

## Puheenvuoro

# RCT-tutkimusten aalto kehitysyhteistyön vaikuttavuuden arvioinnissa

Tiina Pasanen

Vaikuttavuuden arvioinnin (*impact evaluation*) suosio on kasvanut hurjaa vauhtia kansainvälisen kehitysyhteistyön kentällä viimeisen kymmenen vuoden aikana. Satoja kehityskysymyksiin keskittyviä vaikuttavuuden arviointeja tehdään joka vuosi.

Yhtenä suurimmista tekijöistä vaikuttavuuden arvioinnin esiinnousun taustalla nähdään heikentyneen taloustilanteen myötä koventuneet asenteet kehitysyhteistyötä kohtaan. Erityisesti julkista rahaa käyttävillä organisaatioilla ja järjestöillä on tarve todistaa, että kehitysyhteistyöprojekteilla on vaikutusta ja että tukien avulla saavutetaan positiivisia muutoksia avunsaajien elämässä.

Vaikka tilivelvollisuus (*accountability*) rahoittajia ja veronmaksajia kohtaan onkin tärkeä tekijä, se tuskin yksin selittää vaikuttavuuden arvioinnin kasvanutta suosiota. Taustalla on myös sekä rahoittajien että hankkeita toteuttavien järjestöjen halu ymmärtää paremmin hankkeiden toimivuutta ja kehittää niitä paremmiksi. Tilivelvollisuuden voidaan siis nähdä suuntautuvan myös avunsaajia kohtaan. Arviointien tuloksista oppiminen nostetaankin yhä useammin arvioinnin yhdeksi päätavoitteeksi. Vaikuttavuusarviointit ja erityisesti lääketieteestä sosiaalitieteisiin siirtynyttä satunnaistetun kontrolloidun tutkimuksen (RCT) lähestymistapaa<sup>1</sup> hyödyntävät evaluaatiot nähdään keskeisenä keinona saada vastauksia kansainvälisen kehityksen kentällä vallalla olevaan 'mikä todella toimii?' (*what really works*) -ajattelutapaan. Vaikuttavuuden arviointien voidaan siis ajatella tuottavan myös uutta ja luotettavaa todistusaineistoa, jota voidaan hyödyntää yhteiskunnallisessa päätöksenteossa.

Suurin osa vaikuttavuuden arvioinneista perustuu kvantitatiiviseen aineistoon ja ekonometrisiin tekniikoihin. Sekä suosituin että arvostelluin lähestymistapa on edellä mainittu RCT-tutkimus, mutta myös kvasikokeelliset tekniikoita hyödyntävät arvioinnit<sup>2</sup> ovat kasvattaneet suosiotaan. Kokeellisten ja kvasikokeellisten lähestymistapojen suosiota selittää niiden tapa käsitellä avoimesti arvioinnin perusongelmia. Näitä ovat esimerkiksi kysymys siitä, mitä olisi tapahtunut ilman interventiota (*counterfactual*)<sup>3</sup>, valikoituvuuden ongelma (*selection bias*) ja attribuution eli syy-seuraussuhteen todentamisen haaste (*attribution*) (ks. esim. Khandker ym., 2010; Gertler ym., 2011).

Voimakkain RCT-aalto kansainvälisen kehityksen kentällä näyttää olevan ohitse ainakin Euroopassa. Myös kaikkein kiivain vastakkainasettelu RCT-lähestymistavan ja sitä kritisoivien välillä on laantunut. Uskon kuitenkin, että RCT ja niin kutsutut sosiaaliset kokeilut (*social experiments*) tulevat jäämään osaksi vaikuttavuuden arviointia.

Tässä kirjoituksessa esittelen, mitä vaikuttavuuden arvioinnilla tarkoitetaan, mihin RCT-lähestymistavan suosio perustuu ja miten sitä on kritisoitu. Lopuksi pohdin, mitä tästä keskustelusta on opittu ja millaisia uusia suuntauksia vaikuttavuuden arvioinnin kentällä on näköpiirissä.

## MITÄ VAIKUTTAVUUDEN ARVIINNILLA TARKOITETAAN?

Vaikuttavuuden arviointi voidaan määritellä monin eri tavoin. Lähtökohtana määrittelylle on, miten termi vaikutus /vaikuttavuus (*impact*) määritellään.

Laajasti käytetty OECD-DAC:n vaikuttavuuden määritelmä korostaa muutosten pitkäkestoisuutta. Määritelmä nostaa lisäksi esille negatiiviset ja odottamattomat vaikutukset ja on kokonaisuudessaan hyvin laaja:

”[Vaikutukset ovat] positiivisia tai negatiivisia, ensi- tai toissijaisia pitkäkestoisia seurauksia, jotka kehitysyhteistyöhanke on tuottanut joko suorasti tai epäsuorasti, tarkoituksenmukaisesti tai tarkoittamattaan.” (OECD-DAC, 2010:24)<sup>4</sup>

Ero vaikutusten ja tulosten (*outcomes*) välillä on suhteellinen riippuen siitä, miten kehityshankkeen tavoitteet on määritelty. Yleensä tuloksilla viitataan lyhyt- tai keskipitkiin seurauksiin ja vaikutuksilla pitkäkestoisiin tai pysyviin seurauksiin.

Ratkaisevat erot vaikuttavuuden arvioinnin kentällä liittyvät siihen, miten vaikuttavuus mitataan. Tämä puolestaan johtaa kysymykseen mittaamisessa käytettävien lähestymistapojen ja menetelmien hyväksyttävyydestä ja suositeltavuudesta. Esimerkiksi RCT-evaluaatioita ajava vaikutusvaltainen The Abdul Latif Jameel Poverty Action Lab (J-PAL) -instituutti korostaa sitä, miten vaikuttavuuden arvioinnin pitäisi pystyä määrittämään vaikutukset eli esittää tulokset numeroina, jotka kertovat intervention vaikutusten voimakkuudesta.<sup>5</sup> Tämä vaatimus rajoittaa käytettävät menetelmät kokeellisiin ja kvasikokeellisiin lähestymistapoihin, joista edellinen on instituutin mielestä luotettavin ja tarkin. Myös Maailmanpankki ja International Initiative for Impact Evaluation -järjestö (3ie) suosivat kyseisiä lähestymistapoja, vaikkakin erityisesti jälkimmäinen taho puhuu myös monimenetelmällisten (*mixed methods*) vaikuttavuuden arviointien tärkeydestä ja rahoittaa monimenetelmällisiä arviointeja.

Laajempaa määritelmää suosivat tahot eivät halua rajoittaa vaikuttavuuden mittaamista vain yhteen tai muutamaan lähestymistapaan. Esimerkiksi InterActionin oppaassa esitetty määritelmä edustaa tällaista laajempaa näkemystä:

”[Tässä oppaassa] vaikuttavuuden arviointi sisältää minkä tahansa arvioinnin joka systemaattisesti ja empiirisesti tutkii hankkeen vaikutuksia. (Rogers, 2012)<sup>6</sup>

Monet vaikuttavuuden arvioinnin määritelmät (esimerkiksi 3ie-järjestön käyttämä määritelmä<sup>7</sup>) keskittyvät termiin tarkkuus (*rigour*), joka on noussut metodologisten väittelyiden ytimeen. Erityisesti RCT-lähestymistavan kannattajat ovat ottaneet termin haltuunsa korostaen satunnaistamisen avulla saatavaa tilastotieteellistä tarkkuutta vaikuttavuuden mittaamisessa.

### RCT-EVALUAATIOI T EI SATUNNAISTETUT KONTROLLOIDUT TUTKIMUKSET

RCT-evaluaatiot satunnaistavat sen, kuka osallistuu kehityshankkeeseen (koeryhmä) ja kuka ei osallistu (kontrolliryhmä). Kun osallistumista ei satunnaisteta, esimerkiksi mikroluotto-ohjelmassa voivat hankkeeseen osallistuvat olla lähtötasoltaan kyvykkäämpiä tai motivoituneimpia kuin ne, jotka päättävät olla osallistumatta. Tämä puolestaan voi johtaa ylipositiivisiin tuloksiin hankkeen vaikuttavuudesta eikä tuloksia voida yleistää koskemaan laajempaa populaatiota. RCT-lähestymistavan ideana on siis poistaa hankkeisiin liittyvä valikoituvuuden ongelma. Vaikka RCT-kokeiden peruskonsepti onkin yksinkertainen, sen toteuttamiseen liittyy useita erilaisia valintoja sisältäviä vaiheita.

RCT-evaluaatio rakennetaan kehityshankkeeseen alusta alkaen. Kun ongelma, johon halutaan vaikuttaa, on identifioitu ja hankkeesta, jolla siihen aiotaan vaikuttaa, on päätetty, määritellään, miten muutos mitataan. Käytännössä tämä tarkoittaa valintoja siitä, mitkä indikaattorit edustavat haluttua muutosta, mikä on hankkeen kohdejoukko ja satunnaistettava yksikkö ja kuinka satunnaistaminen suoritetaan. Tämän jälkeen kerätään (yleensä kyselytutkimuksella otantamenetelmiä käyttäen) tietoa osallistujien taustamuuttujista ja lähtötasosta (esimerkiksi tuloista, koulutustasosta), jotta varmistetaan satunnaistettavien yksiköiden tilastollinen samankaltaisuus.

Esimerkiksi kuviteltu koulutushanke itäisessä Nepalissa pyrkii lisäämään lasten koulussa käyntiä ja oppimista. Tulosindikaattoreina toimivat lasten poissaolot ja testitulokset. Poissaolot päätetään kerätä viikoittain ja testitulokset vuosittain kahden vuoden ajan.

Yksilöiden sijasta voidaan satunnaistaa erilaisia klustereita, kuten instituutioita (esimerkiksi kouluja) tai maantieteellisiä alueita (esimerkiksi

kyliä tai maakuntia). Satunnaistaminen varmistaa sen, ettei hankkeeseen valita esimerkiksi pelkästään kaikkein aktiivisempia instituutioita tai maantieteellisesti helpommin tavoitettavia alueita. Kontrolliryhmä ei välttämättä jää täysin ilman interventiota, vaan se saattaa saada erilaisen muunnelman samasta hankkeesta. Esimerkiksi koulutushankkeessa testiryhmän lapsia opetetaan uudella tavalla sekä he saavat kouluaterian ilmaiseksi, kun taas kontrolliryhmä saa aterian, mutta opetus jatkuu vanhalla tavalla uuden lähestymistavan sijaan.

Hankkeen päätettyä tai toimittua tietyn ajanjakson verran mitataan valittujen indikaattorien keskiarvot koe- ja kontrolliryhmän välillä ja tutkitaan, ilmeneekö tilastollisesti merkitseviä eroja. Ovatko hankkeen piirissä olleiden lasten poissaolot vähentyneet ja testitulokset keskimäärin parempia kuin kontrolliryhmäläisillä? Havaittua eroa voidaan pitää hankkeen vaikutuksena, sillä satunnaistaminen on kontrolloinut taustamuuttujien vaikutukset ja poistanut oikein sovellettuna selektiivisyyden ongelman.

Yksi laajimmista ja tunnetuimmista satunnaistamista hyödyntäneistä hankkeista on Meksikossa 1990-luvun loppupuolella aloittanut ehdollinen tulonsiirto-/CCT-ohjelma (*Conditional Cash Transfer*) PROGRESA (myöhemmin uudelleen nimetty *Oportunidades*-hankkeeksi). Meksikon hallituksen ja Maailmanpankin yhteisen ohjelman päämääränä oli inhimillisen pääoman kasvattaminen antamalla osallistuville kotitalouksille säännöllisiä tulonsiirtoja lasten koulussa käyntiä ja terveystarkastuksia vastaan. Ohjelmassa paikannettiin äärimmäisen köyhät alueet (*localities*). Yhdelle kolmasosalle identifioituista alueista ohjelman alkamista lykättiin 18 kuukaudella, jolloin se toimi kontrolliryhmänä testialueille. Sekä koulussa käyntiä, terveystarkastuksia että rahan-siirtoja monitoroitiin tarkasti ja säännöllisesti. Tietyn ajanjakson jälkeen tutkijat analysoivat kerätyn aineiston<sup>8</sup>. Ohjelman tulokset olivat sen verran lupaavia, että ohjelman kattavuutta laajennettiin, ja vuonna 2012 se kattoi 6,5 miljoonaa meksikolaiskotitaloutta.

PROGRESA:n ja Brasilialaisen Bolsa Familiarin<sup>9</sup> suosion siivittämänä CCT-ohjelmat ovat levinneet lähes koko Latinalaiseen Amerikkaan, jossa ne ovat olleet kymmenien vaikuttavuuden arviointien kohteena. Viime vuosien aikana

CCT-ohjelmia on sovellettu myös Aasiassa ja Afrikassa, joskin suurin osa Afrikan tulonsiirto-ohjelmista toimii ilman ehtoja (*unconditional*).

### RCT-lähestymistapaan kohdistuvaa kritiikkiä

RCT-tutkimusten suosio on myös herättänyt suurta vastustusta ja kritiikkiä tutkijoiden, organisaatioiden ja arviointija tekevien tahojen parissa. Argumentit liittyvät niin eettisiin periaatteita, käyttäytymisen muutosta kuin toteuttamistakin koskeviin kysymyksiin.

*Eettiset argumentit* tiivistyvät kysymykseen siitä, kuka saa apua ja kuka ei. Kun hankkeeseen osallistumisesta päätetään satunnaisesti valitun populaation sisällä, tarkoittaa se väistämättä, että osa apua tarvitsevasta (eli kontrolliryhmä) jää hankkeen ulkopuolelle. Kuitenkaan harvoin mikään kehityshanke pystyy tavoittamaan kaikkia avunsaajia, mikä onkin RCT-lähestymistavan puolustajien usein käyttämä vasta-argumentti. Tämän lisäksi alueittain tai klustereittain etenevät RCT-tutkimukset ovat yleinen ratkaisu. Tällöin kontrolliryhmä tai -ryhmät pääsevät hankkeen piiriin tietyn ajanjakson jälkeen (esimerkiksi vuoden kuluttua) toimittuaan ensin verrokkina varsinaiselle testiryhmälle.

Usein kuultu argumentti liittyy satunnaistamisen *kalleuteen*. Hyvin suunniteltu ja toteutettu RCT-evaluaatio voikin maksaa useita satojatuhansia euroja. Tavallisessa kyselytutkimusmenetelmiä hyödyntävässä evaluaatiossa kerätään tietoa hankkeesta ja sen osallistujista vain kerran eli hankkeen loputtua. RCT-evaluaatiossa sen sijaan kerätään tietoa niin testi- kuin kontrolliryhmästä sekä hankkeen alussa että lopussa. Tämä voi moninkertaistaa kustannukset. Toisaalta mikä tahansa muu hyvin tehty, laaja ja useampia metodeja hyödyntävä evaluaatio tulee usein hyvin kalliiksi.

*Käyttäytymiseen liittyvät argumentit* kiinnittävät huomiota siihen, että RCT-tutkimuksiin osallistuminen voi muuttaa niin koe- ja kontrolliryhmän jäsenten kuin hanketta toteuttavien työntekijöidenkin käytöstä. Vaikka lääketieteellisten RCT-tutkimusten ideaaliin kuuluu, että osallistujat eivät tiedä kuuluvatko he koeryhmään (joka saa esimerkiksi uuden lääkkeen) vai kontrolliryhmään (joka saa lume-lääkkeen), tämä *blinding*-oletus ei usein toteudu sosia-

litieteellisissä kokeiluissa. Osallistuvat henkilöt tietävät yleensä olevansa osa tutkimusta, ja odotukset hankkeen tuloksista ovat useimmiten hankkeen työntekijöiden ja osallistujien tiedossa. Käyttäytymisen tietoisesta muutoksesta arvioinnin alaisena ollessa on runsaasti tutkimustietoa. Osallistujien käyttäytymisen muutoksesta käytetäänkin yleisesti nimitystä Hawthorne-vaikutus (ks. Levitt ja List 2009.) ja työntekijöiden käyttäytymisen muutoksesta termiä John Henry -ilmiö (ks. Saretsky 1975.).

Kuten osallistujien tietoinen käytöksen muuttuminen tutkimuksen aikana, myös hankkeesta pois jättäytyminen (*drop out-behaviour tai attrition*) voi vaikuttaa RCT-tutkimuksen sisäiseen validiteettiin. Jos esimerkiksi koeryhmän kaikkein köyhimmät osallistujat jättäytyvät systemaattisesti pois hankkeesta, tämä voi vääristää koeryhmän tuloksia. Ratkaisuksi tähän on ehdotettu pois jättäytyneiden seuraamista (Duflo ym., 2008), mutta kustannusten vuoksi tätä harvoin tehdään.

*Ulkoinen validiteetti.* Kaikkein olennaisin argumentti RCT -kokeita vastaan lienee kuitenkin kysymys sisäisen ja ulkoisen validiteetin suhteesta. Sisäisellä validiteetilla tarkoitetaan RCT-koeasetelman kykyä tai pätevyyttä arvioida hankkeen ja tulosten kausaalisuhdetta. Satunnaistaminen tuottaa tilastollisesti samankaltaiset testi- ja kontrolliryhmät ja kontrolloi muut mahdollisesti vaikuttavat tekijät, mikä puolestaan nostaa tutkimuksen sisäistä validiteettia. Ulkoisella validiteetilla tarkoitetaan tulosten yleistettävyyttä laajemmalle. Samalla kun RCT-hankkeiden pääasiallinen huomio on keskittynyt sisäiseen validiteettiin ja selektiivisyyden eliminoimiseen, monet muut olennaiset näkökulmat kuten arvioinnin relevanttius, hyödyllisyys ja erityisesti ulkoinen validiteetti ovat jääneet vähemmälle huomiolle.

Jos tuloksia, jotka seuraavat tiettyä hanketta tietystä kontekstissa, ei pystytä selittämään, on ongelmallista yleistää tuloksia muihin konteksteihin ja olettaa, että samanlainen hanke tuottaisi samanlaisia tuloksia toisaalla. Tulosten taustalla voi esimerkiksi olla tiettyjä mekanismeja, prosesseja tai paikallisia tekijöitä jotka joko aiheuttavat tai mahdollistavat muutoksen tai vaikuttavat muutoksen voimakkuuteen.

Tunnettu amerikkainen tutkija Pritchett on tarttunut RCT-argumentoinnissaan juuri tä-

hän sisäisen ja ulkoisen validiteetin ristiriitaan. Pritchett ja Sandefur (2013) toteavat, että ulkoinen validiteetti on erityisen tärkeä, jos RCT-tutkimuksen tarkoituksena on tehdä laajemmalle ulottuvia yhteiskunnallisia (*policy*) suosituksia.

Kysymys ulkoisesta ja sisäisestä luotettavuudesta onkin johtanut siihen, että ”mikä todella toimii” -tyyppiset kysymykset on kasvavassa määrin korvattu miksi, miten, missä kontekstissa ja kenelle -tyyppisillä kysymyksillä. ”Musta laatikko” (black box) -tyyppiset arvioinnit jotka vain toteavat jonkin joko toimivan tai ei, mutta jotka eivät kykene selittämään prosesseja, mekanismeja tai kontekstia, jossa muutos tapahtuu, rajoittavat oppimista ja intervention kehittämistä sekä tulosten yleistämisen mahdollisuuksia.

### Uudet lähestymistavat ja metodit vaikuttavuuden arvioinnissa

Vuonna 2012 julkaistiin kaksi arviointia koskevaa vaikutusvaltaista raporttia: Stern ym. (2012) ja White ja Phillips (2012). Molemmat raportit hahmottelivat vaihtoehtoisia lähestymistapoja ja metodeja vaikuttavuuden mittaamiseen. Kyse ei ollut niinkään innovatiivisista tai täysin uusista metodeista vaan ennemminkin mahdollisuuksista soveltaa muilla aloilla käytettyjä lähestymistapoja ja metodeja kansainvälisen kehityksen kentällä erityisesti silloin, kun formaalit (eli kokeelliset ja kvasikokeelliset) lähestymistavat eivät ole mahdollisia. Tällaisia tilanteita ovat esimerkiksi monimutkaisiset (*complex*) ja pienen joukon (*small-n*) tapaukset.

Stern ym. (2012) jakavat potentiaaliset lähestymistavat teoriapohjaisiin (esim. realistinen evaluaatio), tapaustutkimus -pohjaisiin (esim. vertaileva kvalitatiivinen analyysi tai kontribuutioanalyysi) ja osallistaviin (esim. osallistava evaluaatio) lähestymistapoihin. Jotkut lähestymistavat eivät välttämättä ole yksistään riittäviä arvioimaan hankkeen ja vaikutusten kausaalisuhdetta, mutta järkevästi yhdisteltyinä ne voivat tuottaa tietoa kausaalisuuden mekanismeista ja olla relevantimpia sekä hyödyllisempiä lähestymistapoja.

White ja Phillips (2012) puolestaan esittelevät kahdeksan potentiaalista lähestymistapaa tutkia pienen joukon (*small-n*) kehitysohjelmien vaikuttavuutta. Osa ehdotetuista menetelmistä on samoja kuin Sternin ym. (2012) raportissa

(kuten kontribuutioanalyysi), ja myös Suomessa tunnetut Outcome mapping- ja Most significant change -menetelmät ovat mukana.

Sekä Sternin ym. ja White ja Phillipsin raportit kuitenkin lähinnä esittelevät uusia mahdollisuuksia. Ne sisältävät hyvin rajoitetun määrän käytännön esimerkkejä siitä, miten kyseisiä menetelmiä tai lähestymistapoja on sovellettu käytännössä vaikuttavuuden mittaamisessa. Viime vuosina onkin aloitettu monia uusia vaikuttavuuden arvioinnin projekteja, joiden päämääränä on tuottaa lisää tietoa uusien lähestymistapojen soveltuvuudesta.

### Monimenetelmällisyys

Uusien ja soveltavien lähestymistapojen lisäksi monimenetelmällisyyden (*mixed methods*) edut on nostettu vaikuttavuuden arvioinnin keskiöön. Koska kaikissa lähestymistavoissa ja menetelmissä on omat sisäsyntyiset vahvuutensa ja heikkoutensa, ajatellaan yleisesti, että yhdistämällä eri menetelmiä voidaan rajoittaa heikkouksien vaikutusta. Millaisia menetelmiä yhdistää ja millä tavalla, ja mitä etuja yhdistämisestä saadaan vaikuttavuuden arvioinnissa, ovat kysymyksiä, jotka vaativat kehittämistä ja lisää käytännön esimerkkejä.

Tällä hetkellä monimenetelmällisyys vaikuttavuuden arvioinnissa tarkoittaa käytännössä usein sitä, että kokeelliseen tai kvasikokeelliseen arviointimalliin yhdistetään yksi tai useampi laadullinen tutkimuselementti kuten avainhenkilöhaastatteluja ja kohderyhmäkeskusteluja.

Vaikka laadullisten menetelmien rooli on useimmiten hyvin rajoitettua monimenetelmällisyyttä käyttävissä vaikuttavuuden arvioinneissa, hyviä esimerkkejä niiden mahdollisuuksista selittää vaikutusten taustoja ja mekanismeja on useita.

Esimerkiksi Adaton (2007) tutkimus ehdollisten rahansiirto-ohjelmien vaikuttavuudesta Nicaraguassa ja Turkissa yhdistää (kvasi)kokeellisen lähestymistavan laadulliseen ja etnografiseen tutkimukseen. Etnografisen kenttä-tutkimuksen osuus ei rajoittunut vahvistamaan (kvasi)kokeellisen puolen tuloksia, vaan sen avulla selitettiin, miksi tietyt muutokset eivät olleet tapahtuneet. Esimerkiksi Nicaraguassa anemia ei laskenut vaikka osallistajat raportoivat antavansa rautatabletteja lapsilleen, ja Turkissa

tyttöjen koulunkäynti ei lisääntynyt odotetulla tavalla. Nicaraguan tapauksessa osallistajat eivät raportoinnista huolimatta jakaneet rautatabletteja koska ne aiheuttivat lapsille vatsakipuja ja Turkissa kulttuuriset asenteet tyttöjen koulunkäyntiä kohtaan erityisesti koulujen sijaitessa kaukana kotoa selittivät sen miksi tyttöjen koulunkäynti ei lisääntynyt.

Tämän lisäksi etnografinen kenttätutkimus paljasti osan mitatuista muutoksista virheelliseksi. Esimerkiksi Nicaraguassa lasten paino ei kasvanutkaan mittauksista huolimatta, sillä vanhemmat pakkojuottivat lapsiaan ennen punnituksia pysyäkseen ohjelmassa.

Laadullisten menetelmien nähdään olevan hyviä kuvaamaan prosesseja, arvoja ja käyttäytymistä. Kiinnostus monimenetelmällisyyttä kohtaan onkin esimerkki edellä mainitusta arvioinnin kehityssuunnasta, jonka mukaan enää ei haluata vain selvittää 'mikä toimii' vaan kuinka, miksi ja missä kontekstissa interventiot toimivat.

### LOPUKSI

Vaikka RCT-tutkimuksiin viitataan joskus 'kultaisena standardina', harva satunnais-kontrollikokeita kannattava organisaatio tai tutkija pitää kyseistä lähestymistapaa ainoana oikeana keinona selittää ja todentaa mitä vaikutuksia hankkeella on, tai kuvittelee, että se sopisi millaisen tahansa hankkeen arviointiin. Yleisesti ajatellaan, että RCT-estimaatit sopivat parhaiten suhteellisen virtaviivaisiin ja yksinkertaisiin hankkeisiin, kuten juuri CCT-ohjelmiin, mutta eivät ole toimiva vaihtoehto kompleksisiin, moniulotteisiin tai pienen joukon hankkeiden arviointiin.

RCT-lähestymistavan kehittämisestä ja siihen kohdistuvasta kritiikistä onkin otettu oppia viime vuosien aikana. Sen kannattajien ja vastustajien ymmärrys on lisääntynyt (ainakin jossain määrin) ja samaan aikaan on noussut tilaa uusille, vaihtoehtoisille tavoille hahmottaa kausaalisia yhteyksiä hankkeen toimintojen ja havaittujen muutosten välillä.

Vaikka nämä uudet vaikuttavuuden arvioinnin lähestymistavat ja menetelmät tulevat mitä todennäköisimmän kehittymään ja täydentämään nykyistä vaikuttavuuden arvioinnin portfolioa, kokeelliset ja kvasikokeelliset menetel-

mät eivät kuitenkaan tule poistumaan. Myös niiden käyttöä ja soveltuvuutta tullaan kehittämään. Sen vuoksi ei ole relevanttia tai hyödyllistä ainoastaan kritisoida niitä tai vaihtoehtoisesti kohdella niitä ainoana mahdollisena vaihtoehtona vaikuttavuuden arvioinnissa.

Olennaista on myös miettiä, mitkä ovat vaikuttavuuden arviointien tavoitteet, kuinka arviointeja hyödynnetään kehityspoliittisessa päätöksenteossa ja kuinka niitä voitaisiin hyödyntää entistä paremmin.

## VIITTEET

<sup>1</sup> Randomised Control Trial (RCT) -termistä voidaan käyttää suomennosta satunnaistetut kontrolloidut kokeet/tutkimukset. Tässä kirjoituksessa käytetään lyhennettä RCT koska sen käyttö on erittäin vakiintunut kansainvälisessä kehityskeskustelussa.

<sup>2</sup> Kvasikokeelliseen lähestymistapaan luetaan kuuluvaksi propensity score matching, difference-in-difference, discontinuity regression ja instrumental variable –nimiset tekniikat.

<sup>3</sup> Counterfactual termi viittaa ns. vastakkaiseen todellisuuteen eli mitä olisi tapahtunut samoille ihmisille samassa ajassa ilman interventiota/kehityshanketta.

<sup>4</sup> “Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.” (OECD-DAC, [2002] 2010: 24)

<sup>5</sup> <http://www.povertyactionlab.org/methodology/what-evaluation/impact-evaluation>

<sup>6</sup> “In this guidance note, an impact evaluation includes any evaluation that systematically and empirically investigates the impacts produced by an intervention.” Rogers PJ. (2012) Introduction to Impact Evaluation. Melbourne, VIC: InterAction, The Rockefeller Foundation, RMIT & BetterEvaluation. <http://www.interaction.org/document/introduction-impact-evaluation>

<sup>7</sup> “Rigorous impact evaluation studies are analyses that measure the net change in outcomes for a particular group of people that can be attributed to a specific program using the best methodology available, feasible and appropriate to the evaluation question that is being investigated and to the specific context.” <http://www.3ieimpact.org/media/filer/2012/05/17/3iefoundingdocument30june2008.pdf>

<sup>8</sup> PROGRESASTA on tehty kymmenittäin tutkimuksia eri näkökulmista, esimerkiksi hankkeen vaikutuksista terveyteen (Gertler 2004), aikuisten ja nuorten työmarkkinoille (Behrman, Parker, ja Todd 2009; Skoufias ja di Maro 2007); koulunkäyntiin (de Janvry ja muut 2006; Schultz 2004; Todd ja Wolpin 2006); ja ruokavaliioon (Behrman ja Hoddinott 2005; Hoddinott ja Skoufias 2004).

<sup>9</sup> Samoihin aikoihin aloitti myös Brasiliassa vastaavanlainen hanke nimeltä Bolsa Familia.

## LÄHTEET

The Abdul Latif Jameel Poverty Action Lab. Haettu sivulta <http://www.povertyactionlab.org/methodology/what-evaluation/impact-evaluation>, 25.1.2014.

Adato, Michelle (2007). Combining survey and ethnographic methods to evaluate conditional cash transfer programs. *Q-Squared Working Paper no. 40*. Q-squared, Centre for International Studies, University Of Toronto.

Behrman, Jere & Hoddinott, John (2005). Programme evaluation with unobserved heterogeneity and selective implementation: The Mexican ‘PROGRESA’ impact on child nutrition. *Oxford Bulletin of Economics and Statistics* 67(4), 547–69.

Behrman, Jere, Parker, Susan & Todd, Petra (2009). Long-term impacts of the oportunidades conditional cash-transfer program on rural youth in Mexico. Teoksessa Klasen, Stephen & Nowak-Lehmann, Felicitas (Eds.), *Poverty, inequality, and policy in Latin America*, (s. 219–70). Cambridge, MA: MIT Press.

Duflo, Esther, Glennerster, Rachel & Kremer, Michael (2008). Using randomization in development economics research: A toolkit. Teoksessa Schultz, T.P. & Strauss, J. (Eds.), *Handbook of development economics*, vol. 4. Amsterdam: Elsevier.

Gertler, Paul (2004). Do conditional cash transfers improve child health? Evidence from

- PROGRESA's control randomized experiment." *American Economic Review, Papers and Proceedings*, 94(2), 336–41.
- Gertler, Paul J., Martinez, Sebastian, Premand, Patrick, Rawlings, Laura B. & Vermeersch, Christel M. J. (2011). *Impact evaluation in practice*. The World Bank.
- Hoddinott, John & Skoufi, Emmanuel (2004). The impact of PROGRESA on food consumption. *Economic Development and Cultural Change*, 53(1), 37–61.
- The International Initiative for Impact Evaluation (2008). *Founding document*. Haettu sivulta <http://www.3ieimpact.org/media/filer/2012/05/17/3iefoundingdocument30june2008.pdf>, 12.1.2014.
- de Janvry, Alain, Finan, Frederico, Sadoulet, Elisabeth & Vakis, Renos (2006). Can conditional cash transfer programs serve as safety nets in keeping children at school and from working when exposed to shocks? *Journal of Development Economics*, 79(2), 349–73.
- Khandker, Shahidur R., Koolwal, Gayatri B. & Samad, Hussain A. (2010). *Handbook on impact evaluation: Quantitative methods and practices*. World Bank.
- Levitt, Steven D. & List, John A. (2009). Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. NBER Working Paper no. 15016.
- OECD-DAC. ([2002] 2010) Glossary of key terms in evaluation and results based management. Development assistance committee Working Party on aid effectiveness. Organisation for Economic Co-operation and Development. Haettu sivulta <http://www.oecd.org/development/peer-reviews/2754804.pdf>, 29.12.2013.
- Pritchett, Lant & Sandefur, Justin (2013). Context matters for size: why external validity claims and development practice don't mix. CGD working paper 336. Haettu sivulta <http://international.cgdev.org/publication/context-matters-size-why-external-validity-claims-and-development-practice-dont-mix>, 30.1.2014.
- Rogers, Patricia (2012) Introduction to impact evaluation. Melbourne, VIC: InterAction, The Rockefeller Foundation, RMIT & BetterEvaluation. Haettu sivulta <http://www.interaction.org/document/introduction-impact-evaluation>, 10.1.2014.
- Saretsky, Gary (1975). The John Henry effect: Potential confounder of experimental vs control group approaches to the evaluation of educational innovations. *The American Educational Research Association's Annual Meeting*. Washington, D.C..
- Schultz, T. Paul (2004). School subsidies for the poor: Evaluating the Mexican PROGRESA Poverty Program. *Journal of Development Economics*, 74(1), 199–250.
- Skoufias, Emmanuel, & di Maro, Vincenzo (2007). Conditional cash transfers, adult work incentives, and poverty. Policy Research Working Paper 3973. World Bank, Washington, DC.
- Stern, Elliot, Stame, Nicoletta, Mayne, John, Forss, Kim, Davies, Rick & Befani, Barbara (2012). Broadening the range of designs and methods for impact evaluations. Report of a study commissioned by the Department for International Development. Working paper 38, DFID. Haettu sivulta <http://www.dfid.gov.uk/Documents/publications1/design-method-impact-eval.pdf>, 29.12.2013.
- Todd, Petra & Wolpin, Kenneth (2006). Assessing the impact of a school subsidy program in Mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility." *American Economic Review*, 96(5), 1384–417.
- White, Howard & Phillips, Daniel (2012). *Addressing attribution of cause and effect in small n impact evaluations: towards an integrated framework*. Towards an integrated framework. The International Initiative for Impact Evaluation (3ie) Working paper 15. Haettu sivulta [http://www.3ieimpact.org/media/filer/2012/06/29/working\\_paper\\_15.pdf](http://www.3ieimpact.org/media/filer/2012/06/29/working_paper_15.pdf), 5.1.2014.