja siihen kohdistuvat odotukset ovat lisääntyneet*. Arviointi kiinnostaa aidosti: sitä ei tehdä vain muodon vuoksi (H2), ja sen uskotaan vaikuttavan yhteisöllisiin ja yhteiskunnallisiin ratkaisuihin (H5). Arviointia pidettiin yhtäältä hyvän organisaation ja sen toiminnan laadukkuuden tunnuksena ja toisaalta melkein kyseenalaistamattomana välttämättömyytenä: ”*Et ihan selvää, et arviointia tarvitaan ja paljon*. (H13)”

Arvioinnin lisääntymisen seuraukset eivät ole olleet vain positiivisia, sillä *koordinointumuutta* kritisoitiin paljon: useat tahot (esim. Karvi, ELY-keskukset, kunnan sivistystoimi) saattavat olla arviointitiedonkeruineen yhtä aikaa liikkeellä lähestyksen samoja toimijoita. Asiaan on maassamme kiinnitetty huomiota esimerkiksi organisoitaessa koulutuksen kansallista ulkoista arviointia (OKM 2012, 17–18; Silvennoinen 2013, 39) sulauttamalla Koulutuksen ja Korkeakoulujen arviointineuvostojen ja Opetushallituksen toimintoja. Myös koulutuksen kansallinen arviointisuunnitelma (Karvi 2016) tähtää ennakointiin ja koordinointiin. Liki kaikilla haastateltavilla oli koordinoinnin puutteesta näkemys ja kokemuksia, ja eräät heistä toivat näkemyksensä kärkevästi julki (alleviivaukset tämän artikkelin kirjoittajan):

Ja sitten ehkä meillä on vähän Suomessa olu sitä koordinointimuutta, että meillä on eri tahot, jotka toteuttaa arviointeja... Ja yleisempänä ilmiönä se, että on sitten samaan yksikköön... hyvin lyhyellä ajalla kohdistunut monia arviointeja, et semmosta paikoin koordinointimuutta, mikä on ihan ilman muuta selvää, että tuottaa sitten ikävän tilanteen henkilökunnalle, kuormittaa sitten kohtuuttomasti. (H4)

... semmosta arviointiähkyä sillä tavalla, että ihmisiä, kaikkia ihmisiä kannustetaan arvioimaan monenlaisia asioita. Et meillä on näitä kyselyitä ruokapöydillä ja joka paikassa, missä tavallaan jokaisen arviointi on tärkeää. (H9)

Se ideaalihan ois se, et meillä ois kansallisesti koordinoitua arviointi niin, että kerättäs sitä dataa, mut sitten ne eri tahot käyttäis sitä. Et yks kysely tai joku, et ei niinku Opetushallitus nyt ja sit joku laitos sillon ja toinen tuollon ja näin, vaan se, et meil ois kansallisella tasolla... se, ettei yhtään turhaa arviointia. (H14)

Vapaassa demokratiassa arviointien rajoittaminen on tietysti vaikeaa. Esimerkiksi tutkijoiden suunnitelmiin ei voida puuttua kansallisessa ohjauksessa, vaikka painoaloista satuttaisiin pääsemään yhteisymmärrykseen. Silti kaikkien arviointitoimeksiantajien ja -hankintapäätöksiä tekevien olisi hyvä muistaa ohje, että ”vähemmän voi olla enemmän”.

Kehittäminen ja tuloksellisuus samanaikaistrendeinä

Chelimskyn (1997, 21) mukaan arvioinnilla on periaatteessa kolme pääfunktiota: tilivelvollisuuden toteuttaminen, tiedon tuotanto ja kehittäminen. Näistä haastateltavat näkivät selvimmän muutoksen tapahtuneen tilivelvollisuudesta *kehittämisen suuntaan*, joka puhutti tulosluvun alussa esitellyistä kymmenestä muutosalusumasta määrällisesti kaikkein eniten. Haastattelemani ammattiarvioijat kuvasivat, että

... enemmän haetaan sitä, et miten me voidaan kehittää... tän kyseisen sektorin tai organisaation tai instituution tai instrumentin vaikuttavuutta ja toimivuutta... Aikaisemmin... oli enemmän sitä taipumusta, et aateltiin, että se ulkoinen arvioitsija tekee ulkoisen arvioinnin ja sit tuo jostain viisauden jostain norsunluutornistaan... Et nythän on paljon enemmän menty siihen, että sitä vuoropuhelua käydään koko ajan. (H2)

... minun mielestä on ollu erittäin positiivista tämä, että on lähdetty painottamaan... hyvin näkyvästi ja tavoitteellisesti tätä kehittävää arviointia, joka lähtee sitten tämmösestä toisen tyyppisestä ku joku mitattavuuteen perustuva, kvantitatiiviseen orientaatioon perustuva raporttien laatiminen... Se synnyttää sitä dynamiikkaa ja se on minusta erittäin tärkeää... Mä nään tämmösenä paradigman muutoksena. (H6)

Pattonin (2011, 2013) mukaan kehittävä arviointi tähtää viime kädessä siihen, että arvioinnista tulee osa organisaation tavanomaista toimintaa. Hän puhuu evaluatiivisen kulttuurin juurruttamisesta, jolloin (ulkoista) arvioijaa ei tarvita, vaan yhteisön jäsenet osaavat itse seurata ja sen pohjalta uudistaa omaa toimintaansa. Arviointitutkimus käsitteellistää samaa asiaa arviointikyvykkyydellä (esim. Lennie ym. 2015). Myös haastateltavani kuvasivat tätä ajattelutavan muutosta perinteisten laatutyön elementtien (esim. Demingin ympyrä) itseohjautuvana käytönä (H5) tai arviointitiedon arkipäiväistymisenä (H7). Arvioijasta sanottiin tulevan kriittinen ystävä, joka auttaa strukturoimaan organisaation itse kokoaman seurantatiedon hyötykäyttöä (H2).

Vaikka kehittämisen vahvistuminen oli yksi selvä haastattelujen muutosjuonne, arviointiammattilaiset tunnistivat myös tilivelvollisuushenkistä arviointikulttuuria, kuten Vedungin (2010) historiatarkastelun uusimmat aallot olettavat. Puheessa toistui lisäänntyneet *vaikuttavuuden ja tehokkuuden vaatimukset*. Tämän ilmiön tunnistivat parhaiten pisimpään arvioijina työskennelleet. Seuraavia sitaatteja lukiessa ei voi välttyä Vedungin (2010, 273) ajatukselta tieteellisen arviointiaallon paluusta:

Ehkä yks teema on tämmönen vaikuttavuusarviointi ja sen nousu yhdeksi tämmöseksi megatrendiksi. Ja se varmaan jatkossakin myös näkyy, et päätöksenteon tueksi halutaan saada aitoa vaikuttavuustietoa... Ja se haaste on nykyäänkin arvioitsijoilla, et miten he voi saada vaikuttavuustietoa, jos se arviointiprojekti kestää vuoden-kaksi. Ja nyt tää tämän hetkinen suuntaus on se, että projekteissa täytyy itse alkaa mittaamaan vaikuttavuutta. Se tuntuu kyl aika mahdolltomalta. Et siitä tulee ehkä vähän semmossa vasemmalla kädellä heitettyä tietoa. (H7)

Vaativustaso on noussu, et se ei ole vaan sitä kehittämistä, vaan tulee myös se indikaattoreitten ja tämmösten tulos... evidence-based, mitataan taloudellista vaikuttavuutta. Et se ei oo hävinny, vaan entisestään kyllä taas terävöityny... Ennen saatto tuntuu, et se riitti, et mitä ne toimijat itse koki tapahtuneen

muutosta, ihmisten oma kokemus... Nyt se ei oikein enää oookkaan, vaan halutaan, et sillä arvioinnilla vaikutetaan niihin rakenteisiin ja osotetaan vielä päättäjille tuloksia, että kanattaako ottaa tämä vaihtoehto (H12)

Aito vaikuttavuuden arviointi on metodisesti vaativaa, ja usein sen perusteista tingitään tietoisesti tai tiedon puutteen vuoksi esitellen tosiasiaa tuloksia eikä vaikutuksia, vaikuttavuudesta puhumattakaan (Dahler-Larsen 2005, 18–20; Paasio 2006, 100–102; Stame 2010, 378). Hektistytvä koulutuspoliittinen päätöksenteko ajaa arvioijia ahtaalle vaatimuksissaan yksiselitteisesti toimeenpantavista ratkaisuesityksistä, joita on aidosti vaikea tehdä ihmisyhteisöjen kompleksisesta syy-seuraussuhteiden verkostosta. Turhautunut keskustelu satunnaistetuista koe-kontrolliasetelmista laajeni etenkin Yhdysvalloissa opetusalan ”No child left behind”-ohjelman julkaisemisen jälkeen (esim. Franco 2007).

Hertting ja Vedung (2012) analysoivat artikkelissaan tulosvastuu- ja verkostoperustaisen arvioinnin yhteensovittamista. Tulosvastuullinen arviointi perustuu siihen, että johto kontrolloi alaistensa toimia tilanteessa, jossa on hajautettu päätäntävaltaa. Monimutkaistuvien arviointikysymysten hallinnassa yhden arvioijan asian- tunteemukseen perustuva käytäntö on haavoituttava, ja siksi verkostoitumista pidetään tavoiteltavana. Siinä horisontaalinen yhteistyö, jaettu vastuu ja vastavuoroinen monitorointi ovat ensisijaisia vertikaalisen päätöksentekoketjun sijasta. (Emt., 38.)

Kvalitatiivinen, prosesseja ensisijaistava metodinen muutostrendi

Jotkut haastatelluista kuvasivat muutoksia edelle selostetulla tavalla kehittävä arvioinnin käsitteellä, mutta toiset puhuivat arviointimenetelmistä todeten niiden *muuttuneen kvantitatiivisista kvalitatiivisiksi*. Kehittävä arvioinnin menetelmävarantoon laadulliset tiedonhankintatavat (tapauskuvaukset, ryhmä/teemahaastattelut, narratiivit, selonteot, päiväkirjat) soveltuvat hyvin silloin, kun pyritään antamaan toimijoille kuvailevaa palautetta ja tavoittamaan myös tarkoittamattomia, positiivisia sivuvaikutuksia eli

ilmaantumista (emergence), kuten Jalonen ym. (2013) sen sanoittavat. Kehittämisteemassa otin sitaatit henkilöiltä H2 ja H6, nyt ääneen pääsivät metodia pohtineet H9 ja H15:

... että arviointimenetelmänä tämmönen, tosiaan niinku tutkimusmenetelmistäkin tietysti ollaan siirrytty tästä... kvantitatiivisena... mutta sitten... laadullisena. Ni ehkä se kuvaa myöskin sitä kehitystä tässä arviointiasiaassakin, et tavallaan asioiden näkeminen kokonaisuuksina niin, että siihen vaikuttaa ne eri näkökulmat ja ne tuodaan esille. (H9)

... että harvassa on semmoset pelkät numerouskovaiset ja tämmöset. Mut se, mikä muutos ehkä on tapahtunu tai tuntuu, että tollasessa määrällisessäkin tiedossa ei enää hirveesti korosteta kyselyiden tämmöstä tilastollista luotettavuutta... Tämmöset isot tilaajatahot... pitää semmosiakin tuloksia hyvänä, vaikka on pieni otos ja ei oo kontrolloitu taustamuuttujia, eikä merkitsevyyttä oo laskettu desimaalien tarkkuudella, että tämmönen muutos musta ehkä on tapahtunu. Semmoset menetelmäintoilijat ja semmoset, jotka kysyy aina merkitsevyyttä, ni ne on kadonnu. (H15)

Kvalitatiiviseen tai kehittävään arviointiin huomionsa kiinnittäneet arvioijat puhuivat siitä, miten arviointi on muuttunut *tuotoksista prosessia korostavaksi*. Kirjallisuudessa prosessiarviointi tunnustetaan omana lähestymistapanaan ja kuvataan sitä eräänlaiseksi ”vastaliikkeeksi” tieteellisen arvioinnin aallolle, joka tavoitteita ja tuloksia vertailevana sivuutti prosessitiedon. Sekä kehittävä että responsiivisen tai reaaliaikaisen (Abma 2005; Nolan & Lo 2012) arvioinnin puoltajat tähdentävät prosessin aikaisen palautteen merkitystä: esimerkiksi ohjelma-arviointi ei tulisi tyypistä projektin päättövaiheen summaatiiviseen tietoon, jota ei voida käyttää hanketömminnan parantamiseen. Eräs julkisen sektorin kokenut arvioija kuvasi, miten

... itse on... päässy tekemään semmosia hyviä arviointeja, joissa on aito intressi taustalla. Että niitä oikeesti hyödynnetään ja luetaan, niissä on se prosessillisäarvo usein sitä raporttilisäarvoa suurempi. (H5)

Jo viime vuonna tuli... tarjouspyyntöjä konsultatiivisesta arvioinnista, prosessiarvioinnista, osin ehkä myös sitten kehittävästä arvioinnista. Eli tämmösiä arviointeja, missä projektin kanssa keskustellaan kaikki tavallaan alusta lähtien. Täsmennetään tavoitteita, jäsennetään päämäärää, tavoitteita, keinoja. Arvioidaan, mitkä ovat mittarit, millä voidaan tuloksia arvioida, mitkä ovat kriteerit. Ja jumpataan tätä prosessia kuntoon (H7)

Alkinin (2012) mukaan metodinen tulokulma arviointiin on aina ollut tärkeä. Dialogisen arviointiaallon (Vedung 2010) myötä tapahtunut asteittainen luopuminen kokeellis-kvantitatiivisista arviointiasetelmista lisäsi laadullista tiedonkeruutekniikkaa ja lähestymisfilosofiaa, mikä paikantuu myös yhteiskuntatieteellisillä aloilla tapahtuneeseen tutkimusmenetelmien kvalitatiiviseen paradigmamurrokseen. Metoditeema kohdentuu tärkeään peruskysymykseen arvioinnin, tutkimuksen ja arviointitutkimuksen eroista (Basson 2009; Leimu 2010; Linnakylä & Atjonen 2008). Hyvä arviointi noudattaa soveltuvien osien tieteellisen tutkimuksen periaatteita (läpinäkyvyys, perusteltavuus, metodinen osuvuus, systemaattisuus), vaikka se on yleensä käytännöllisempi sekä rahoittajien aloitteeseen ja aikaan sidotumpi kuin tutkimus. Nykyinen metodinen joustavuus ja triangulointi tekevät arviointi-tutkimus -erottelun entistä vaikeammaksi – tai toisaalta kyseenalaistavat rajanvedon tarpeellisuutta.

Resurssit, osaaminen ja kilpailu – havainnot toimintaympäristön muutoksista

Arviointi on varsin työvoimavaltainen asiantuntijatehtävä, jota ei voi automatisoida koneiden hoidettavaksi. Alati monimutkaistuvat ja odotuksiltaan suurenevat arviointitoimeksiannot vaatisivat aikaa perehtyä kunnolla arvioitavaan kohteeseen. Siten voitaisiin torjua myös lisääntyviä arviointieettisiä ongelmia (Morris 2008, 16–22). Haastatellut kuvasivat kuitenkin työnsä tapahtuvan nykyisin aiempaa tiukemmissa aikajärahakehyksissä. Projektien aikataulut olivat kiristyneet (H1), ja arvioinnin toteuttaminen hektistynyt (H11). Arvioijat konkretisoivat,

miten arviointien ”päällä et voi istua sitten kah-
ta vuotta. Siinä on tiukat aikarajat ja resurssit.
(H7)” tai että on ”nopeemmin tehtäviä hankkeita
pienemmällä rahalla... Nopea, tälle alkaa [nap-
sauttaa sormiaan] ja keväällä valmis (H12)”.

Arviointien lisääntymisen myötä opetus- ja
kasvatusalalle on tullut uusia toimijoita, joista
osa on puhtaasti liiketaloudellisin periaattein
toimivia yksityisiä tai monikansallisia yrityksiä.
Arvioijille ei Suomessa ole kunnollista perus- tai
täydennyskoulutusta toisin kuin esimerkiksi
Yhdysvalloissa ja Kanadassa on yliopistollisia ar-
vioinnin maisteriohjelmaa. Osaamisen kirjo voi
olla suuri ja keskustelu ammatin professioluon-
teesta siksi hyvin aiheellinen (Jacob & Boisvert
2010). Haastattelemieni arvioijien oma koulu-
tustausta oli heterogeeninen, ja siksi he pohtivat
eriateisen kriittisesti arvioijakenttää: Yhtäältä
arviointien hankinta- ja tilaajaosaamisen nähi-
tiin parantuneen ja hankekilpailutusikäntönnön
tulleen entistä tutummaksi ja järkipäisem-
mäksi (mm. H1, H2). Toisaalta puhuttiin kentän
polarisoitumisesta (H7), kun suuret kansain-
väliset konsulttiyritykset valtaavat markkinoi-
ta, ja pienet toimijat jäävät helposti alakynteen
kiristyvässä kilpailussa (H3).

Osaamisen laadusta kannettiin huolta pu-
humalla, miten ”arviointia tehdään aika paljon
semmosilla sapluunoilla, joissa on jo-
tenkin samat kysymyksenasettelut ja jopa mel-
kein samat kysymykset hankkeista riippumatta...
(H15)”. Tämä yksityisyrittäjänä toimiva arvioija
tarttui klassiseen arviointiosaamisdilemmaan,
riittääkö hyvänä arvioijana toimimiseen ar-
viointimenetelmän hallinta vai pitääkö tuntea
myös substanssi. Kansainvälistäkin kokemusta
hankkinut ohjelma-arvioija H3 katsoi, ettei
hänen tarvitse ymmärtää hankkeen kohteena
olevaa ilmiötä, koska projektin tavoitteiden
tietäminen riittää pätevään, tavoitteet-tulokset
-vertailulla tehtävään arvottamiseen.

Monenlaiset toimijat, niukkenevat voima-
varat ja uusliberalistinen imagon rakentamisen
into olivat näkyneet kilpailun ja valtakamppai-
lun lisääntymisenä. Kilpailun yksinkertaisesti
todettiin koventuneen selvästi (H12) tai mu-
rehdittiin valtataistelujen saaneen liikaa sijaa jo
pitkän aikaa:

*Hintataso alkaa olla niin alhainen, ja kilpailu
jotenkin säädytöntä niin, että esimerkiksi mi-
nä ja muutamat vanhemmat kollegamme, ni
ei me kovin montaa toimeksiantoa ees vuo-
teen viitsit tehdä tai sitten ihmiset erikoistuu.
(H3)*

*Arvioinnille on ollu haitallista se, että ... tälla-
set valtakysymykset ovat olleet... tähän on lii-
tetty niin valtavasti näitä intressejä, ja niihin
on hukattu valtavasti aikaa ja henkisiä voi-
mavaroja. En tiedä, mikä sitä sitten ylläpitää
tämmöstä jatkuvaa jännitettä ja intohimoa,
että tää ei ole rauhoittunu näitten kymmenien
vuosien aikana. (H6)*

Hankkeiden kilpailutusta pidettiin toisinaan
turhana ajan ja voimavarojen hukkaamisena
("hirveesti aikaa käytetään tarjousten väentämi-
seen ja kaiken maailman selvitysten tekemiseen",
H1), vaikka kohteena olevan erityisteeman pä-
teviä arvioijia olisi tiedetty olevan maassa vain
yksi tai enintään kourallinen. Olen valinnut
tämän artikkelin haastattelusitaatteihin vain
selvästi muutos- tai tulevaisuuskuvausten yh-
teydessä ilmaistut valta-aiheet kannanotot.
Kokonaisuutena ottaen valta ja intressiristiriidat
olivat huomattavasti tätä tunnistetumpia, kuten
olen raportoinut aiemmassa osatutkimuksessa-
ni (Atjonen 2015b).

Arvioinnin tulevaisuutta koskevat ennakoinnit

Arvioinnin tulevaisuutta koskeva haastattelu-
teema osoittautui melko vaikeaksi. Muutosten
takautuva tarkastelu sujui arvioijilta kohtuulli-
sin ponnisteluin, mutta kun käänsin keskustelun
kehityksen ennakointiin, empivyyks ja vastausten
katkoksellisuus lisääntyivät. Määrällisesti kat-
soen arvioinnin tulevaa kehitystä koskevia valis-
tuneita arvauksia (f = 31) löytyi selvästi vähem-
män kuin tähänastisia muutoksia (f = 82), ja ne
esitettiin monesti toteavasti ja lyhyesti, samaan
vastaukseen jopa päinvastaisia otaksumia
upottaen.

Tietystä hajanaisuudestaan huolimatta
kommentoinnit olivat temaattisesti saman-
kaltaisia kuin tähän asti koetut muutokset.
Tulevaisuusodotuksista jäsenyi yhdeksän pien-

teemaa, joista kuhunkin kajottiin 1–3 haastattelussa:

Arvioijat ennakoivat, että tulevaisuudessa arviointi korostaa kehittämistä, hyödyntämistä, osallistavuutta ja itsearviointikyvykkyyttä. Arvioinnin otaksuttiin jatkavan lisääntymistään ja painottavan enenevästi tilivelvollisuutta. Arvioinnin toimijakentän polarisointumisen voi jyrketä, oppilaitokset eriarvoistaa ja tarkastajajärjestelmä palata.

Havaituissa tähänastisissa muutoksissa nostettiin usein esille arvioinnin kehittävä ja evaluatiivisen kulttuurin edistämisen tehtävä. Niitä pidettiin myös tulevaisuuden painoaloina puhumalla konsultatiivisuudesta (H8) ja itsearvioinnista (*”et sitä osaamista ei tarvii tilata ulkopuolelta”,* H2) sekä esittämällä toivomus arvioinnista enenevästi suoraan koulutuksen järjestäjiä palvelevana toimintana, ei hallinnon tai kansallisen ohjauksen intresseihin keskittyvänä (H6). Haastateltavat sanoivat, että *”ei riittä, että katsoo menneisyyteen, vaan suunta täytyy olla aina tulevaisuudessa (H7)”* tai että arvioinnin tulisi olla *”oikeesti oppiva prosessi (H3)”*. Menetelmällisesti puhuttiin esimerkiksi osallistavuuden sekä asiakas- ja käyttäjänäkökulman vahvistumisesta (H2).

Jo muutoksien kuvauksessa ilmeni, miten samanaikaisesti on sekä tilivelvollisuus- että kehittämisorientoitunutta (Chelimsky 1997) arviointia. Sama päällekkäistyminen oli läsnä myös tulevaisuuspuheessa. Tilivelvollisuus sai yhden haastateltavan pyörtämään puheensa kesken virkkeen, ja toinen aprikoi arviointien tosiasiallista motiivia:

Must tuntuu et se [arvioinnin tulevaisuus] on ehkä kakspiippunen... meinasin sanoa ensin niin, että yhä enemmän mennään siihen, et arviointiosaamista on siel organisaation sisällä, mutta toisaaltahan me ollaan menossa enemmän siihen, että kaikki ulkoistetaan. (H3)

Kyllä mulla siis se suurin huoleni on tässä just..., et miksi ylipäätään, miksi arviointia. Ja sit... jos kerron, ni käytetäänkö sitä. Ja sit se onkin jo melkein kehäpäätelmä, et jos se ei ollu

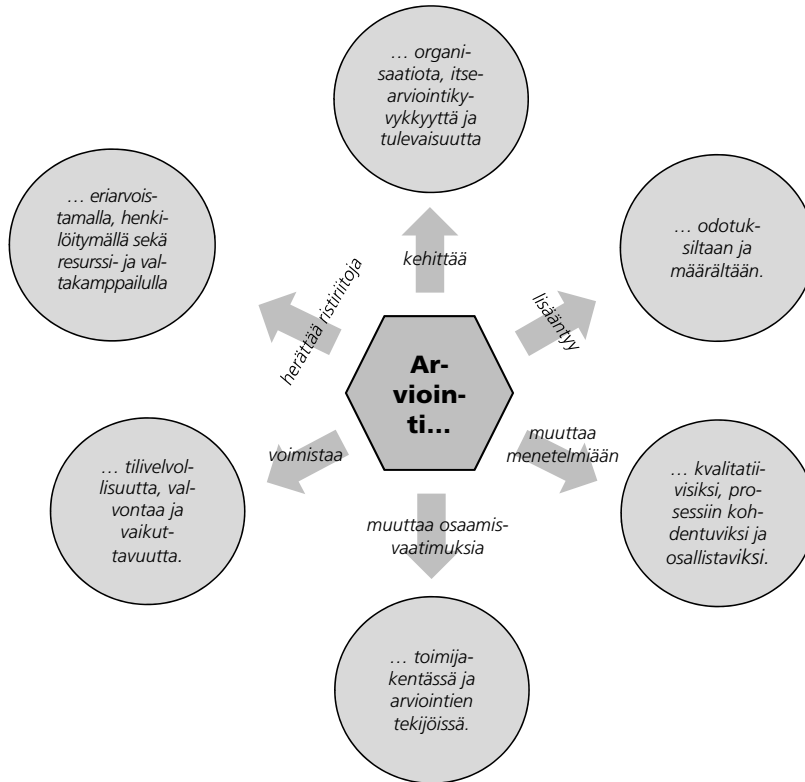
käytettävyytlähtökohtanen alunperinkään, vaan se on... joko puhtaasti tilivelvollisuutta tai sit vieläkin rankemmin et... sit ollaan oikeutettuja julkiseen tukeen. Koska on evidenssillä todennettu että. (H3)

Arvioinnin lisääntymisen ja toimijakentän polarisointumisen uskottiin (H7, H8) jatkuvan, kuten edellä totesin koetuista arviointimuutoksista. Uusina tulevaisuusskenaarioina pohdittiin arvioinnin laadun heikkenemisestä seuraavaa tarvetta korvata se jollain muulla toiminnalla, ettei eriarvoistuminen lisääntyisi; kaksi arvioijaa (H6, H14) pelkäsi jopa paluuta jonkinlaiseen koulutoimen tarkastajajärjestelmään (Lyytinen & Lukkarinen 2010). Arvioijat pohtivat tulevaisuutta ennakoidessaan ihmisten tai organisaatioiden voimistuvaa luokittamista voittajiin ja häviäjiin:

Mutta se, että kaikkia asioita, ihmisiä... kaikesta annetaan pisteitä ja pistetään järjestykseen, ni se on tietysti aika mielenkiintosta. Ja ehkä tää tämmönen mentaliteetti, mitä toi media viljelee, että arvioidaan ja nimenomaan arvostellaan, ja sitten heitetään heikot pelistä pois. Ni se on ehkä sellanen, että en mä sitä tietysti kyllä pidä hyvänä. (H9)

Et jos esimerkiks kouluvalinta lisääntyy, ja erilaiset koulut eriarvoistuu hyvin paljon ja arvioinnit viel tukee tätä kehitystä, ni arviointi saa aika huonon leiman sit kyl siinä... Miten arvioinnin roolin sais pidetty sellasena, että sen tehtävänä ois nimenomaan olla kehittämäs asioita eteenpäin, eikä niin, että tehtäs näitä ranking-listoja tai sulettas kouluja... Et miten päästään eroon tästä, tai miten pysytään siinä tilanteessa, ettei oo pakko antaa ranking-listoja? (H9)

Turun ja Helsingin yliopistojen tutkimusryhmien raportti ”Lohkoutuva peruskoulu” (Sepänen ym. 2015) puhuu mittavien tutkimusaineistojen kautta karua kieltä kouluvalintojen yhteyksistä esimerkiksi kotien sosiaaliseen pääomaan, tulotasoon, summittaisiin imagoletuksiin ja yhteiskuntaluokkaan. Arvioinnin sisäsyntyinen pyrkimys vertailemiseen voi siis uusliberalistisessa ajassa ruokkia esineellistä-



Kuvio 1. Suomalaisen arvioinnin muutos- ja tulevaisuustrendit (osana globaaleja kehityskulkuja) arviointiammattilaisten kuvaamina

vää ihmiskäsitystä ja perustua pohjoismaiselle hyvinvointivaltioajattelulle vieraisiin, koveneviin arvoihin. Jos koulutuspoliitikkoja alkaa kovasti arviotuttamaan ja mediaa julkistuttamaan, koulu- ja koulutusvalinnat voivat johtaa seurauksiin, joita kukaan ei olisi halunnut.

Jälkikäteen arvioiden tulevaisuusteema olisi voinut tuottaa kiinnostavia tuloksia, jos olisin osannut fasilitoida arvioijien pohdintaa paremmin. Moniarvoistuvassa jälkimodernisessa yhteiskunnassa arviointitrendivaihtoehtoja on lukuisasti (ks. esim. Meyer & Stockmann 2016), ja niistä on tutkimustyönkin perusteella vaikea tehdä todennäköisyysvalintoja, jotka eivät jäisi mielipiteiksi. Ymmärrän hyvin haastattelvien varovaisuutta, koska omakohtaisen arviointityön tuottama kokemusavaruus on lopulta suppea, vaikka työtä olisi tehnyt vuosia. Haastattelusarjan loppupuolella oivalsin, että haastattelurungon neutraalin kysymyksen

”Millaiseksi näet (alasi) arvioinnin *tulevaisuuden*?” sijasta olisi puhunnan lisäämiseksi kannattanut henkilökohtaistaa kysymystä: ”Ajattelepa, että jos olet arviointitöissä vielä 10 tai 15 vuoden kuluttua, mitä arvioinnille silloin kuuluu”. Vaikka aineisto nyt ei osoittautunut rikkaaksi, tulevaisuus-näkökulman sisällyttämällä vältän artikkelin rajautumisen pelkkiin menneisyyden muutoskuvauksiin.

POHDINTA

Tutkimuksen tulokset voidaan kiteyttää kuviksi 1, josta voi tunnistaa useita sisäisesti ja keskenään ristiriitaisia arviointitodellisuuksia, kuten edelle olen kansainvälisen kirjallisuuden avulla ennakooinut ja tuloksiin tulkintoina lomittanut. Sain artikkelin alussa luettelemiini Pattonin (2010, 44) kuuteen trendiin haastatteluaineistosta selkeästi tukea lukuun ottamatta

pohdintaa sen professioluonteesta (Jacob & Boisvert 2010) ja transdisiplinäärisestä tehtävistä (Scriven 2013). Haastateltavat eivät puhuneet arviointiviitekehityksen systeemisestä luonteesta eksplisiittisesti, mutta tunnistin sitä ”rivien välistä” erityisesti Pattonin (2011) edustaman kehittävän arvioinnin konseptin ansiosta. Alkinin (2012) historiallisen tarkastelun kolme juonetta – metodit, arvot ja käyttö – esiintyivät haastateltavien puheessa varsin tasapuolisesti.

Kuvio ilmentää Santiagon (2013) erittelemiä OECD-maiden koulujen ja opettajien arviointitrendejä ja niitä muotoavia kontekstitekijöitä. Silti sekä Santiagon analyysi että Shewbridgen (2013) koulutuspoliittinen katselmus nostavat voimakkaammin esiin koulutusstandardeihin ja -indikaattoreihin, mittaaamiskulttuuriin, julkisen sektorin tulostuuseen, arvioinnin kaupallistumiseen sekä median lisääntyvään kiinnostukseen liittyvää kehitystä. Ero selittyy sillä, ettei opetus- ja kasvatusalan arviointi-innoston ole Suomessa toistaiseksi ylittänyt OECD-maiden mittoihin, eikä varsinkaan kontaminoitunut sen huonoimmilla lieveilmiöillä. Suomalaiselle koulutusjärjestelmälle ominainen, vuosikymmenien ajan rakennettu luottamuskulttuuri on ylläpitänyt kehittävää ja vuorovaikutuksellista arviointiotetta.

Arvioinnin monipuolinen kehittäminen ja erilaisten konseptien kokeileminen on tärkeää, ettei ”arviointiähky tai -mania” (ts. kaikkea halutaan arvioida nopeasti ja vähin kustannuksin) johda kapea-alaisuvaan arviointinäkemukseen ja arvioinnin tyypistymiseen pelkäsi kontrollin välineeksi. Jaan myös Azzamin ja Jacobsonin (2015) huolen siitä, että arviointiin kohdistuu tutkimusta tarvitaan kipeästi lisää entistä laadukkaiden prosessien ja tuotosten turvaamiseksi.

Jotkut haastateltavat näkivät arvioinnin ja sen muutokset melko ongelmattomina ja työnsä toimeksiantojen kyseenalaistamattomana toimeenpanemisena. Osa suhtautui tutkijamaisen kriittisesti työhönsä ja siinä ilmenneisiin muutoksiin pohtien arvioijaetiikkansa venyvyyttä kamppailussa toimijakentillä (Desautels & Jacob 2012). Arvioijan yhteiskuntatieteellinen peruskoulutus näyttäisi avaavan silmiä hallinnan, vallan, resurssikamppailujen ja intressiritiriitojen tunnistamiselle paremmin kuin esimerkiksi psykologinen tai kasvatustieteellinen kompe-

tenssiperusta. Ellei mainstream-arvioinnin perusteita tunne ja kyseenalaista, on vaikea tunnistaa vaihtoehtoisia kehittäviä lähestymistapoja.

Haastateltavajoukon heterogeenisuus on nähtävä tutkimukseni vahvuutena, vaikka tulokinnassa jonkinasteinen pidättyväisyys on useimmiten eduksi. Mitä useammin olen lukeut haastattelulitteraatioita eri osatutkimuksia varten ja keskeiseen tutkimuskirjallisuuteen peilaten, sitä kirkkaammin olen arvostanut aineiston kykyä nostaa tärkeitä opetus- ja kasvatusalan arviointiteemoja. Onnistuin saavuttamaan erittäin hyvän asiantuntijajoukon. Joidenkin pitkään arvioitsijoina toimineiden ääni kuuluu tuloksissa luontaisesti toisia enemmän, koska he voivat argumentoida näkemyksiään lukuisin esimerkein. Toisaalta 4–6 vuotta työtä tehneillä on herkkyyttä kysellä rutiineja ravistelevia miksi-kysymyksiä.

Monen arvioijan kokemus myös muun kuin opetus- ja kasvatusalan arvioinnista toi näkemyksiin uudenlaista perspektiiviä, vaikka tulokset oikeuttavat enintään yleismerkityksisiin (ei yleistäviin) opetus- ja kasvatusalan arviointipohdintoihin. Toisaalta kyseessä on yksi julkisen sektorin merkittävä toimiala, jolta kummunneiden havaintojen voimaa voi koetella generalistisilla arviointitrendeillä (esim. Albæk 1998; Alkin 2012; Jacob ym. 2015; Vedung 2010; Kushner 2011; Patton 2010) tai alan ulkopuolisena lukijana kenties varoen rinnastaa muihinkin yhteiskuntasektoreihin.

Pidän tämän artikkelin tärkeimpänä tehtävänä ylläpitää kriittistä keskustelua siitä, mihin suuntaan Suomi yhtenä OECD-maana haluaa koulutusalan arviointia viedä, sillä muutoskuvauksista valottuu useita sellaisia ylikansallisia trendejä (esim. kilpailullisuus, resurssiniukuus, kulttuurierojen sivuutus, ”mittausajattelun” kapea-alaisuus), jotka eivät suomalaisen kehittämisen ja luottamuskulttuurin kannalta näytä välttämättä toivottavilta.

Vaikka tietoperustainen arviointi on arvioinnin valtavirtana monin tavoin oikeutettu, sen suoraviivainen tavoittelu ja toimeenpano voivat yksinkertaistaa liikaa kasvatus- ja koulutusprosessien monimuotoisuutta ja edistää mekanistista ihmiskäsitystä. Johansson ym. (2015, 83–84) jatkavat Vedungin esittelemän neljännen eli EBP-aallon (evidence-based practice) analyysia ja väittävät, että se on johtanut asiakkaiden

ja käyttäjien osallistamisen sivuuttamiseen vaikuttavuusarvioinneissa. Heidän artikkelinsa otikon ilmaisu ”vanishing client” tarjoaa kolkkoa tulevaisuutta kehittävän arvioinnin kannattajille.

Kaksi pitkään Suomen PISA-osuudesta vastannutta suomalaista arviointiasiantuntijaa ennakoi vuonna 2010 eräiden kansainvälisten tutkijoiden (Chelimsky, Donaldson, Mark) vana-vedessä arviointitulevaisuutta näin:

Arviointi nähdään nykyistä painokkaammin monitahoisena ja -menetelmäisenä toimintana, jolla on erilaisia tarkoituksia, kysymyksenasetteluja ja toisiaan täydentäviä lähestymis- ja käyttötapoja. Siinä on myös käytettävissä rikas metodinen variaatio. Arviointi on sallivampaa ja uuteen innostavaa, ja se pitää tervetulleina niin tulevaan kuin menneisyyteenkin katsovia hankkeita. Arvioinneissa sovelletaan ja yhdistetään aikaisempaa ennakkoluulottomammin kvantitatiivisia ja kvalitatiivisia metodeja, rakennetaan tieteidenvälisiä yhteyksiä ja luodaan kanavia arviointitulosten tehokkaaseen levittämiseen ja hyödyntämiseen. (Väljærvi & Kupari 2010, 28.)

LÄHTEET

- Abma, Tinneke (2005). Responsive evaluation: Its meaning and special contribution to health promotion. *Evaluation and Program Planning*, 28(3), 279–289.
- Albæk, Erik (1998). Knowledge, interests and the many meanings of evaluation: a developmental perspective. *Scandinavian Journal of Social Welfare*, 7(2), 94–98.
- Alkin, Marvin (2012). *Evaluation roots*. London: Sage.
- Atjonen, Päivi (2014). School-focused evaluation in focus. *Nordic Studies in Education*, 34(4), 279–297.
- Atjonen, Päivi (2015a). *Kehittävä arviointi kasvatusalalla*. Joensuu: Kirjokansi.
- Atjonen, Päivi (2015b). “Your career will be over” – Power and contradictions in the work of educational evaluators. *Studies in Educational Evaluation*, 45, 37–45.
- Azzam, Tarek & Jacobson, Miriam (2015). Reflections on the future on the research on evaluation. *New Directions on Evaluation*, 148, 103–116.
- Basson, Ray (2009). Evaluation research in education – Developments in qualitative approaches. *Evaluation Research in Education*, 10 (1), 55–70.
- Boyles, Nancy (2013). Closing in on close reading. *Educational Leadership*, 70(4), 36–41.
- Chelimsky, Eleanor (1997). The coming transformations in evaluation. Teoksessa Chelimsky, Eleanor & Shadish, William (eds.) *Evaluation for 21st century* (s. 1–26). Sage: Thousand Oaks.
- Contandriopoulos, Damien & Brousselle, Astrid (2010). Evaluation models and evaluation use. *Evaluation*, 18(1), 61–77.
- Dahler-Larsen, Peter (2005). *Vaikuttavuuden arviointi. Hyvät käytännöt -menetelmäkirja*. http://www.julkari.fi/bitstream/handle/10024/77071/vaikuttavuuden_arv.pdf?sequence=1 (luettu 2.1.2017)
- Desautels, Geoffroy & Jacob, Steve (2012). The ethical sensitivity of evaluators: A qualitative study using a vignette design. *Evaluation*, 18(4), 437–450.
- Guba, Egon & Lincoln, Yvonna (1989). *The fourth generation evaluation*. Newbury Park: Sage.

Kuvaus tavoittaa paljon siitä, mistä haastateltavani kertoivat. Ennustetta voi jo nyt pitää melko ihanteellisena ajassa, jossa NPM-hengessä (New Public Management) tietuon perustuva, uusliberalistinen ja yhdenmukaistamiseen mieltynyt arviointi tiukentaa otettaan ylikansallisenä trendinä. Kansallisesti merkittävä koulutuspoliittinen kysymys kuuluukin: Millaisin perustein otteen sallitaan voimistuvan luottamuskulttuurin koulutus-Suomessa?

Toisaalta Väljærven ja Kuparin kuvaama tulevaisuusnäky sisältää riskin arvioinnin laadun vaarantumisesta. Luovasti mutta kehnosti yhdistetyt menettelyt voivat tuottaa osaamattomissa käsissä epäpäteviä tuloksia. Jos niiden toimeenpanosta kiinnostuvat valikoivoin tarkoituksiperin liikkeellä olevat päättäjät, lopputulos voi olla onneton. Arvioinnin laadunvarmistus kehkeytyy koko ajan tärkeämmäksi periaatteelliseksi ja käytännölliseksi ongelmaksi, joka vaatii osaavia ja valtaintresseistä riippumattomia ratkaisijoita. Laadukkaan arviointikoulutuksen tarve on suuri, jotta nopeasti lisääntyvän toiminnan dysfunktioita voitaisiin estää tai vähentää.

- Franco, Suzanne (2007). Reauthorization of NCLB: Time to reconsider the scientifically based research requirement. *Nonpartisan Education Review*, 6(6), 1–14.
- Herttig, Nils & Vedung, Evert (2012). Purposes and criteria in network governance evaluation: How far does standard evaluation vocabulary take us? *Evaluation*, 18(1), 25–44.
- Jacob, Steve, Speer, Sandra & Furubo, Jan-Eric (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation*, 21(1), 6–31.
- Jalonen, Harri, Lindell, Juha, Puustinen, Alisa & Raisio, Harri (2013). Yhteistyön kääntöpuoli – kun itseorganisoituminen epäonnistuu ja ilmaantuminen yllättää. *Hallinnon tutkimus*, 32(4), 284–300.
- Johansson, Kersti, Denvall, Verner & Vedung, Evert (2015). After the NPM Wave: Evidence-based practice and the vanishing client. *Scandinavian Journal of Public Administration*, 9(2), 69–88.
- Karvi. 2016. Koulutuksen arviointisuunnitelma vuosille 2016–2019. <http://karvi.fi/publication/koulutuksen-arviointisuunnitelman-vuosille-2016%E2%88%922019/> (luettu 2.1.2017).
- Kushner, Saville (2011). A return to quality. *Evaluation*, 17(3), 309–312
- Leimu, Kimmo (2010). Koulutuksen arviointityön kehystekijöitä tiedon vaikuttavuuden kannalta. Teoksessa Korkeakoski, Esko & Tynjälä, Päivi (toim.) *Hyötyä ja vaikuttavuutta arvioinnista* (s. 31–60). Koulutuksen arviointineuvoston julkaisuja 50.
- Lennie, June, Tacchi, Jo, Wilmore, Michale & Koirala, Bikash (2015). A holistic, learning-centred approach to building evaluation capacity in development organizations. *Evaluation*, 21(3), 325–343.
- Linnakylä, Pirjo & Atjonen, Päivi (2008). Arviointi, tutkimus ja arviointitutkimus koulutuksen tietotuotannossa. Teoksessa Korkeakoski, Esko & Silvennoinen, Heikki (toim.) *Avaimia koulutuksen arvioinnin kehittämiseen* (s. 79–96). Koulutuksen arviointineuvoston julkaisuja 31.
- Lundahl, Lisbeth, Erixon Arreman, Inger, Holm, Ann-Sofi. & Lundström, Ulf (2013). Educational marketization the Swedish way. *Education Inquiry*, 4(3), 497–517.
- Lyytinen, Heikki K. & Lukkarinen, Esko (2010). Arvioinnin lakia niin kuin sitä luetaan historiasta nykypäivään. Teoksessa Korkeakoski, Esko & Tynjälä, Päivi (toim.) *Hyötyä ja vaikuttavuutta arvioinnista* (s. 77–98). Koulutuksen arviointineuvoston julkaisuja 50.
- Meyer, Wolfgang & Stockmann, Reinhard (2016). The future of evaluation. Global trends, new challenges, shared perspectives. London: Palgrave MacMillan.
- Morris, Michael (2008). Ethics and evaluation. Teoksessa Morris, Michael (ed.). *Evaluation ethics for best practice* (s. 1–29). New York: Guildford.
- Nolan, Clare & Lo, Fontane (2012). *Getting real about real-time evaluation*. <http://blueavocado.org/content/getting-real-about-real-time-evaluation> (luettu 2.1.2017).
- OECD (2013). Overview. Teoksessa OECD. *Synergies for better learning. An international perspective on evaluation and assessment*. (s. 17–28). OECD.
- OKM (2012). *Kansallisen koulutuksen arviointikeskuksen perustaminen*. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 19.
- Ozga, Jenny, Dahler-Larsen, Peter, Segerholm, Christina & Simola, Hannu (2011). *Fabricating quality in education. Data and governance in Europe*. London: Routledge.
- Paasio, Petteri (2006). Yleinen ja erityinen viitekehys arvioinnista. *Hallinnon tutkimus*, 25(3), 92–107.
- Patton, Michael (2010). Future trends in evaluation. Teoksessa Segone, Marco (ed). *From policies to results. Developing capacities for country monitoring and evaluation system* (s. 44–57). http://www.theoryofchange.org/wp-content/uploads/toco_library/pdf/2010_-_Segone_-_From_Policy_To_Results-UNICEF.pdf (luettu 2.1.2017).
- Patton, Michael (2011). *Developmental evaluation*. New York: Guilford.
- Patton, Michael (2013). The future of evaluation in society. Top ten trends plus one. Teoksessa Donaldson, Stewart (ed.) *The future of evaluation in society* (s. 45–62). Charlotte, NC: Information age publishing.
- Santiago, Paulo (2013). Trends in evaluation and assessment. Teoksessa OECD. *Synergies for better learning. An international perspective on evaluation and assessment* (s. 35–56). OECD.
- Scriven, Michael (2013). The foundation and future of evaluation. Teoksessa Donaldson, Stewart (ed.) *The future of evaluation in society* (s. 11–44). Charlotte, NC: Information age publishing.
- Shewbridge, Claire (2013). Education system evaluation: Informing policies for system improvement Teoksessa OECD. *Synergies for better learning. An international perspective on evaluation and assessment* (s. 583–655). OECD.

- Silvennoinen, Heikki (2013). Koulutuksen arviointineuvoston asema arvioinnin kentällä ja koulutuspolitiikassa. Teoksessa Knubb-Manninen, Gunnel, Niemi, Hannele & Pietiläinen, Ville (toim.) *Kansallinen arviointi kohti tulevaisuutta* (s. 37–48). Koulutuksen arviointineuvoston julkaisuja 63.
- Simola, Hannu (2015). Arvioinnin valta, voima ja kunnia. Teoksessa Simola, Hannu. *Koulutusihmeen paradoksit. Esseitä suomalaisesta koulutuspolitiikasta*. Tampere: Vastapaino, 213–293.
- Simola, Hannu, Rinne, Risto, Varjo, Janne & Kauko, Jaakko (2013). The paradox of education race: how to win the ranking game by sailing to headwind. *Journal of Education Policy*, 28(5), 612–633.
- Seppänen, Pii, Kalalahti, Mira, Rinne, Risto & Simola, Hannu (2015). *Lohkoutuva peruskoulu. Perheiden kouluvalinnat, yhteiskuntaluokat ja koulutuspolitiikka*. Kasvatusalan tutkimuksia 68.
- Stame, Nicoletta (2010). What doesn't work? Three failures, many answers. *Evaluation*, 16(4), 371–387.
- Vaismoradi, Mojtaba, Turunen, Hannele & Bondas, Terese (2013). Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences*, 15(3), 398–405.
- Vedung, Evert (2010). Four waves of evaluation diffusion. *Evaluation*, 16(3), 263–277.
- Virtanen, Petri (2005). Neljännen sukupolven evaluationin jälkeinen arviointitutkimus? *Hallinnon tutkimus*, 23(2), 16–29.
- Väljäärvi, Jouni & Kupari, Pekka (2010). Koulutuksen arvioinnin näkökulmat ja arviointien hyödyntäminen. Teoksessa Korkeakoski, Esko & Tynjälä, Päivi (toim.) *Hyötyä ja vaikuttavuutta arvioinnista* (s. 21–29). Koulutuksen arviointineuvoston julkaisuja 50.