

11

Vaikuttavuustarkastuksen kompastuskivet

Tomi Rajala

Vaikuttavuustarkastuksessa voidaan varmentaa kansalaisille, että valtiovallan toiminta tuottaa haluttuja vaikutuksia yhteiskunnassa. Vaikuttavuustarkastuksia tehdään kuitenkin sekä Suomessa että kansainvälisesti hyvin vähän verrattuna laillisuus- ja tilintarkastuksiin emmekä juurikaan tiedä, miksi vaikuttavuustarkastuksia on tehty hyvin vähän Suomessa verrattuna laillisuus- ja tilintarkastuksiin. Syitä etsitään suorittamalla kvalitatiiviseen synteesiin pyrkivä kirjallisuuskatsaus. Tässä tutkimusmenetelmässä temaattisen analyysin avulla syntetisoidaan kansainvälisen tutkimuskirjallisuuden esille nostamia keskeisiä vaikuttavuuden mittaamisen sekä tulkitsemisen haasteita ja tutkitaan näiden haasteiden mahdollista ilmentymistä kotimaisessa kontekstissa suoritetuissa laillisuus- ja tilintarkastuksissa. Analyysin lopputuloksena saadaan kvalitatiivinen synteesi tekijöistä, jotka haittaavat vaikuttavuustarkastuksia enemmän kuin laillisuus- ja tilintarkastuksia Suomessa. Nämä tekijät olivat tarkastuskriteerien muodostaminen, vaikuttavuusmittaamisen tekninen toteuttaminen, vaikuttavuusinformaation tulkinta ja vaikuttavuustarkastuksen aiheuttama epätasa-arvo. Tutkimustulokset avaavat ovia empiirisille ja teoreettisille jatkotutkimuksille sekä vaikuttavuustarkastuksen käytännön kehittämiselle.

Johdanto

Valtion perimmäisenä tarkoituksena on tuottaa haluttuja vaikutuksia yhteiskuntaan (Scott 1996, 33). Perinteisesti on pidetty tärkeänä myös sitä, että julkinen sektori toimii läpinäkyvästi eli tuottaa esimerkiksi informaatiota tuoteista vaikutuksista veronmaksajille (ks. esim. Hood 2007). Ajatus siitä, että vaikuttavuusinformaatio on olennainen osa julkisen sektorin toimintaa, on hyvin vanha (ks. Bachman 1912) ja vakiintunut (ks. Van Dooren, Bouckaert & Halligan 2015) sekä laajasti hyväksytty (Smith 1996). Tarkastus kytkeytyy

läheisesti sekä julkisen sektorin vaikuttavuuteen että läpinäkyvyyteen, koska tarkastustoiminta informaation tuottajana ja varmentajana palvelee sekä julkisen sektorin vaikuttavuuden kehittymistä että kansalaisten informaatiotarpeita (VTV 2010a).

Vaikuttavuustarkastuksen tehtävänä on tuottaa ja varmentaa julkisen sektorin vaikuttavuusinformaatiota sekä tukea julkisten palveluiden vaikuttavuuden kehittymistä. Tärkeistä tehtävistä huolimatta vaikuttavuustarkastuksia tehdään kansainvälisesti ja kansallisesti yllättävän vähän verrattuna tilintarkastuksien ja laillisuustarkastuksien määrään (ks. Funnell 1998; Jacobs 1998; Radcliff 1998; Guthrie & Parker 1999; Pollitt 2003; Lonsdale 2008; Skarbak 2009; Grönlund, Svärdesten & Öhman 2011). Vaikuttavuustarkastuksien vähäinen lukumäärä vaikuttaa sekä julkisen sektorin läpinäkyvyyteen että vaikuttavuuden kehittämiskykyyn. Tuloksellisuuskirjallisuuden tunnistama vaikuttavuustarkastuksien vähäisyys ei siis ole mikään vähäpätöinen ongelma.

Nykyinen tuloksellisuustarkastukseen liittyvä tutkimuskirjallisuus ei ole pohtinut kovinkaan systemaattisesti sitä, mistä vaikuttavuustarkastusten alhainen lukumäärä voisi johtua. Yksi merkittävä puute aiemmassa tutkimusperinteessä on se, että tuloksellisuustarkastusta tutkineet tieteenharjoittajat eivät ole järjestelmällisesti tunnistanee niitä haasteita, jotka ilmenevät vaikuttavuustarkastuksissa ja puuttuvat vastaavasti tilintarkastuksista (engl. financial audit) ja laillisuustarkastuksista (engl. compliance audit). Toisaalta nykytutkimus ei kerro meille sitä, mitkä kaikille kolmelle tarkastustyypille yhteiset haasteet ovat sellaisia, jotka aiheuttavat eniten ongelmia juuri vaikuttavuustarkastukselle. Tämä artikkeli pyrkii selvittämään juuri näitä haasteita ja sitä kautta täyttämään tutkimuskentässä olevaa tyhjiötä.

Tutkimuskirjallisuus on pääasiassa analysoinut tuloksellisuustarkastusta kolmesta eri näkökulmasta käsin ja nämä ovat seuraavat (Longsdale, Wilkins & Ling 2011):

1. Fundamentaaliset ideat ja käsitteet, jolloin on tarkasteltu tuloksellisuustarkastuksen roolia ja rationaalisuutta sekä siihen liittyviä keskeisiä käsitteitä. Erot tuloksellisuustarkastukseen liittyvissä institutionaalisissa ennakkoehdossa eli lainsäädännössä, informaation saatavuudessa ja niin edelleen.

2. Tuloksellisuustarkastuksen käytännöt, joilla viitataan tarkastusmetodeihin- ja tekniikkoihin, tarkastuskohteisiin- sekä kysymyksiin, tarkastusnormeihin ja tarkastustuloksien käyttämiseen.
3. Tuloksellisuustarkastuksen erilaisia käytäntöjä selittävät kulttuuritekiäjät ja lainsäädäntöön liittyvät muuttajat.

Käsillä oleva tutkimus pyrkii jatkamaan tuloksellisuustarkastuksen erilaisia käytäntöjä selittävää perinnettä. Tutkimuksen tarkoituksena ei kuitenkaan ole löytää eri maiden välisiä kulttuurieroja tai hahmottaa kansallisen lainsäädännön vaikutusta erilaisiin käytäntöihin vaikuttavuustarkastuksessa. Tavoitteena sitä vastoin on löytää tekijöitä, jotka voisivat selittää sitä, miksi vaikuttavuustarkastamista ei ole juurikaan harrastettu Suomessa verrattuna tilintarkastukseen ja laillisuustarkastukseen. Nämä vaikuttavuustarkastamisen puuttumista selittävät tekijät suomalaisessa kontekstissa liittyvät toki tiiviisti niin tuloksellisuustarkastuksen käytäntöjä ohjaaviin rationaliteetteihin kuin vaikuttavuustarkastusten esiintymistiheyteen sekä kansallisesti että mahdollisesti myös kansainvälisestikin. Tutkimuksen rajautumista suomalaiseen tarkastusympäristöön voidaan perustella sillä, että syvällisen analyysin tekeminen vaikuttavuustarkastuksen haasteiden ilmenemisestä kotimaisessa tilintarkastuksessa ja laillisuustarkastuksessa on hyvä tehdä ensin ennen kuin siirrytään kansainvälisiin vertailuihin. Artikkelin pituuden puitteissa ei ole mahdollista toteuttaa sekä kotimaista että kansainvälistä analyysia ja tästä syystä jokin rajausvalinta on tehtävä.

Tutkimuksen tavoitteeseen pyritään pääsemään teemoitellun kirjallisuuskatsauksen avulla. Tutkimus käyttää hyväkseen abduktiivista päättelyä (ks. Peirce 1998), konstruktivistista tutkimusotetta (Guba & Lincoln 1998) ja meta-evaluaatiota (Pawson & Tilley 1997). Haasteiden löytämiseksi käytetään metasynteesiä, joka on yksi meta-evaluaation metodeista (ks. lisää Salminen 2011). Metasynteesin avulla pyritään kuvailemaan ja selittämään etsittyjä haasteita. Kuvailu ja selittäminen tehdään uudelleen tulkitsemalla aiempia tutkimustuloksia. Konstrukttiivinen tutkimusote viittaa siihen, että vaikuttavuustarkastuksen haasteita kuvataan tässä tutkimuksessa aiempien kansainvälisten tutkimusten esittämien argumenttien avulla, jolloin ollaan kiinnostuneita siitä, minkälaista kuvaa vaikuttavuuden mittaamisen ja tulkinnan haasteista

tiede on aikaisemmissa tutkimuksissa rakentanut. Näiden tieteellisten argumenttien avulla pyritään tarkastelemaan kahta asiaa. Ensinnäkin, valitut tieteelliset argumentit kuvaavat vaikuttavuuden mittaamiseen ja tulkintaan liittyviä ongelmia, joiden voidaan ajatella hankaloittavan myös vaikuttavuuden tarkastuskäytäntöjen muodostamista Suomessa. Toiseksi, argumenttien avulla pyritään luomaan vertailuasetelma, jossa verrataan sitä, esiintyykö vaikuttavuustarkastuksen yleisesti tunnetut ongelmakohdat Suomessa tapahtuvassa tilintarkastuksessa ja laillisuustarkastuksessa. Kuva Suomessa tapahtuvasta laillisuus- ja tilintarkastuksesta luodaan käyttämällä hyväksi tutkimuskirjallisuutta, kansallisia lakeja sekä Suomessa käytössä olevia kansainvälisiä standardeja. Näin menettelemällä pyritään löytämään ne tekijät, jotka voivat selittää vaikuttavuustarkastusten vaikean läpiviemisen kotimaisessa tarkastuskentässä. Vaikuttavuustarkastusta vaivaavien uniikkien haasteiden löytäminen ja ymmärtäminen voi auttaa hahmottamaan sitä, miksi vaikuttavuustarkastuksia on tehty Suomessa niinkin vähän kuin niitä on tehty.

Tutkimuksen keskeisimpänä tuloksena nousi esille neljä isompaa kokonaisuutta, jotka vaikeuttavat vaikuttavuustarkastamista. Nämä kokonaisuudet ovat tarkastuskriteerien muodostaminen, vaikuttavuusmittaamisen tekninen toteuttaminen, vaikuttavuusinformaation tulkinta ja vaikuttavuustarkastuksen aiheuttama epätasa-arvo ja ne muodostuvat joukosta pienempiä haasteita. Vaikuttavuustarkastusta vaivaavien ongelmien systemaattinen tunnistaminen auttaa sekä käytännötoimijoita että tutkijoita vaikuttavuustarkastuksen kehittämiseen tähtäävässä toiminnassa, koska ongelmien muodostaman kokonaisuuden tunnistaminen on ensimmäinen askel kohti näiden ongelmien tutkimista ja ratkaisemista.

Artikkelin teksti tästä eteenpäin jakautuu kuuteen osaan. Näistä ensimmäisessä osassa määritellään tilintarkastus, laillisuustarkastus ja vaikuttavuustarkastus käsitteinä. Toisessa osassa voidaan tarkastuskriteerien muodostamista tilintarkastuksessa, laillisuustarkastuksessa ja vaikuttavuustarkastuksessa. Niin sanottu kolmas osio on omistettu mittausmenetelmien tarkasteluun näissä kolmessa edellä mainitussa tarkastustyyppissä. Neljäs osa tarkastelee tulkintojen tekemistä eri tarkastustyypeissä. Viidennessä osiossa pohditaan tilivelvollisten eriarvoista asemaa, joka saattaa helpommin syntyä

vaikuttavuustarkastamisen kuin tilintarkastuksen ja laillisuustarkastuksen sivutuotteena. Viimeinen eli kuudes osa vetää yhteen artikkelin keskeisimmät havainnot ja pohtii relevantteja tutkimuskysymyksiä

Mitä tilintarkastus, laillisuustarkastus ja vaikuttavuustarkastus oikein tarkastavat?

Nykyisen kirjanpitolain (1336/1997) 3 luvun 1 §:n mukaan tilikaudelta on laadittava tilinpäätös, joka sisältää tilinpäätöspäivän taloudellista asemaa kuvaavan taseen, tuloksen muodostumista kuvaavan tuloslaskelman, varojen hankintaa ja niiden käyttöä selvittävän rahoituslaskelman (jos kirjanpito-velvollinen on suuryritys tai yleisen edun kannalta merkittävä yhteisö) sekä taseen, tuloslaskelman ja rahoituslaskelman liitteinä olevat tiedot. Tilintarkastus keskittyy tarkastamaan sitä, onko entiteetin tilinpäätös laadittu lakien sekä asetusten mukaisesti, ja sisältääkö tilinpäätöksessä oleva talousinformaatio olennaisia virheitä, jotka ovat aiheutuneet joko petoksesta tai inhimillisen virheen johdosta (ISSAI 100, kohdat 22 ja 30). Käytännössä tilintarkastus toteutetaan niin, että tilintarkastaja hankkii tilintarkastusevidenssiä tilinpäätöksessä esitetyistä tiedoista ja niiden esittämistavasta, tilikaudella tapahtuneista liiketapahtumista ja muista tapahtumista sekä tilinpäätöksen tilien saldoista voidakseen todeta tilinpäätöksen oikeellisuuden (Tomperi 2009, 85–86).

Laillisuustarkastuksessa on pyrkimyksenä keskittyä siihen, onko tarkastettava toiminta standardien, lakien ja asetusten, normien, budjettipäätösten, toimintapoliitikoiden, vakiintuneiden käyttäytymiskoodistojen mukaista ja vastaako toiminta yleisiä hyvän hallinnon periaatteita ja hyviä virkamieskäytäntöjä (ISSAI 100, kohta 22). Laillisuustarkastuksen aiheita Valtiontalouden tarkastusvirastossa ovat talousarviotulot- ja menöt, tase, vastuut, hallinnointiprosessit sekä muut asiat ja keskeisimpiä aihealueita ovat valtionavut, hankinnat ja tulot (VTV 2015). Kunnissa laillisuustarkastuksen yhteydessä tarkastetaan seuraavia asioita (L 365/1995, 73 §; L 410/2015, 123 §):

1. onko kunnan hallintoa hoidettu lain ja valtuuston päätösten mukaisesti;

2. antavatko kunnan tilinpäätös ja siihen kuuluva konsernitilinpäätös tilinpäätöksen laatimista koskevien säännösten ja määräysten mukaisesti oikeat ja riittävät tiedot kunnan tuloksesta, taloudellisesta asemasta, rahoituksesta ja toiminnasta;
3. ovatko valtionosuuksien perusteista annetut tiedot oikeita;
4. onko kunnan sisäinen valvonta ja riskienhallinta sekä konsernivalvonta järjestetty asianmukaisesti.

Tilintarkastuksen ja laillisuustarkastuksen suhteesta Kosonen (2005) toteaa, että tilintarkastus on perimmäiseltä luonteeltaan laillisuustarkastusta. Tilintarkastuksessa tarkastetaan siis myös esimerkiksi lain noudattamista, talousarviotuloja- ja menoja, kunnan hallintoa, sisäistä valvontaa ja niin edelleen.

Huomionarvoista näyttäisi olevan se, että tilintarkastus ja laillisuustarkastus tarkastavat siis tilikaudella tapahtunutta toimintaa, kuten panoskäyttöä (toimintamenoja, investointeja tai omaisuuden käyttöä yms.), tuotoksien tekemistä (esim. liikeloudelliset suoritteet), ja toimintaprosesseja (sisäistä valvontaa jne.). Vaikuttavuustarkastuksessa sen sijaan tulisi ensisijaisesti tarkastaa loppusuoritteista syntyneitä vaikutuksia eikä pelkästään, jos ollenkaan, loppusuoritteiden määrää (vrt. Smith 1996; Hatry 2006). Vaikuttavuustarkastuksessa voidaan käytännössä tarkastaa joko loppusuoritteita (toimintaa) ja niiden vaikutuksia tai pelkkiä saavutettuja vaikutuksia, jolloin loppusuoritteita ei vaikuttavuustarkastuksessa tarkasteltaisi ollenkaan. Tämä on selkeä paradigman muutos verrattuna tilintarkastukseen ja laillisuustarkastukseen, joissa on selkeästi fokus toiminnantarkastamisessa ja toiminta itsessään on joko sanktioitavaa, palkittavaa tai neutraalia. Toki tilintarkastukseen ja laillisuustarkastukseen liittyy seurauksien pohtiminen esimerkiksi vastuuvapauden epäämisen kontekstissa, mutta tilintarkastuksessa lähtökohta on se, että tarkastus kohdistetaan nimenomaan tilikaudella tapahtuneeseen toimintaan ja seurauksia pohditaan vasta siinä vaiheessa, jos toiminta ei vastaa oletettua.

Jotta vaikutuksia voitaisiin tarkastaa, pitäisi pystyä määrittelemään se, minkälaisista vaikutuksista olemme kiinnostuneita. Vaikutuksien tarkastamisessa voidaan tarkastaa hyvin monen tyyppisiä vaikutuksia riippuen siitä, miten vaikutus halutaan määritellä. Erilaisia tapoja luokitella vaikutuksia on tunnistettu lukuisia. Virtanen ja Vakkuri (2016) puhuvat käyttäytymisvaiku-

tuksista. Käyttäytymisvaikutusten vastinparina voidaan nähdä ihmisen toimintaympäristöön kohdistuvat vaikutukset (McClure 1982). Vedung (1997) on tunnistanut lukuisia erilaisia vaikutuksia, jotka eroavat jossain määrin jo edellä mainituista. Nämä vaikutukset ovat seuraavan tyyppisiä (Vedung 1997):

1. asiakas- ja yhteiskuntavaikutukset
2. pitkän-, keskipitkän-, ja lyhyenaikavälin vaikutukset
3. organisaation ulkoiset ja sisäiset vaikutukset
4. positiiviset ja negatiiviset vaikutukset
5. odotetut ja odottamattomat vaikutukset
6. tahalliset ja tahattomat vaikutukset
7. subjektiiviset ja objektiiviset vaikutukset
8. kvalitatiiviset ja kvantitatiiviset vaikutukset.

Yeung ja Matheison (1998) toteavat, että vaikutuksia voi syntyä eri yhteiskuntasektoreilla (esim. koulutussektori, ympäristösektori, talouselämä jne.). Martin (1997) kirjoittaa, että vaikutuksia voi ilmetä palvelu- ja ohjelmasolla, organisaatio- ja virastotasolla sekä yhteisö- ja valtiotasolla. Jokainen toiminto, jota tilintarkastus ja laillisuustarkastus tarkastavat, voi teoriassa synnyttää kaiken tyyllisiä yllä eriteltyjä vaikutuksia, jotka voivat syntyä eri sektoreilla ja organisaatiotasolla yhtäaikaaisesti tai viiveellä. Vaikuttavuustarkastus onkin luonteeltaan kompleksisempaa kuin tilintarkastus tai laillisuustarkastus.

Tarkastuskriteerien muodostaminen

Tilintarkastuksessa tarkastuskriteerit tulevat standardeista ja muista normatiivisista dokumenteista (ks. Lindeberg 2007). Lisäksi tilintarkastus voi saada lisätarkastuskriteerejä talousarvioista, strategisista suunnitelmista ja niin edelleen (esim. L 365/1995; L 410/2015). Laillisuustarkastuksessa tarkastaja saa tarkastuskriteerinsä sellaisilta auktoriteeteiltä, kuten esimerkiksi lainsäädäntö sekä johtosääntö, ja vertaa näitä kriteereitä sitten itse toimintaa (ISSAI 100). Vaikuttavuustarkastuksen kriteerit voivat tulla normatiivisista dokumenteista kuten esimerkiksi strategisesta suunnitelmasta tai talousarvioista (VTV 2010b). Merkittävänä painopiste-erona tilintarkastukseen ja laillisuustarkastukseen on kuitenkin se, että standardeissa ja lainsäädännössä harvemmin

määritellään eksakteja vaikuttavuustavoitteita- tai mittareita (ks. Lindeberg 2007), koska niiden tarkka määrittäminen vaatisi samanlaista konsensusta tarkastuskohteista ilmiönä kuin tilintarkastuksessa ja laillisuustarkastuksessa on. Lisäksi vaikuttavuustarkastuskohteiden tarkka määrittäminen laissa vähentäisi normien joustavuutta. Poliittista vallan vaihtelua mukaileva sykli eli esimerkiksi talousarviosykli vaikuttaisikin olevan lähtökohtaisesti vaikuttavuustavoitteiden- ja mittarien tarkalle määrittämiselle joustavampi vaihtoehto kuin lainsäädäntöön ja standardien luomiseen kytkeytyvä sykli.

Arens, Elder ja Beasley (2005) toteavat, että vaikuttavuustarkastuksessa käytetyt tarkastuskriteerit voidaan praktiikassa muodostaa neljällä tapaa: historiallisen suorituskyvyn, verokkir ryhmän suorituskyvyn, suoritusasostandardien tai tarkastajan ja entiteetin johdon välisen sopimuksen perusteella. Näissä jokaisessa tavassa näyttäisi olevan sisäänkirjoitettuna oletuksena se, että vaikuttavuus ilmiönä ja vaikuttavuusmittari pystytään määrittämään, koska vaikuttavuusilmiön kehitystä ei voi esimerkiksi verrata aikapisteessä t_2 aikapisteeseen t_1 , jos tarkasteltavaa ilmiötä ei ole määritetty. Demokraattisesti toimivassa valtiossa perusteltua olisi, että kansalaiset tai heidän edustajansa määrittäisivät viime kädessä nämä vaikutukset sekä oikeat mittarit, jos mittarien määrittäminen korostaa tai rajaa pois tarkasteltavia vaikutuksia. Ei nimittäin vaikuta tarkoituksenmukaiselta, että esimerkiksi ei-demokraattisesti valitut Valtiontalouden tarkastusviraston vaikuttavuustarkastajat itse määrittäisivät sen, mitä vaikutuksia julkisen sektorin pitäisi yhteiskunnalle tuottaa. Asianmukaista ei liene sekään, että vaikuttavuustarkastaja ottaisi kantaa siihen, millä mittarilla kyseessä olevia vaikutuksia tulisi kuvata niissä tilanteissa, joissa mittareiden valinta edellyttää selkeästi poliittiseksi tulkittavia arvovalintoja. Nykyisen kuntalain hengen vastaista tuntuisi olevan myös esimerkiksi se, että tarkastuslautakunta alkaisi määritellä valtuustoa ja kunnanhallitusta ohjaavia vaikuttavuustavoitteita- ja mittareita (ks. L 365/1995; L 410/2015)

On turvallista väittää, että tarkastuskriteerien muodostaminen epäonnistuu, jos vaikutusten ja vaikuttavuusmittarien määrittäminen muodostuu ylipääsemättömäksi poliittiseksi ja/tai ammatilliseksi ongelmaksi. Poliittisten ja ammatillisten ongelmien takia vaikuttavuuden tarkastajat eivät yksinkertaisesti pysty johtamaan tarkastuksessa käytettyjä kriteereitä mistään ja praktiik-

kassa vaikutuksien tarkastaminen muodostuu mahdottomaksi. Tämän tyyppinen ongelma syntyy aina, kun julkinen sektori ei pysty määrittelemään niitä vaikutuksia, joita se haluaa kansalaisilleen tuottaa, eikä niitä vaikuttavuusmitareita, joiden avulla vaikuttavuutta tulisi seurata.

Julkisen sektorin vaikuttavuuteen liittyy paljon subjektiivisuutta ja vaikuttavuusilmiön käsitteellisessä määrittelemisessä ei aina löydetä konsensus-ta (ks. Kurunmäki & Miller 2011) muun muassa siksi, että vaikutukset ovat moniulotteisia, usein kvalitatiivisia luonteeltaan ja niitä on vaikeata kuvata yksittäisen kvantitatiivisen mittarin avulla (Carlin & Guthrie 2003). Vaikeudet vaikuttavuusilmiön käsitteellisessä määrittelemisessä ovat hyvin tiedossa monissa valtioissa (ks. Carlin & Guthrie 2003). Esimerkiksi Heinrich (2002) havaitsi, että yksi vaikeimmista tehtävistä on vaikuttavuusmittaussysteemin käsitteellinen määrittely Yhdysvaltojen liittovaltion julkisjohtajien mukaan.

Julkisella sektorilla on harvoin laajaa konsensus-ta siitä, mikä on toiminnan haluttu vaikutus, koska erilaiset sidosryhmät asettavat erityyppisiä odotuksia hyvinvointipalveluille (Smith 1996). On verrattain yleistä, että poliittiset puolueet kilpailevat keskenään erilaisten toimintapolitiikkaehdotusten ja poliittisten ongelmien avulla (vars. Spoon & Kluver 2014). Eriävät toimintapolitiikat ja poliittiset ongelmat näyttäisivät kytkeytyvät erilaisiin yhteiskunnallisiin vaikutuksiin, jolloin puolueilla voi olla hankaluuksia löytää yhteisesti tärkeitä vaikutuksia. Joskus poliittiset tavoitteet, kuten työttömyyden vähentäminen, kuulostavat saman vaikutuksen tavoittelulta, mutta tarkoittavat eri asiaa eri puolueille (Lindblom 1959). Tässä kappaleessa esitetyistä tieteellisistä argumenteista voimmekin päätellä sen, että erilaiset poliittiset tavoitteet vaikeuttavat yhteisten vaikuttavuustavoitteiden nimeämistä ja sitä kautta ilmiön käsitteellistämistä ja mittaamista.

Yhteisien vaikuttavuustavoitteiden nimeäminen ei ole vain poliittinen ongelma, sillä poliitikoilla ja virkamiehellä on usein myös eri näkemys toiminnan halutuista vaikutuksista (esim. Smith 1995; Agranoff & McGuire 2001). Se, mikä on manageriaalisesta näkökulmasta järkevä vaikuttavuustavoite, ei välttämättä ole sitä poliittisesti (ks. Ter Bogt 2004). Vaikuttavuustavoitteiden nimeämistä ja käsitteellistämistä hankaloittavat myös toimintaympäristössä

tapahtuvat muutokset ja tätä aspektia on varsin vähän, jos ollenkaan, problematisoitu systemaattisesti aiemmassa kirjallisuudessa. Ajankohdassa t₁ asetettu tavoitetaso voi tuntua hetkessä t₂ kohtuuttomalta, jos julkisen sektorin oman toiminnan merkitys vaikutukseen on tunnettujen tai tuntemattomien ulkoisten tekijöiden takia muuttunut olemattomaksi. Kaiken kaikkiaan edellä mainituista käsitteellistämisen ongelmista voidaan johtaa seuraavan tyyppinen johtopäätös: mittaaminen tai ylipäätään vaikuttavuusmittareiden kehittäminen muodostuu mahdottomaksi tehtäväksi jo lähtökohtaisesti, jos julkisella sektorilla ei ole yhteistä näkemystä siitä, mitä vaikutuksia tulisi seurata ja mitkä ovat tavoitteet vaikutusten osalta.

Myös vaikuttavuusmittarien valinta on arvolatautunutta toimintaa, jossa vaikuttavuusmittarit saattavat palvella yhtä intressiryhmää enemmän kuin jotain toista intressiryhmää (Smith 1996). Todellisuutta onkin se, että oikeista vaikuttavuusmittareista ei useinkaan löydetä konsensusta eri intressiryhmien välillä (Newcomer 2015). Jos konsensuksen puute aiheuttaa esteen vaikuttavuusmittarin luomiselle, on suhteellisen helppoa päätellä, että vaikuttavuusinformatiota ei synny, jos mikään mittari ei kyseistä informaatiota tuota. Lisäksi tarkastuskriteerien tutkiminen sekä tilivelvollisen toiminnan oikeutuksen toteaminen ei ole mahdollista vaikuttavuustarkastuksessa ilman vaikuttavuusinformatiota. On lisäksi hyvä tiedostaa, että vaikuttavuustavoitteiden suurpiirteinen ja epätarkka määrittäminen (vrt. esim. työttömyyden vähentäminen) jättää mittarien valintaan paljon poliittista latausta.

Erilaisten syytöksienvälttelystrategioiden (engl. blame avoidance strategies) käyttö poliittishallinnollisessa toimintaympäristössä ei yhtään helpota vaikuttavuustavoitteiden ja vaikuttavuusmittareiden määrittämistä (vrt. Hood 2007). Syytöksienvälttelystrategioiden käyttämisen perimmäinen motiivi on syytöksien välttäminen ja näiden strategioiden käyttäminen on verrattain yleistä julkisella sektorilla (Hood 2010; Weaver 1986). Syytöksienvälttelystrategioiden käyttö on suosittua (Hood 2010), koska esimerkiksi poliitikot voivat kokea suuria imagotappiota ja sitä kautta vaalihäviöitä, jos he joutuvat syntipukiksi huonoista tuloksista (Jones & McDermott 2004). Huonot tulokset voivat myös vähentää viraston saamaa rahoitusta ja sitä kautta palveluksessa olevaa henkilöstöä (Van de Walle & Van Dooren 2010). Julkisohtajien huole-

na on se, että poliitikot, intressiryhmät sekä media käyttävät vaikuttavuusinformaatiota hyökätäkseen julkisjohtajia vastaan (Wholey & Hatry 1992). Sekä julkisjohtajien että poliitikkojen kokema huoli on perusteltua (Hood 2007), sillä politiikassa on hyvin yleistä, että negatiivinen informaatio synnyttää enemmän toimintaa ja vaikutuksia kuin positiivinen informaatio (mm. Rozin & Royzman 2001). Myös media saattaa kärsiä negatiivisuusharhasta, jolloin helpommin raportoidaan huonoista tuloksista kuin hyvistä (Virtanen & Vakuri 2016).

Mittausmenetelmien vakiintuneisuus

Toiminnan vaikutusten syntyminen tai syntymättä jääminen on usein hankalampi todentaa ja mitata kuin toiminnan tapahtuminen tai tapahtumatta jääminen. Toiminta on joko tapahtunut tai jäänyt tapahtumatta. Tällöin tarkastus toteaa toiminnan tai toiminnan puutteen olevan joko asianmukaista tai asiatonta vertaamalla toimintaa ja toiminnan puutetta tarkastuskriteerien edellyttämään toimintaan. Toki toiminnan asianmukaisuuden tarkastamista saattaa haitata se, että toimintaa ei ole dokumentoitu mitenkään, mutta tässäkin tapauksessa tarkastaja voi huomauttaa tilivelvollista siitä, että dokumentointitoimintaa ei ole suoritettu. Tällöin toiminnan dokumentointi pitäisi tulla kuntoon sillä, että tarkastaja ohjeistaa tilivelvollista mittaamaan ja dokumentoimaan toimintansa hyväksi havaituilla metodeilla.

Toiminnassa havaituille puutteille, kuten esimerkiksi dokumentaation puuttumiselle, löytyy aina vastuuhenkilö, koska tilintarkastus ja laillisuustarkastus tarkastavat lähtökohtaisesti toimintaa, joka on jonkin toimijan vastuulla (esim. VTV 2010a; L 365/1995, 73 §; L 410/2015, 123 §). Tilintarkastus ja laillisuustarkastus eivät voi siis kohdistua ilmiöön, johon ei liity jonkun toimijan konkreettista vastuuta jostakin toiminta-alueesta tai vastuuta tehdä jotakin. Toiminnasta vastuussa oleva voi olla kollektiivi tai yksilö ja tätä vastuutahoa nimitetään yleisesti tilivelvolliseksi (ks. Mulgan 2003). Tässä tutkimuksessa tilivelvollisuuden käsite määrittyy kolmen ominaisuuden kautta (ks. Schedler 1999). Ensinnäkin, A on tilivelvollinen B:lle silloin, kun A on velvoitettu informoimaan B:tä omista menneistä ja tulevista päätöksistään ja

toimistaan. Toiseksi, tilivelvollisuuteen kuuluu myös se, että A:n tulee kyetä perustelevaan oikeutuksensa tehdä kyseiset päätökset ja toimet. Kolmanneksi, A:ta voidaan joko palkita tai rangaista riippuen siitä, onko hän toiminut oikeutetusti tai epäoikeutetusti. (Schedler 1999.) Tilintarkastus ja laillisuustarkastus pyrkivät siis kaikki varmentamaan sekä tilivelvollisen tuottamaa informaatiota että tilivelvollisen toimintaa ja sen oikeutusta. Vaikuttavuustarkastuksessa tarkastetaan tilivelvollisen tuottamaa vaikuttavuusinformaatiota, hänen toimintansa vaikutuksia ja mahdollisesti myös vaikutuksia aiheuttaneen toiminnan oikeutusta. Kaikissa näissä kolmessa tarkastusmuodossa voi kuitenkin toisinaan olla vaikeata hahmottaa sitä, kuka tai ketkä ovat tilivelvollisia (tästä ongelmasta jatketaan myöhemmin lisää).

Tilintarkastus- ja laillisuustarkastusevidenssiä voi hankkia seitsemällä erilaisella hyvin vakiintuneella tavalla: yksityiskohtaisilla tarkastuksilla, havainnoilla, ulkopuolisilla vahvistuksilla, uudelleen laskennalla, uudelleen suorittamisella, analyttisillä toimenpiteillä ja tiedusteluilla (Halonen & Steiner 2009; Ricchiute 2005, 206–207). Dataa voidaan kerätä muun muassa tilastotieteellisen otannan avulla tai ei-tilastotieteellisellä otannalla ja nämä molemmat menetelmät ovat yhtä hyväksyttäviä (Arens ym. 2005, 439). Tilintarkastusevidenssiä hankitaan tarkastuskohteen tilinpäätöksessä esitetyistä tiedoista ja niiden esittämistavasta, tilikaudella tapahtuneista liiketapahtumista ja muista tapahtumista sekä tilinpäätöksen tilien saldoista, jotta voidaan todeta tilinpäätöksen oikeellisuus (Tomperi 2009, 85–86). Tilintarkastuksessa käytetyt tiedonkeruumenetelmät nojaavat hyvin säänneltyihin tapoihin kerätä tietoja (Lindeberg 2007, 342). Laillisuustarkastuksessa on tilintarkastuksen menetelmät käytössä. Tilintarkastuksessa ja laillisuustarkastuksessa siis käytetään laajaa hyväksyntää nauttivia mittausmenetelmiä, joiden avulla pystytään suhteellisen hyvin varmistamaan tilivelvollisen toiminta ja sen oikeutus verrattuna vaikuttavuustarkastukseen.

Tarkastusevidenssin hankkiminen vaikutuksista on usein hankalampaan kuin tilintarkastuksen kohteena olevien toimintojen tarkastaminen johtuen mittausteknisistä haasteista. Mittaustekniset haasteet viittaavat tässä kontekstissa yhtäältä mittausteknologioiden sekä nykyisen mittausosaamisen puutteisiin ja toisaalta vaikuttavuusmittaamisen korkeisiin kustannuksiin.

Mittaushaasteita esiintyy siitakin huolimatta, että toiminnan vaikutuksien tarkastamisessa alkuasetelma näyttää hyvin samanlaiselta kuin toiminnan tarkastamisessa: vaikutuksia on joko syntynyt tai jäänyt syntymättä ja ne ovat joko asianmukaisia tai asiattomia tarkastuskriteerien mukaan. Vaikutuksien mittaamisen haasteet voivat kuitenkin aiheuttaa sen, että tarkastaja ei pysty toteamaan syntyneitä ja syntymättä jääneitä vaikutuksia kohtuullisella tai edes riittävällä (kohtalaisella) varmuudella. Vakavat vaikuttavuusmittaamisen ongelmat kykenevät myös aiheuttamaan sen, että tarkastaja ei pysty ohjeistamaan tilivelvollista mittaamisessa siinä määrin, että julkinen sektori pystyisi tuottamaan ja dokumentoimaan mittaustuloksia.

Yleiset mittaamisteoreettiset haasteet vaikeuttavat toiminnan vaikutuksien mittaamista, mikä väistämättä heijastuu myös tilivelvollisen aikaansaamien vaikutuksien tarkastamiseen. Aiempi kirjallisuus on tunnistanut mittausteoreettisiksi haasteiksi validiteetin- ja reliabiliteetin (Hand 2004), kompleksisuuden (Humphreys 2014; Smith 1995), vertailtavuuden (Andrew & Hildebrand 1976; Harris 1995; Cardy & Leonard 2011), vastauskadon (Miller & Smith 1983), vinoutuneen otoksen (Lee 2005) ja rajoitetun rationaalisuuden (ks. Meklin & Vakkuri 2006). Rajoitettu rationaalisuus tässä artikkelissa on ymmärretty niin, että ihminen ei ole kaikkietävä ja näin ollen hän ei aina kykene ratkaisemaan muita edellä mainittuja mittausteoreettisia haasteita, jotka voivat vaikeuttaa tuntuvasti tilivelvollisen toiminnan vaikutusten todentamista.

Yleisellä tasolla validiteetti ja reliabiliteettiongelmaa voidaan lähestyä kahdesta tulokulmasta (ks. Hand 2004, 129). Ensinnäkin voimme tarkastella sitä, onko mittausmenetelmä validi (eli pätevä) ja reliaabeli (eli luotettava). Vaikuttavuusmittaamisen vaikeutta pohtiva artikkelin osio pohtii juuri tätä. Toiseksi, voimme miettiä sitä, ovatko tuloksista johdetut päätelmät valideja ja reliaabeleja. Tällöin mietimme tulkinnan oikeellisuutta, jota pohdimme lisää myöhemmin tämän artikkelin siinä osiossa, joka pohtii vaikuttavuusinformaation tulkintaan liittyviä ongelmia.

Kompleksisuus ja kausaliteetit

Tilivelvollisuus vaikuttavuudesta voi kaatua vaikutuksien mittaamisen mahdolluuteen. Toiminnan vaikutukset voivat olla ilmiönä hyvin monitahoisia, jolloin monimutkaisuus voi yksinkertaisesti estää niiden validin ja reliaabelin mittaamisen (Smith 1996). Aiempi tutkimuskirjallisuus tuntee tapauksia, joissa sekä julkisen että yksityisen sektorin toiminnassa ei ole pystytty monimutkaisia vaikutuksia mittaamaan (Van der Valk & Iwaarden 2011; Newcomer 2015). Esimerkiksi ihmisten hyvinvointia on hyvin vaikea mitata kokonaisvaltaisesti (Conceição & Bandura 2008), sillä se pitää sisällään kaikki ihmiselämään liittyvät aspektit (McGillivray 2007). Ihmiselämään liittyviä laajoja kokonaisuuksia kuten ihmisoikeudet sekä ympäristötekijät (ks. esim. Summer 2006) ja näiden reliaabeli sekä validi mittaaminen on haastavaa. Kyvyttömyys mitata vaikutuksia estää sekä tilivelvollisen raportoinnin että vaikuttavuustarkastuksen toteuttamisen. Tällöin vaikuttavuustarkastuksessa voitaisiin ainoastaan todeta, että toiminnan vaikutuksia ei pystytty todentamaan eikä näin ollen tarkastamaan.

Vaikka vaikutukset saataisiinkin eksaktisti mitattua, voi tilivelvollisen toiminnan oikeutuksen todentaminen kaatua kausaliteettien reliaabeliin ja validiin tunnistamiseen sekä mittaamiseen. Kausaaliyhteyksien todentaminen julkisen sektorin toiminnan ja syntyneiden vaikutuksien välillä on usein problemaattista (Midwinter 2009; Hatry 2002). Problemaattisuus johtuu siitä, että vaikutuksia selittävästä julkisen sektorin ulkopuolisista tekijöistä on usein hankala saada informaatiota (Miller & Fox 2007) ja nämä tekijät voivat selittää syntyneitä vaikutuksia paljon paremmin kuin mikään mitattu julkisen sektorin toiminta (Anthony & Young 1988). Lisävaikeutta kausaliteettien mittaamiseen tuo myös se, että useat julkisen sektorin toimintaohjelmat linkittyvät toisiin toimintaohjelmiin ja lukuisiin toimintaohjelmien alaisiin alaohjelmiin (Diamond 2005) sekä erilaisiin toimintapolitiikoihin ja virastoihin (Newcomer 2015). Usein myös julkisen sektorin tuottamien palveluiden vaikutukset myös syntyvät yhteistyössä asiakkaiden kanssa (Talbot 2010) ja asiakkaiden toimintaa on huomattavasti hankalampi mitata kuin julkisen sektorin omaa toimintaa.

Vaikuttavuusmittaussysteemin pitäisi pystyä todellisuudessa erottelmaan validisti ja reliaabelisti eri toimintaohjelmien, toimijoiden ja julkisen sektorin ulkopuolisten tekijöiden kontribuutioarvo mitatuille vaikutuksille. Jos vaikuttavuusmittaristo ei pysty tätä erottelua tekemään, on tilivelvollisuuden todentajilla käsissään niin sanottu attribuutio-ongelma (Bovaird 2014). Attribuutio-ongelmassa on kyse siitä, että tarkastaja ei kykene arvioimaan tilivelvollisen toiminnan merkitystä vaikutuksille (ks. Mayne 2007). Attribuutio-ongelman ytimessä on tietämättömyys siitä, kenen tai keiden toimijoiden toiminta on saanut aikaan havaitun vaikutuksen (Talbot 2010, 48). Täsmällisten kausaaliteettisuhteiden tunnistamattomuus johtaa reaali maailmassa aina jonkinasteiseen attribuutio-ongelmaan (Bovaird 2014). Vaikuttavuustarkastuksessa tehdyt johtopäätökset ja niiden oikeudenmukaisuus ovat kyseenalaisia silloin, kun attribuutio-ongelma on läsnä tilivelvollisen toiminnan vaikutusten todentamisessa. Sekä laillisuustarkastuksessa että tilintarkastuksessa attribuutio-ongelma kytkeytyy lähinnä valvontavastuun asianmukaisen suorittamisen todentamiseen, jolloin pohditaan sitä, mikä vastuu valvontavastuussa olevalla taholla on inhimillisen virheen, laiminlyönnin tai petoksen syntymiseen. Muuten kausaalisyhteyksien todentaminen toiminnan ja siitä vastuussa olevan tilivelvollisen välillä on suhteellisen ongelmatonta näissä tarkastusmuodoissa.

Ongelmat kausaaliteetin tunnistamisessa heijastuvat myös vertailtavuuteen ja sitä kautta toiminnan vaikuttavuuden reliaabeliin ja validiin arvioimiseen. Yleisesti ottaen vaikuttavuustuloksia voi vertailla julkisen organisaation omaan historiaan, asetettuun tavoitetasoon, tai verrokkiryhmän kehitykseen nähden (Kates, Marconi & Mannle 2001). Mittaamatta jääneet julkisen sektorin ulkopuoliset muuttujat voivat kuitenkin rapauttaa vertailun, jos kyseisten muuttujien kontribuutioarvo havainnoituihin vaikutuksiin vaihtelee merkittävästi ajassa tai vertailtavien yksikköjen välillä (ks. Scutchfield & Keck 2003, 209). Vertailuasetelman perusteella annetut sanktiot tai palkkiot voivat olla epäoikeudenmukaisia silloin, kun vertailuasetelma on pahasti vääristynyt ulkoisten muuttujien takia. Tilintarkastuksessa ja laillisuustarkastuksessa toimintaympäristön muutokset harvoin estävät tilivelvollista toimimasta lakien ja säädösten mukaisesti, joten toiminnan vertailtavuus tarkastuskriteereihin on hyvin varmistettu.

Vaikuttavuusilmiön emergentti perusluonne (Funnell & Rodgers 2011) asettaa omat haasteensa tilivelvollisen toiminnan oikeutuksen todentamiselle. Vaikutuksien kompleksinen ja emergentti luonne aiheuttaa usein sen, että vaikutuksien syntymekanismia ei tarkalleen ottaen tunneta ja osata ennustaa. Verrattaessa esimerkiksi tilintarkastukseen, jonka tärkeimpiä tehtäviä oli petosten tai inhimillisten virheiden aiheuttamien merkittävien virheiden havaitseminen (Issai 200), on vaikutusten mittaaminen huomattavasti vaikeampaa. Tämä johtuu siitä, että staattisimpina ilmiönä sekä petos että inhimillinen virhe ovat paremmin tunnettuja ja ymmärrettyjä kuin vaikuttavuus. Summaten voisi todeta, että meillä on olemassa vakiintunut määritelmä sille, mitä petos, inhimillinen virhe ja lain rikkominen ovat ja tiedämme myös miten ne syntyvät, mutta emme valitettavan usein tiedä sitä, miten vaikutus pitäisi määritellä, ja kuinka vaikutukset oikein lopulta syntyvät.

Vastauskato ja otosvinoumat

Vaikutukset voivat olla perusluonteeltaan kvalitatiivisia (Carlin & Guthrie 2003) ja kvalitatiivista dataa usein kerätään kyselytutkimusten avulla. Vastauskato on yksi tyypillisimmistä ongelmista, joka vaarantaa kyselytutkimusten validiteetin ja reliabiliteetin (Bethlehem, Cobben & Schouten 2011). Bethlehem ym. (2011) mukaan on olemassa kahden tyyppistä vastauskatoa:

1. vastaajakato on kyseessä silloin, kun kyselytutkimukseen valittu henkilö ei vastaa ollenkaan kyselyyn ja
2. asiakohtainen kato kuvaa sellaista tilannetta, jossa vastaaja on joihinkin kysymyksiin vastannut ja joihinkin jättänyt vastaamatta.

Tilivelvollisen toiminnan vaikutusten todentaminen saattaa olla hyvin hankalaa, jos vastauskato heikentää mittaustuloksen validiteettia ja reliabiliteettia. Vastauskadon ollessa tarpeeksi suuri, voivat saadut kyselytulokset olla täysin sattumaa. Tässä tapauksessa tilivelvollisen toiminnan vaikutuksista ei voida sanoa oikeastaan mitään, koska reliaabeli ja validi todistusaineisto kannanotolle puuttuu. Vastauskadon ongelma on todellinen, kun vaikuttavuusinformaation onnistunut kerääminen riippuu julkisen sektorin ulkopuolisten sidosryhmien vastausinnokkuudesta. Kyselytutkimusten otoksen edus-

tavuuden vaarantaa myös se, että julkinen sektori tai vaikuttavuustarkastaja ei osaa identifioida oikeita ihmisiä, joilta palvelujen vaikutuksia pitäisi pystyä tiedustelemaan. Tällöin esimerkiksi kyselytutkimuksen otos on vinoutunut niin pahasti, että kyselyn tulokset eivät heijastele validisti ja reliaabelisti relevantin kohdepopulaation näkemyksiä palvelujen vaikutuksista (ks. Lee 2001). Tähän niin sanottuun identifikaatio-ongelmaa törmätään usein julkisen sektorin vaikuttavuusmittaamisessa (Smith 1996) ja sen aiheuttamat epäluotettavat tulokset vaarantavat tilivelvollisen toiminnan oikeutuksen todentamisen vaikuttavuustarkastamisessa.

Tilintarkastuksessa ja laillisuustarkastuksessa ei käytetä laajoja kyselytutkimuksia. Kun tilintarkastuksessa ja laillisuustarkastuksessa käytetään otantaa, tarkastellaan aina fyysisiä asioita kuten pankkitilille panoja, myyntilaskuja, velkojen balansseja tai rahayksiköitä (ISA 530). Fyysisiä yksiköitä tarkasteltaessa identifikaatio-ongelmaa ei esiinny, sillä tarkastettava populaatio on hyvin tarkastajan tiedossa. Myös erilaiset otosvinoumat on helpompi arvioida, kun identifikaatio-ongelma ei aiheuta päänvaivaa.

Mittaamisen kustannukset ja nykyiset informaatiojärjestelmät

Tilivelvollisuuden reliabeli ja validi todentaminen aikaansaatuisten vaikutuksien osalta voi kaatua mitaamisen aiheuttamiin kustannuksiin. Vaikutuksien mitaamisen sekä arvioinnin kustannukset voivat olla hyvin korkeat (Kristensen ym. 2002; Lowe 2013). Kaikkien relevanttien muuttujien mitaaminen pitkällä aikavälillä vaatii useasti täsmällisiä kvalitatiivisia sekä kvantitatiivisia mitausmenetelmiä (Schalock 2001), jolloin kustannukset yleensä nousevat tuntuvasti (Lowe 2013). Korkeat kustannukset vaikuttavuustarkastuksessa voivat tarkoittaa sitä, että tarkastus tuottaa enemmän kustannuksia kuin hyötyjä yhteiskunnalle, jolloin koko toiminnan suorittamisen mielekkyys voidaan asettaa kyseenalaiseksi (ks. Needles & Powers 2007). Kokonaisvaltaisessa tilintarkastuksessa ja laillisuustarkastuksessa voidaan olettaa olevan alhaisemmat kustannukset jo lähtökohtaisesti kuin kokonaisvaltaisessa vaikuttavuustarkastuksessa, koska mitattavia muuttujia on paljon vähemmän. Toiminnan

yksikertaisuus verrattuna vaikutuksien monimutkaisuuteen on kustannuksia keventävä tekijä tässä ajattelussa.

Jos laajamittaiset menetelmät vaikutuksien seuraamiseksi valitaan, voivat julkisen sektorin työntekijät joutua omistamaan enemmän aikaansa raportoinnille. Tällöin heille jää itseasiassa vähemmän aikaa hyvinvointituotteiden- ja palveluiden tekemiselle (ks. Keevers, Treleaven, Sykes & Darcy 2012). Tilivelvollisen toiminnan oikeutuksen todentamiseen johtava raportointi sekä muu selvitystyö toiminnan perusteluista saattavat siis laskea tilivelvollisen tuottamien palvelujen määrää ja laatua. Kun näin tapahtuu, puhutaan niin sanotusta tilivelvollisuusparadoksista (Dubnick 2005), joka pitäisi myös vaikuttavuustarkastuksessa osata ottaa huomioon. Yhtenä tilivelvollisen vaikuttavuustarkastamisen mahdollistamisen haasteena onkin siis tilivelvollisuusparadoksin tunnistaminen ja välttäminen. Tilintarkastuksessa ja laillisuustarkastuksessa paradoksi uhka ei ole niin suuri kuin vaikuttavuusmittaamisessa, koska toiminta on usein ilmiönä paljon yksinkertaisempi kuin toiminnan seurauksena syntyneet vaikutukset.

Laajamittaisen ja kalliin vaikuttavuusmittausjärjestelmän sijaan voidaan myös valita vain joitakin mittareita seuraamaan syntyneitä vaikutuksia, jotta mittaamisen kustannukset eivät kohoaisi liiaksi (Lowe 2013). Mittarien rajoittamisen mukana tulee kuitenkin seuraavan tapainen ongelma: jokainen rajaus voi merkittävästi vaikuttaa tilivelvollisen toiminnan oikeutuksen arvioimiseen. Sama toiminta voi näyttää sekä oikeutetulta että epäoikeutetulta yhtä aikaa riippuen siitä, millä vaikuttavuusmittarilla asiaa tarkastellaan (ks. Swanson 2012). Vaikuttavuusmittari valinta voi siis ratkaista sen, näyttääkö tilivelvollisen toiminta oikeutetulta vai epäoikeutetulta. Ongelmaksi voi siis muodostua mittarin valinnan tekeminen tai mittarin legitimizeettivaje eri toimijoiden tekemisissä tulkinnoissa, jolloin vaikuttavuustarkastuksen luotettavuus voidaan asettaa kyseenalaiseksi. Se seikka, että tulkintaan vaikuttavuusmittarien tuottaman informaation tärkeydestä vaikuttaa aina se, kuka toimija asiaa tarkastelee ja missä asiayhteydessä (ks. Kunda 1990; Van Maanen & Schein 1979), ei yhtään helpota olennaisien vaikuttavuusmittarien valitsemista ja epäolennaisien mittarien poissulkemista.

Nykyiset entiteettiperustaiset ja oikeussubjektipohjaiset informaatiojärjestelmät perustuvat perinteisiin linjaorganisaatioihin ja näiden informaatiojärjestelmien kehittäminen tapahtuu entiteettien sisäisinä projekteina (esim. Ryan & Walsh 2004). Entiteettiperusteisia informaatiojärjestelmiä voidaan perustella sillä, että suoritteet tuotetaan ja määrärahat käytetään juuri näissä entiteeteissä (ks. Mayne 2007). Entiteettipohjaiset informaatiojärjestelmät palvelevat hyvin tilintarkastusta sekä laillisuustarkastusta, koska molemmat tarkastusmuodot tarkastavat toimintaa, joka on rajattavissa aina yksittäisen entiteetin vastuulle (ks. lisää tietoa tarkastuskohteista L 1141/2015).

Vaikuttavuustarkastuskin voi pohjautua entiteetteihin silloin, kun vaikutuksen syntyminen on yksittäisen entiteetin toiminnan tulosta. Usein vaikutukset kuitenkin syntyvät usean entiteetin ponnistelujen tuloksena ja esimerkiksi poikkisektorillisen yhteistyön kautta, jolloin vaikutuksen tuottamiseen on osallistunut useita entiteettejä eri hyvinvointipalvelusektoreilta. Poikkisektorillisten vaikutusten seuraaminen vaatisi kuitenkin informaatiojärjestelmän, joka seuraisi ja raportoisivat vaikuttavuus- ja kustannusinformaatioita yli entiteettirajojen (Hodges 2012). Poikkisektorillista ja yli entiteettirajojen operoivaa vaikuttavuustarkastusta rajoittaa siis merkittävästi nykyiset entiteettiperustaiset informaatiojärjestelmät. Lisäksi entiteettiperustainen informaatiojärjestelmä pystyy usein vain rajallisesti seuraamaan julkisen sektorin ulkopuolisia tekijöitä kuten esimerkiksi yksityisen sektorin palvelutuottajien toimintoja sekä vaikutuksia.

Vaikuttavuusinformaation tulkinta todentamisen haasteena

Vaikuttavuusinformaation todentamisen haasteet ovat jaoteltu tässä tekstissä kolmeen kategoriaan: epätäydellisen informaation tulkintaan liittyviin haasteisiin, informaatiotulvaan sekä vaikuttavuusinformaation selektiivisen valikoinnin aiheuttamiin komplikaatioihin.

Epätäydellisen mittausinformaation tulkinta

Käytännössä vaikutusten tulkitseminen on usein haastavaa (Smith 1996). Haastavuus johtuu monimutkaisista kausaalisuhteista, vastauskadon ja otos-

vinoumien arvioimisesta, vaikutusviiveestä ja mahdollisen tilivelvollisuusparadoksin arvioimisesta. Monimutkaiset kausaalisuhteet viittaavat tässä siihen, että on vaikea hahmottaa täsmällisesti sitä, kuinka julkisen sektorin tuotokset ja julkisen sektorin ulkopuoliset tekijät ovat vaikuttaneet havaittuun vaikutukseen (ks. esim. Mascarenhas 1996). On usein epäselvää, mikä on eri ohjelmien, toimintapolitiikkojen ja virastojen merkitys sekä kytkös havaittuihin vaikutuksiin (Newcomer 2015). Tämä epäselvä kytkös julkisen sektorin toiminnan ja sen vaikutusten välillä synnyttäisi jossain määrin spekulatiivisen asetelman tilivelvollisuuden toiminnan oikeutuksen todentamiseen ja tämä väistämättä vaikuttaisi käsitykseen tarkastuksen oikeudenmukaisuudesta. Tilintarkastuksessa ja laillisuustarkastuksessa kausaliitteongelmat vaivaavat oikeastaan vain valvontavastuun selvittämiseen liittyviä tilanteita.

Edellä kuvatut mittaamisen ongelmat tekevät vaikuttavuustuloksien tulkinnan ongelmalliseksi, joka väistämättä heijastaa myös aina jollain tasolla toiminnan oikeutuksen todentamisen laatuun. Jos esimerkiksi vaikuttavuusmittari tai vaikuttavuuden mittaaja ei ole arvioinut syystä tai toisesta vastauskadon merkitystä mittaustiedon reliabiliteetille ja validiteetille, voidaan tilivelvollisuuden osalta tehdä virheellisiä johtopäätöksiä. Ongelma voi olla myös se, että tilivelvollisuuden todentajat, kuten tilintarkastaja tai kunnan valtuusto, eivät osaa tulkita mittaustietoon liittyvän vastauskadon merkitystä tai mittaustietoa ja vaikuttavuuden mittaajat eivät välitä tilivelvollisuuden todentajille riittävästi tietoa vastauskadosta tai sen merkityksestä. Käytännön ongelmaksi voi hyvin muodostua se, että ilman oikeutusta toiminut viranhaltija voi säästyä sanktioilta ja vastaavasti oikeutetusti toiminut viranhaltija voi joutua aiheettomien sanktioiden kohteeksi, jos tilivelvollisen toiminnan oikeutuksen todentaminen perustuu vastauskadon turmelemaan tietoon. Aiheettomat sanktiot ja palkkiot luovat epäoikeudenmukaisuutta tilivelvollisten aktuaaliseen toimintaympäristöön. Lisäksi läpinäkyvyys kansalaisten suuntaan heikkenee toiminnan tuloksista, jos vastauskadon ja otosvinauman merkitystä ei tiedoteta. Kansalaiset voivat siis tehdä äänestyspäätöksiä väärin perustein ja sitä kautta muun muassa poliittinen tilivelvollisuus toimintamekanismina voi vääristyä reaali maailmassa. Tilintarkastuksessa ja laillisuustarkastuksessa ei vastauskadon ongelmaa esiinny juurikaan, jolloin myös näiden

tarkastusmuotojen raporttien tulkitseminen ei muodostu samalla tapaa ongelmaksi.

Vaikutusviiveellä viitataan siihen aikaväliin, joka on varsinaisen toiminnan ja siitä syntyvän vaikutuksen välillä. Seurausten tarkastamista vaikeuttaa seurauksien syntymisajankohdan määrittäminen ja tilivelvollisuuden kiinnittyminen tilikauteen. Tilivelvollisen toiminnan oikeutuksen tarkastelu voi vääristyä jo pelkästään sen perusteella, missä ajankohdassa vaikutukset pääteetään ottaa tarkastelun alle. Teoreettisesti tarkasteltuna vaikutukset jaoteltiin kolmeen eri kategoriaan aiemmin tässä artikkelissa: välittömiin vaikutuksiin, keskipitkän aikavälin vaikutuksiin ja lopullisiin vaikutuksiin (Vedung 1997). Vedungin (1997) teoreettinen kategorisointi demonstroi ennen kaikkea sitä, että vaikutukset voivat syntyä eri ajankohdissa. Yksi vaikutusten todentamisen haasteista onkin se, missä ajankohdassa tulkinta vaikutusten saavuttamisesta tulisi tehdä. Oikean ajankohdan tunnistaminen ei ole helppoa ja väärän ajankohdan valitseminen voi johtaa vääriin johtopäätöksiin tilivelvollisen toiminnan vaikutuksista. Liian varhaisen ajankohdan valinta tarjoaa mahdollisuuden sille, että tilityksen vastaanottajat sanktioivat tai palkitsevat liian aikaisin tilivelvollista. On esimerkiksi vaikea perustella jälkikäteen sitä, että tilivelvollista sanktioitiin saavuttamatta jääneistä positiivista vaikutuksista ajankohdassa t_1 , jos tilivelvollisen toiminnan aikaansaamat positiiviset vaikutukset syntyivätkin vasta myöhemmin ajankohdassa t_2 . Myös toiminnan palkitseminen saattaa olla ajankohdassa t_1 kyseenalaista, jos tilivelvollisen toiminnan positiiviset vaikutukset kumoutuvat myöhemmin ajankohdassa t_2 kyseisen toiminnan aiheuttamalla negatiivisilla vaikutuksilla.

Tilivelvollisuuden todentamisen lykkääminen pitkän ajanjakson päähän ei ole sekään täysin ongelmatonta, sillä sekä tilivelvollinen että tilintarkastaja joutuisivat tulkitsemaan pitkällä aikavälillä tapahtunutta vaikutusten kehityskulkua, joka voi olla hyvin monimutkainen ilmiö. Mahdollisen informaatiohälyn lisäksi tilivelvollinen joutuisi todellisuudessa odottamaan koko tämän ajan vastuuvapauden myöntämistä, mikä voi olla jo itsessään kohtuuton lähtökohta. Toiminnan ja sen vaikutusten syntyminen välinen pitkä ajanjakso nostaaakin esille seuraavan tapaisen vaikean kysymyksen: kuinka pitkän aikaa tilivelvollinen on yhtäältä velvollinen raportoimaan toimintansa vaikutus-

ta ja toisaalta vastuussa toimintansa vaikutuksista? Näyttääkin siis siltä, että vaikuttavuustarkastuksessa tarkastusajankohdan määrittäminen kykenee aiheuttamaan ilkeän ongelman tarkastajalle, tilivelvolliselle ja tarkastuksen tilaajalle eli julkiselle sektorille.

Tilikausajattelu aiheuttaa suurempia haasteita vaikuttavuustarkastuksessa kuin laillisuus- ja tilintarkastuksessa. Sekä tilintarkastuksessa että laillisuustarkastuksessa tarkastettava toiminta on sijoitettavissa sille tilikaudelle, jolloin toiminta tapahtui. Tilintarkastusta ja laillisuustarkastusta suoritettaessa tiedämme useimmiten sen, milloin toimintaa on tapahtunut ja milloin ei. Vaikuttavuustarkastukseen usein sisältyy paljon enemmän epävarmuutta siitä, milloin toiminnan vaikutus realisoituu. Hankalaksi vaikuttavuustarkastuksen tekeekin se, että toiminta ja sen vaikutukset voivat syntyä aivan eri tilikausilla ja etukäteen on vaikea ennustaa sitä, millä tilikaudella vaikutukset tulevat näkyviin. Epävarmuus kykenee syömään vaikuttavuustarkastuksen tuloksien validiteettia ja reliabiliteettia.

Jo vaikutusten mittaamisen yhteydessä esiin nostettu tilivelvollisuusparadoksi aiheuttaa myös omat ongelmansa. Tilityksen vastaanottajat, jotka eivät kykene tulkitsemaan tilivelvollisuusparadoksin merkitystä vaikutusten syntymiselle, saattavat arvioida väärin tilivelvollisen toiminnan oikeutuksen. Tulkinta tilivelvollisuusparadoksin merkityksestä on erityisen haastavaa silloin, kun raportoinnin merkitystä hyvinvointipalvelutuotannolle ja sitä kautta syntyneille vaikutuksille ei ole mitattu. Tilivelvollisen sanktioimista tilanteessa, jossa raportointivelvoitteet ovat aiheuttaneet havaitut negatiiviset vaikutukset, ei voida pitää hyväksyttävänä käytäntönä.

Informaatiotulva ja vaikuttavuusinformaation selektiivinen valikointi

Informaatioähkyn on todettu aiheuttavan ongelmia tilintarkastuksessa (Cornelius 1980; Chewning & Harrell 1990; Chung & Monroe 2001) ja laillisuustarkastuksessa (esim. Rudolph & Welker 1998; Nielsen & Madsen 2009). Koska toiminnasta syntyneet vaikutukset ovat toimintaa monimutkaisempi ilmiö, kykenee kattava vaikuttavuustarkastusevidenssi tuottamaan haitalli-

semman informaatioöhkyn kuin tilintarkastus- tai laillisuustarkastusevidenssi, jolloin myös tuloksien tulkinta vaikeutuu.

Yleisesti ottaen päätöksentekijän kokeman informaatioöhkyn on havaittu huonontavan päätöksenteon laatua (Hahn, Lawson & Lee 1992). Vaikuttavuustarkastukseen liittyvä päätöksenteko saattaa vaarantua, jos vaikuttavuustarkastaja kärsii informaatioöhkystä. Toisaalta myös tilityksen laatu voi olla huono, jos tilivelvollinen on kärsinyt informaatioöhkystä tilitystä antaessaan. Huonolaatuinen tilitys jo lähtökohtaisesti heikentää vaikuttavuustarkastajan mahdollisuuksia arvioida validisti, reliabelisti ja oikeudenmukaisesti tilivelvollisen toiminnan oikeutusta. Informaatioöhkyn merkitys korostuu, mitä moniulotteisimpia vaikutuksia tarkastelemme tilivelvollisuuden todentamisen yhteydessä. Nykyisen kaltaisessa globaalissa ja monimutkaisessa maailmassa on vaikeata ennaltaehkäistä informaatioöhkyn syntyminen vaikuttavuustarkastuksessa ilman, että samalla ei aiheuttaisi informaation valikoinnista aiheutuvia ongelmia.

Päätöksenteossa informaatioöhkyltä voi toki välttyä valikoimalla suuresta tietomassasta vain osan, jota sitten käytetään päätöksenteon tukena (Simon 1979). Koska vaikuttavuustarkastajan tehtäväksi ei voida lukea suotuisien yhteiskunnallisten vaikutusten määrittämistä, jää vaikutusten valikointi poliittisille toimijoille ja todennäköisesti jossain määrin praktiikassa myös virkamiehille. Poliittisten toimijoiden tai virkamiesten suorittama informaation valikointi on kuitenkin jossain määrin kyseenalainen lähtökohta vaikuttavuustarkastukselle. Valikointia hankaloittaa se, että ihmisillä on taipumus etsiä ja käyttää sellaista tietoa, joka tukee heidän aiempia uskomuksia ja arvojaan (esim. Taber & Lodge 2006). Jos esimerkiksi poliittisesti motivoitunut tiedon etsintä toteutuu tilivelvollisuuden todentamisessa, saattavat annetut sanktiot tai palkkiot perustua aiempien uskomusten ja arvojen pohjalle eivätkä niinkään siihen viestiin, jonka vaikuttavuusinformaatio kokonaisuudessa kertoo tilivelvollisen toimien oikeudenmukaisuudesta. Kun poliittiset toimijat tai virkamiehet määrittävät tarkastusinformaation määrittelemällä tavoitteen ja mittarit, voimmeko puhua objektiivisesta vaikuttavuustarkastuksesta kansalaisten tai kuntalaisten näkökulmasta katsottuna? Tosioloissa saattaisikin ongelmaksi muodostua se, miten virkakoneisto ja poliitikot kehystäisivät kuvan

julkisen sektorin vaikuttavuudesta suotuisia mittareita valikoimalla. Tämä vaarantaisi hyvinvoinnin kehittymisen yhteiskunnassa, koska negatiiviset asiat eivät tulisi kansalaisten tietoisuuteen ja tämä saattaisi vaikuttaa äänestyskäyttäytymiseen epäsuotuisalla tavalla.

Tarkastajan on vaikea olla objektiivinen, jos hänelle tuotettu informaatio on lähtökohtaisesti puolueellista. Toisaalta vaikuttavuustarkastajan on hankala ottaa kantaa mittarin oikeellisuuteen, jos kannanottoon liittyy poliittisia arvovalintoja. Ongelmallista vaikuttavuustarkastamisen näkökulmasta on myös se, että kenen pitäisi tehdä tulkinta käytettävästä informaatiosta silloin, kun eri vaikuttavuusmittarit antavat ristiriitaisen ja jopa täysin päinvastaisen kuvan julkisen sektorin vaikuttavuudesta. Jos politiikan aiheuttamat negatiiviset vaikutukset ovat ennustettavissa jo etukäteen, valitsevatko poliittiset toimijat sellaisia mittareita, jotka myös tuovat vaikuttavuustarkastajan tietoon nämä vaikutukset? Kun annamme arvioitavan tahon eli luottamushenkilöiden ja/tai virkamiesten päättää vaikuttavuustarkastuksessa käytetyt vaikuttavuusmittarit, saammeko tietoon aiheutuneita negatiivisia vaikutuksia? Tiedon valikointi voi aiheuttaa myös ongelmia tilivelvollisen oikeusturvan kannalta katsottuna, jos kokonaisvaikutuksista vain valikoidaan joitakin vaikutuksia tarkastelun alle. Miten turvaamme reilun käsittelyn tilivelvollisille näissä tosielämän tilanteissa?

Vaikuttavuustarkastuksen merkitys tilivelvollisten oikeudenmukaiselle kohtelulle

Tilivelvolliset voidaan asettaa eriarvoiseen asemaan, jos joillekin tilivelvollisille asetetaan vain suoritteisiin, toiminnan lainmukaisuuteen, toimintaprosesseihin ja/tai määrärahojen käyttämiseen liittyviä tilivelvollisuusvelvoitteita ja joillekin toisille julkisen sektorin toimijoille määritetään tilivelvollisuusvaikutuksista. Epätasa-arvo johtuu siitä, että suoritteiden sekä toimintaprosessien tuottamista, toimintojen lainmukaisuuden varmistamista ja määrärahojen käyttöä tilivelvollinen pystyy hallitsemaan paremmin kuin toiminnan tuloksena syntyneitä vaikutuksia (esim. Irwin 1996). Kyvyttömyys mitata joidenkin julkisen sektorin toimialojen vaikutuksia ollenkaan voi käytän-

nössä estää tilivelvollisuuden kytkemisen syntyneisiin vaikutuksiin kyseessä olevilla toimialoilla. Tilivelvollisuuden kytkeminen vaikutuksiin epäonnistuu myös silloin, kun vaikuttavuusmittaaminen ei kykene ratkaisemaan edellä esitettyä attribuutio-ongelmaa. Jos jollakin julkisen sektorin toimialueella esiintyy attribuutio-ongelmaa tai mittauskyvyttömyyttä, voi koko julkisen sektorin vaikuttavuustarkastussysteemi rapautua epätasa-arvoisuutensa takia. Epätasa-arvoa syntyy, jos tällaisissa tilanteissa osalle asetetaan tilivelvollisuus toiminnasta ja osalle vaikutuksista riippuen siitä, voidaanko toimijoiden vaikutuksia mitata. Toisaalta epätasa-arvoa aiheuttava vaikuttavuustarkastussysteemi voi myös helposti jäädä kokonaan implementoimatta, jos tiedossa on jo etukäteen, että kyseinen systeemi luo epätasa-arvoa julkisen sektorin työntekijöiden välille.

Vaikka kaikille julkisen sektorin toimijoille saataisiinkin tilivelvollisuus vaikutuksista, voi tämän tapainen tilivelvollisuussysteemi tuntua epäoikeudenmukaiselta johtuen vaikutuksien hallitsemattomasta luonteesta. Tilintarkastuksessa ja laillisuustarkastuksessa ei synny epätasa-arvoa tai epäoikeudenmukaisuuden tunnetta eri tilivelvollisten välille samassa määrin kuin vaikuttavuustarkastuksessa, koska jokainen tilivelvollinen on näissä tarkastusmuodoissa vastuussa omasta toiminnastaan, joka on tilivelvollisten hallinnassa huomattavasti paremmin kuin haluttujen vaikutusten synnyttäminen.

Yhteenveto

Elämme maassa, jossa julkiselta sektorilta odotetaan haluttujen vaikutuksien synnyttämistä ja vaikuttavuusinformaation tuottamista. Vaikuttavuustarkastamisen nykytila näyttää kuitenkin tutkimuskirjallisuuden valossa siltä, että vaikuttavuustarkastuksia tehdään kansainvälisesti ja kansallisesti katsottuna hyvin vähän verrattuna tilintarkastuksiin ja laillisuustarkastuksiin. Koska nykyinen tuloksellisuuskirjallisuus ei ole systemaattisesti tarkastellut vaikuttavuustarkastuksen haasteita ja tutkinut näiden haasteiden ilmenevyyttä tilintarkastuksessa ja laillisuustarkastuksessa, lähdettiin tässä artikkelissa tutkimaan sitä, mitkä vaikuttavuustarkastuksien haasteista ilmentyvät Suomen kontekstissa samanlaisina myös tilintarkastuksissa ja laillisuustarkastuksissa

ja mitkä eivät. Tutkimuskysymystä lähestyttiin vaikuttavuustarkastamisen haasteiden kautta. Paikantamalla ensin vaikuttavuustarkastusta vaivaavat kompastuskivet kansainvälisestä tutkimuskirjallisuudesta ja vertaamalla näitä kompastuskiviä Suomessa suoritettavaan tilintarkastukseen ja laillisuustarkastukseen, yritettiin löytää niitä mahdollisia haasteita, jotka erottavat vaikuttavuustarkastuksen tilintarkastuksesta ja laillisuustarkastuksesta.

Toteutetun kirjallisuuskatsauksen ja meta-evaluaation avulla löytyi lukuisia mahdollisia haasteita eli toisin sanoen tekijöitä, jotka voivat selittää vaikuttavuustarkastuksien vähäisen lukumäärän. Nämä tekijät ovat kerätty kootusti taulukkoon 1. Kuten taulukosta 1. voidaan havaita, on mahdollisia syitä vaikuttavuustarkastusten vähäiselle määrällä monia. Tutkimuksen keskeisimmäksi tuloksi nousi neljä isompaa kokonaisuutta, jotka hankaloittavat vaikuttavuustarkastusta. Nämä haasteet liittyvät:

1. tarkastuskriteerien muodostamiseen (ei poliittista konsensusta halutuista vaikutuksista),
2. vaikuttavuusmittaamisen tekniseen toteuttamiseen (attribuutio-, vertailtavuus, kustannusongelma kompleksisuus- ja informaatiojärjestelmäongelma),
3. vaikuttavuusinformaation tulkintaan (epätäydellinen mittaustieto, informaatiohäky ja haitallinen vaikuttavuusinformaation välittäminen) ja
4. vaikuttavuustarkastuksen aiheuttamaan epätasa-arvoon (vaikutuksia ei voi useinkaan hallita tai mitata samaan tapaan kuin omaa toimintaansa, jolloin vaikuttavuustarkastus voi kärsiä legitimititeettivajeesta, koska se voi aiheuttaa pahimmillaan tilivelvollisten epäoikeudenmukaista kohtelua).

Koska tämä tutkimus oli ensimmäisiä askeleita vaikuttavuustarkastuksen haasteiden kokonaisvaltaisessa kartoittamisessa ja ymmärtämisessä, sisältyy taulukossa 1 oleviin syihin luonnollista epävarmuutta. Siksi on syytä korostaa, että tämän tutkimuksen merkittävimmät löydökset avaavat ennen kaikkea uusia ovia sekä empiirisille että teoreettisille tutkimuksille. Ensinnäkin taulukon 1. kokoamat haasteet odottavat empiiristä testaamista ja toimivat näin ollen mahdollisina tutkimushypoteeseina tuleville vaikuttavuustarkastusta

Taulukko 1. Vaikuttavuustarkastuksen kootut haasteet

Tarkastustyyppi	Mitä tarkastus tulkitä pääasissa ja onko tarkastuskohde stabiili ja mitattavissa?	Onko tarkastuskriteerien muodostamisessa yleensä ongelmia?	Onko tarkastuskohteesta ja mittausmenetelmästä konsensusa?	Vaivako attribuutio-ongelma tarkastusta mitaamisessa ja tuosten tulkitsemisessä?	Kärsiikö tarkastus vastauskäden ja vinoutuneen jakautuman ongelmista tai vertailtavuus-ongelmista?	Voiko tarkastaminen tuntua epäoikeudenmukaiselta tilivelvoisen mielestä?	Tukevatko nykyiset informaatiojärjestelmät tarkastamista?	Rastitaako informaatio-ähy tai informaation valikoitua tarkastamista?
Tilintarkastus	Toiminta ja tarkastuskohde on suhteellisen stabiili (esim. petoksen syntymekanismi on stabiili sekä tiedossa ja fyysisiä toimintaa on myös helpompi mitata kuin vaikutuksia)	Ei, koska lait ja muut normatiiviset dokumentit ovat jo olemassa ja niiden taustalla on suhteellisen vahva konsensus	Suhteellisen vakintuneet käsitteet petoksesta ja inhimillisestä virheestä sekä niiden mitaamisesta ovat olemassa	Kyllä: valvontavastuukysymyksissä	Ei juurikaan verrattuna valvontavastuutarkastukseen	Ei sinä määrin kuin vaikutus- ja tilivelvoisen oma toiminta on paremmiin hallittavissa	Tukevat hyvin, koska toiminta esitetään ja tilikausittain	Usein vähemmän kuin vaikuttavuustarkastusta
Lailisuus-tarkastus	Toiminta ja tarkastuskohde on suhteellisen stabiili (esim. lain rikkomisen syntymekanismi on stabiili sekä tiedossa ja fyysisiä toimintaa on myös helpompi mitata kuin vaikutuksia)	Ei, koska lait ja muut normatiiviset dokumentit ovat jo olemassa ja niiden taustalla on suhteellisen vahva konsensus	Vakintuneet käsitteet lainittamisesta sekä sen mitaamisesta ovat olemassa	Kyllä: valvontavastuukysymyksissä	Ei juurikaan verrattuna valvontavastuutarkastukseen	Ei sinä määrin kuin vaikutus- ja tilivelvoisen oma toiminta on paremmiin hallittavissa	Tukevat hyvin, koska toiminta esitetään ja tilikausittain	Usein vähemmän kuin vaikuttavuustarkastusta
Vaikutavuus-tarkastus	Vaikutukset ja tarkastuskohde on usein varsin kompleksinen ja emergentti, jolloin mitaaminen on hyvin vaikeata ja kallimpaa kuin tilintarkastuksessa ja lailisuustarkastuksessa	Usein on, koska poliittisesti päätetyistä vaikuttavuustavoitteista sekä mittareista ei päästä konsensusukseen eikä niitä näin ollen ole lakinkaan kirjattu	Ei useinkaan vähäa konsensusa siitä, mitä vaikutukset ovat tai miten niitä pitäisi mitata	Kyllä: enemmän tai vähemmän suuressa osassa tarkastuskohteita, jos ei peräti kaikissa tarkastuskohteisissa riippuen vaikutavuus-käsitteen määrittelyistä	Kyselyihin nojautuen vaikutuksissa kärsii vastustuksesta ja väkijoukosta ja jakautumasta ja monimuutuksissa vaikutuksissa kärsii vertailtavuus-ongelmista	Vaikutusten hallittavuus on heikompi kuin oman toiminnan ja sanktionneen tai pakittamisen tuloksista on kyseenalaista	Tukevat huomosti silloin, kun vaikutusten syntyminen alajakoitaa on valkua leitä ja vaikutukset syntyvät eritteittajojen operoivan yhteistyön tuloksena	Usein enemmän kuin tilintarkastusta ja lailisuustarkastusta

tutkiville tutkimuksille. Toiseksi, tilivelvolliset mahdollisesti epäoikeudenmukaiseen asemaan asettava vaikuttavuustarkastus on käsittelemisen arvoisen aihealue varmasti myös tilintekovastuuta tarkastelevassa tiedekentässä. Toisaalta tilintekovastuukirjallisuuden kannalta olisi varmasti myös relevanttia käsitellä vaikuttavuustarkastuksen yhteydessä tapahtuvaa todentamisprosessia, jossa tilivelvollinen yhdessä tilityksen vastaanottajan kanssa pohtii toiminnan oikeutusta, todentamista sekä hyväksymistä. Nämä tilintekovastuuta koskevat aiheet omaavat sekä teoreettisesti että empiirisesti relevantteja tutkimuskohteita.

Kolmanneksi, tässä tutkimuksessa esille nostetut mittaustekniset haasteet antavat vaikuttavuusmittaamista pohtiville tutkijoille selkeitä tutkimuskohteita.

Vaikuttavuusmittaamisen tutkimusalueella kaivataan tämän tutkimuksen mukaan kipeästi uusia teoreettisia malleja ja empiirisiä sovelluksia, joiden avulla voitaisiin mitata vaikuttavuutta kustannustehokkaasti ja mielekkäästi. Neljänneksi esille nostettu vaikuttavuustarkastuksen ja poliittisen toiminnan yhtymäpinta kaipaisi uutta tutkimusta. Tällöin tulevat teoreettiset tai empiiriset tutkimukset voisivat muun muassa pohtia poliitikkojen monia rooleja sekä tehtäviä ja niiden merkitystä vaikuttavuustarkastuksen tekemiselle. Toisaalta mielenkiintoista olisi myös pohtia sitä, kuinka vaikuttavuustarkastuksen objektiivisuusvaatimus ja muut hyvään tilintarkastustapaan liittyvät vaatimukset voivat toteutua tässä hyvin poliittisessa tarkastuskontekstissa, jota leimaa ennemminkin informaatiotulva kuin sen puute.

Nämä edellä nostetut tutkimuskohteet ovat vain muutamia esimerkkejä niistä kiinnostavista tutkimusaihioista, joita nousee esille tämän tutkimuksen tuloksista. Artikkelin yhtenä tarkoituksena on avata se tutkimuksellinen tie, joka toivon mukaan johtaa parempaan ymmärrykseen julkisen sektorin vaikuttavuustarkastuksesta. Tälle ymmärrykselle on tilausta, sillä vaikuttavuustarkastaminen on kansainvälisten ja kansallisten tutkimusten mukaan hyvin vähäistä nykyisellään ja tiedämme vielä verrattain vähän siitä, miksi asia on näin. Nykyinen tietämyksen taso johtuu sekä vaikuttavuustarkastuksen että yleisimmin vaikuttavuusilmiön tutkimisen haasteista. Tämä tutkimus on pyrkinyt nostamaan tätä tietämyksen tasoa tuomalla esiin joukon syitä, jotka

voivat selittää vaikuttavuustarkastuksien vähäistä lukumäärää kotimaisessa kontekstissa. Näitä syitä pitäisikin pystyä jatkossa tutkimaan, jotta voimme kehittää vaikuttavuustarkastusta tulevaisuudessa. Aiheen jatkotutkimuksia kuitenkin tulevat vaivaamaan edelleen samat haasteet kuin ennenkin. Näitä haasteita ovat muun muassa aiheen poliittinen herkkyyys ja vaikutusten mittaamisen yleinen vaikeus, josta on usein kiusallisen vaikea puhua. Tutkimista nämä haasteet eivät saisi kuitenkaan estää, sillä vaikuttavuuden tarkastamisella on tärkeä funktio yhteiskuntamme hyvinvoinnin luomisessa nyt ja tulevaisuudessa.

Lähteet

- Agranoff, R. & McGuire, M. 2001. American federalism and the search for models of management. *Public Administration Review* 61(6), 671–681.
- Andrew, C.O. & Hildebrand, P.E. 1976. *Planning and conducting applied research*. New York: MSS Information Corp.
- Anthony, R.N. & Young, D.W. 1988. *Management control in nonprofit organizations*. Homewood, Ill: Irwin.
- Arens, A.A., Elder, R.J. & Beasley, M.E. 2005. *Auditing and Assurance Services: An Integrated Approach*. 10th ed. Upper Saddle River: Pearson Prentice Hall.
- Bachman, F.P. 1912. Attaining Efficiency in City School Systems. *Annals of the American Academy of Political and Social Science* XLI (May): 158–175.
- Baekgaard, M. & Serritzlew, S. 2015. Interpreting performance information: Motivated reasoning or unbiased comprehension. *Public Administration Review* 76(1), 73–82.
- Bethlehem, J., Cobben, F. & Schouten, B. 2011. *Handbook of nonresponse in household surveys* (Vol. 568). John Wiley & Sons.
- Bovaird, T. 2014. Attributing Outcomes to Social Policy Interventions – ‘Gold Standard’ or ‘Fool’s Gold’ in Public Policy and Management? *Social Policy & Administration* 48(1), 1–23.
- Cardy, R.L. & Leonard, B. 2011. *Performance management: Concepts, skills, and exercises*. ME Sharpe.
- Carlin, T. & Guthrie, J. 2003. Accrual output based budgeting systems in Australia The rhetoric-reality gap. *Public Management Review* 5(2), 145–162.
- Chewning, E.G. & Harrell, A.M. 1990. The effect of information load on decision makers’ cue utilization levels and decision quality in a financial distress decision task. *Accounting, Organizations and Society* 15(6), 527–542.

- Chung, J. & Monroe, G.S. 2001. A research note on the effects of gender and task complexity on an audit judgment. *Behavioral Research in Accounting* 13(1), 111–125.
- Conceição, P. & Bandura, R. 2008. *Measuring subjective wellbeing: A summary review of the literature*. United Nations Development Programme (UNDP) Development Studies, Working Paper.
- Cornelius, C. J. 1980. Variation in accounting information load: The effect on loan officers' predictions of bankruptcy. *Accounting Review*, 36–49.
- Diamond, M.J. 2005. *Establishing a performance management framework for government* (No. 5–50). Washington, D.C: International Monetary Fund.
- Dubnick, M. 2005. Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review* 28(3), 376–417.
- Funnell, W. 1998. Executive coercion and state audit: a processual analysis of the responses of the Australian audit office to the dilemmas of efficiency auditing 1978–84. *Accounting, Auditing & Accountability Journal* Vol. 11 No. 4, 436–58.
- Funnell, S.C. & Rogers, P.J. 2011. *Purposeful program theory: Effective use of theories of change and logic models*. San Francisco, CA: Jossey-Bass.
- Grönlund, A., Svärdesten, F. & Öhman, P. 2011. Value for money and the rule of law: the (new) performance audit in Sweden. *International Journal of Public Sector Management* 24(2), 107–121.
- Guba, E.G. & Lincoln, Y.S. 1998. Competing paradigms in qualitative research. In N.K. Denzin & Y.S. Lincoln (Eds) *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks: Sage Publications, 191–215.
- Guthrie, J. & Parker, L. 1999. A quarter of a century of performance auditing in the Australian federal public sector. *Abacus* Vol. 35 No. 3, 302–32.
- Hahn, M., Lawson, R. & Lee, Y.G. 1992. The effects of time pressure and information load on decision quality. *Psychology & Marketing* 9, 365–378.
- Halonen, K. & Steiner, Maj-Lis 2009. *Tilintarkastusprosessi käytännössä*. Helsinki: WSOYpro.
- Hand, D.J. 2004. *Measurement theory and practice*. London: Wiley.
- Harris, J. 1995. Service efforts and accomplishments: a primer of current practice and an agenda for future research. *International Journal of Public Administration* 18(2–3), 253–276.
- Hatry, H.P. 2002. Performance measurement: Fashions and fallacies. *Public Performance & Management Review* 25(4), 352–358.
- Hatry, H.P. 2006. *Performance measurement: Getting results*. Washington: The Urban Institute Press.
- Heinrich, C.J. 2002. Outcomes-based performance management in the public sector: Implications for government accountability and effectiveness. *Public Administration Review* 62(6), 712–725.
- Hodges, R. 2012. Joined-up government and the challenges to accounting and accountability researchers. *Financial Accountability & Management* 28(1), 26–51.

- Hood, C. 2007. What happens when transparency meets blame-avoidance? *Public Management Review* 9(2), 191–210.
- Hood, C. 2010. *The blame game: Spin, bureaucracy, and self-preservation in government*. Princeton University Press.
- Humphreys, P. 2014. *The Chances of Explanation: Causal Explanation in the Social, Medical, and Physical Sciences*. Princeton: Princeton University Press.
- Irwin, T. 1996. *An analysis of New Zealand's new system of public sector management*. Public management occasional papers no 9. Performance management in government. Contemporary Illustrations. Paris: Organisation for Economic Co-operation and Development.
- Jacobs, K. 1998. Value for money auditing in New Zealand: competing for control in the public sector. *British Accounting Review* Vol. 30 No. 4, 343–60.
- Jones, D. R. & M. L. McDermott. 2004. The Responsible Party Government Model in House And Senate Elections. *American Journal of Political Science* 48 (1), 1–12.
- Kates, J., Marconi, K. & Mannle, T.E. 2001. Developing a performance management system for a Federal public health program: the Ryan White CARE ACT Titles I and II. *Evaluation and Program Planning* 24(2), 145–155.
- Keevers, L., Treleaven, L., Sykes, C. & Darcy, M. 2012. Made to measure: Taming practices with results-based accountability. *Organization Studies* 33(1), 97–120.
- Kosonen, L. 2005. *Vaarinpidoista virtuaali aikaan: Sata vuotta suomalaista tilintarkastusta*. Acta Universitatis Lappeenrantaensis.
- Kristensen, J.K., Groszyk, W. & Bühler, B. 2002. Outcome-focused management and budgeting. *OECD Journal on Budgeting* 1(4), 7–34.
- Kunda, Z. 1990. The case for motivating reasoning. *Psychological Bulletin* 108, 480–498.
- Kurunmaki, L. & Miller, P. 2011. Regulatory hybrids: Partnerships, budgeting and modernising government. *Management Accounting Research* 22(4), 220–241.
- Lee, C.V. 2005. Public Health Data Acquisition. Teoksessa Novick, Mayes & Morrow (toim.) *Public Health Administration: Principles for Population-based Management*, 297–328.
- Lonsdale, J. 2008. Balancing independence and responsiveness: a practitioner perspective on the relationships shaping performance audit. *Evaluation* Vol. 14 No. 2, 227–48.
- Lonsdale, J., Wilkins, P. & Ling, T. (Eds). 2011. *Performance auditing: Contributing to accountability in democratic government*. Edward Elgar Publishing.
- Lindblom, C.E. 1959. The Science of 'Muddling Through.' *Public Administration Review* 19(2), 79–88.
- Lindeberg, T. 2007. The ambiguous identity of auditing. *Financial Accountability & Management* 23(3), 337–350.
- Lowe, T. 2013. New development: The paradox of outcomes – the more we measure, the less we understand. *Public Money & Management* 33(3), 213–216.
- Martin, L.L. 1997. Outcome budgeting: A new entrepreneurial approach to budgeting. *Journal of Public Budgeting, Accounting & Financial Management* 9(1), 108.

- Mascarenhas, R.C. 1996. Searching for efficiency in the public sector: Interim evaluation of performance budgeting in New Zealand. *Public Budgeting & Finance* 16, 13–27.
- Mayne, J. 2007. Challenges and Lessons in Implementing Results-Based Management. *Evaluation* 13(1), 87–109.
- McGillivray, M. 2007. Human Well-being: Issues, Concepts and Measures. Teoksessa McGillivray (toim.) *Human Well-Being: Concept and Measurement*. Basingstoke, UK: Palgrave MacMillan.
- Midwinter, A. 2009. New development: Scotland's Concordat – An assessment of the new financial framework in central–local relations. *Public Money & Management* 29(1), 65–70.
- Miller, H.T. & Fox, C.J. 2007. *Postmodern public administration*. New York: M.E. Sharpe.
- Miller, L.E. & Smith, K.L. 1983. Handling Nonresponse Issues. *Journal of Extension* 21 (5), 45–50.
- Mulgan, R.G. 2003. *Holding power to account: Accountability in modern democracies*. Houndmills, Basingstoke etc.: Palgrave Macmillan.
- Needles, B.E. & Powers, M. 2006. *Financial accounting*. Boston, MA: Houghton Mifflin Co.
- Newcomer, K.E. 2015. From outputs to outcomes. Teoksessa Guy & Rubin (toim.) *Public Administration Evolving: From Foundations to the Future*. New York: Routledge, 125–158.
- Nielsen, C. & Madsen, M.T. 2009. Discourses of transparency in the intellectual capital reporting debate: Moving from generic reporting models to management defined information. *Critical Perspectives on Accounting* 20(7), 847–854.
- Pawson, R. & Tilley, N. 1997. *Realistic evaluation*. London: Sage.
- Peirce, C.S. 1998. *The essential peirce: selected philosophical writings*. Bloomington: Indiana University Press.
- Pollitt, C. 2003. Performance audit in Western Europe: trends and choices. *Critical Perspectives on Accounting* Vol. 30 Nos 1/2, 157–70.
- Radcliff, V.S. 1998. Efficiency audit: an assembly of rationalities and programmes. *Accounting, Organization and Society* Vol. 23 No. 4, 377–400.
- Ricchiute, D.N. 2005. *Auditing*. 8th ed. Thomson South-Western, Mason.
- Rozin, P. & Royzman, E.B. 2001. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review* 5(4), 296–320.
- Rudolph, H.R. & Welker, R.B. 1998. The effects of organizational structure on communication within audit teams. *Auditing* 17(2), 1.
- Ryan, C. & Walsh, P. 2004. Collaboration of public sector agencies: reporting and accountability challenges. *International Journal of Public Sector Management* 17(7), 621–631.
- Salminen, A. 2011. *Mikä on kirjallisuuskatsaus? Jobdatus kirjallisuuskatsauksen tyyppeihin ja hallintotieteen sovelluksiin*. Vaasan yliopiston julkaisuja.

- Schedler, A. 1999. Conceptualizing Accountability. Teoksessa Schedler, Diamond & Plattner (toim.) *The Self-Restraining State: Power and Accountability in New Democracies*. London: Lynne Rienner Publishers, 13–28.
- Scutchfield, F.D. & Keck, C.W. 2003. *Principles of public health practice*. Clifton Park: Thomson/Delmar Learning.
- Schalock, R.L. 2001. *Outcome-Based Evaluation*. New York: Kluwer Academic/Plenum Publishers.
- Scott, G.C. 1996. *Government reform in New Zealand*. Washington, DC: Internat. Monetary Fund.
- Simon, H.A. 1979. Rational decision making in business organizations. *The American economic review*, 493–513.
- Skærbæk, P. 2009. Public sector auditor identities in making efficiency auditable: the National Audit Office of Denmark as independent auditor and modernizer. *Accounting, organizations and Society* Vol. 34 No. 8, 971–87.
- Smith, P. 1995. On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18(2–3), 277–310.
- Smith, P.C. 1996. *Measuring outcome in the public sector*. New York: Taylor & Francis.
- Spoon, J.J. & Kliver, H. 2014. Do parties respond? How electoral context influences party responsiveness. *Electoral Studies* 35, 48–60.
- Summer, A. 2006. Economic wellbeing and non-economic wellbeing: A review of the meaning and measurement of poverty. United Nations University World Institute for Development Economics Research, Research Paper. No. 2004/30. April.
- Taber, C.S. & Lodge, M. 2006. Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science* 50(3), 755–769.
- Talbot, C. 2010. *Theories of Performance: Organizational and Service Improvement in the Public Domain*. Oxford University Press.
- Ter Bogt, H.J. 2004. Politicians in Search of Performance Information? – Survey Research on Dutch Aldermen’s Use of Performance Information. *Financial Accountability & Management* 20(3), 221–252.
- The International Standards of Supreme Audit Institutions, ISSAI 100 Reporting standards in Government Auditing, International Organization of Supreme Audit.
- Tomperi, S. 2009. *Tilintarkastus: Normeista käytäntöön*. Helsinki: Edita.
- Van de Walle, S. & Van Dooren, W. 2010. How is information used to improve performance in the public sector? Exploring the dynamics of performance information. Teoksessa Walshe, K., Harvey, G. & Jas, P. (toim.) *Connecting knowledge and performance in public services: From knowing to doing*. Cambridge: Cambridge University Press, 33–55.
- Vakkuri, J. & Meklin, P. 2006. Ambiguity in performance measurement: A theoretical approach to organizational uses of performance measurement. *Financial Accountability & Management* 22(3), 235–250.

- Van der Valk, W. & Van Iwaarden, J. 2011. Monitoring in service triads consisting of buyers, subcontractors and end customers. *Journal of Purchasing and Supply Management* 17(3), 198–206.
- Van Dooren, W., Bouckaert, G. & Halligan, J. 2015. *Performance Management in the Public Sector*. New York: Routledge.
- Van Maanen, J. & Schein, E.H. 1979. *Toward a theory of organizational socialization*. Cambridge: Mass: M.I.T.
- Vedung, E. 1997. *Public policy and program evaluation*. New Jersey: Transaction Publishers.
- Virtanen, P. & Vakkuri, J. 2016. *Julkisen toiminnan tuloksellisuusarviointi*. Tietosanoma Oy, Tallinna Raamatutrukikoda.
- VTV 2010a. *Tarkastusohjeiden yleinen osa*. Valtiontalouden tarkastusviraston tarkastusohjeita. Helsinki: Edita.
- VTV 2010b. *Tuloksellisuustarkastusohje*. Valtiontalouden tarkastusviraston tarkastusohjeita. Helsinki: Edita.
- VTV 2015. *Laillisuustarkastusohje*. Valtiontalouden tarkastusviraston tarkastusohjeita. Helsinki: Edita.
- Weaver, R.K. 1986. The politics of blame avoidance. *Journal of Public Policy* 6(04), 371–398.
- Wholey, J.S. & Hatry, H.P. 1992. The case for performance monitoring. *Public Administration Review*, 604–610.
- Yeung, O.M. & Mathieson, J.A. 1998. *Global benchmarks: Comprehensive measures of development*. Washington D.C: Brookings Institution Press.