

Arvioita suomalaisen yliopistotutkimuksen tuloksellisuudesta

■ Mika Nieminen ja Otto Auranen

Jälkiteollisten maiden tiedepolitiikassa uutta tietoa luovat organisaatiot nähdään taloudellisen menestyksen ja sitä kautta yhteiskunnallisen hyvinvoinnin kannalta tärkeinä toimijoina. Viime vuosien arviointibuumi (Neave 1998) ja yliopistojen tuloksellisuuden korostaminen ja mittaaminen ovat ilmeisiä seurauksia siitä strategisesta asemasta, johon tietoa tuottavat organisaatiot on politiikassa nostettu. Mitä tärkeämmiksi ja laajemmiksi koulutuksen ja tutkimuksen investoinnit ovat tulleet, sitä yksityiskohtaisemmin on haluttu varmistua rahoituksen tehokkaasta, vaikuttavasta ja asianmukaisesta käytöstä. Samalla lisääntynyt arviointi on ollut viesti vähentyneestä luottamuksesta tietoa tuottavia organisaatioita kohtaan ja poliittisten ohjaustarpeiden lisääntymisestä.

1990-luvun vaihteesta alkaen arvioinnit kotiutuivat vaihtelevin toteutustavoin ja tavoittein useisiin kansallisiin järjestelmiin. Eräs tunnetuimpia kansainvälisiä esimerkkejä tutkimuksen arvioinnista lienee Isossa-Britanniassa käyttöön otettu Research Assessment Exercise (RAE) -järjestelmä (ks. <http://www.rae.ac.uk/>). Arviointia hyödynnetään tavalla tai toisella tutkimus- ja yliopistojärjestelmien ohjaamisessa kymmenissä eri maissa kautta maailman (Jongbloed & Vossensteyn 2001; Geuna & Martin 2003).

Globalisaatiokehityksen myötä yliopistot kilpailevat kuitenkin lisääntyvästi kansainvälisillä koulutuksen ja tutkimuksen markkinoilla. Näiden markkinoiden benchmarking- ja markkinointitarpeita on kehitetty palvelemaan lukuisia erilaisia kansainvälisiä ”ranking-listoja”. Vertailuja voivat hyödyntää niin yliopistot oman toimintansa kehittämisessä kuin rahoittajat hoidossa koulutuksen ja tutkimuksen investointien vaikuttavuutta. Kansainvälises-

sä ympäristössä myös opiskelijat ja työnantajat tarvitsevat tietoa valintojensa ja rekrytointiensä pohjaksi. Tätä tarvetta voidaan tyydyttää erilaisen vertailujen avulla (Holmes 2006). Kansainväliset yliopistovertailut ovat olleet erityisesti 2000-luvulla esiin noussut ilmiö ja niistä eniten keskustelua lienevät aiheuttaneet Sanghain Jiao Tong -yliopistossa laadittu maailman yliopistojen vertailu (<http://www.arwu.org/>) sekä *Times Higher Education* -lehden yliopistojen kansainvälinen vertailu (<http://www.timeshighereducation.co.uk/>). Näiden lisäksi on löydettävissä useita erilaisiin aineistoihin ja toteutustapoihin perustuvia kansainvälisiä ja kansallisia yliopistojen vertailuja (esim. http://en.wikipedia.org/wiki/College_and_university_rankings).

Erityisesti Suomen kaltaisen pienen ja suhteellisen perifeerisen tiedemaan yliopistojärjestelmää koskevalle politiikkakeskustelulle on ollut tyypillistä ”kansainvälisen tason” ja ”kansainvälisten huippuyliopistojen” seuraaminen ja nostaminen politiikan tavoitteeksi. Kansallista tasoa ajatellaan korotettavan puolestaan mm. tuloksellisuuden arvioinnilla ja palkitsemisella. Tämä ajattelumalli on näkynyt kautta linjan viime vuosien yliopistopolitiikassa. Parhailaan opetusministeriössä kehitellään myös aiempaa kattavampaa indikaattorikonaisuutta korkeakoulujen toiminnan seuraamiseksi ja ohjaamiseksi.

Yliopistojen tuloksellisuuden arvioinnille ja mittaamiselle sekä näistä käytävälle keskustelulle on siis monitahoista ajankohtaista kysyntää. Perustelluista syistä aihe on todettu kuitenkin metodisesti vaikeaksi. Muun muassa aineiston luotettavuuteen, laskentaan ja tulkintaan saattaa liittyä useita ongelmia (esim. Billaut 2009; Florian 2007; Holmes 2006; Van Raan 2005; Weingart 2005; Sipilä & Vuorinen 2005; Adam 2002).

Lisäksi kenttä on yliopistopoliittisesti miinoitettu. Erilaisten epäonnistuneiksi ja puutteelliseksi todettujen rankinglistojen ja tuloksellisuusmittausten maailmassa on silti edelleen perusteltua pyrkiä etsimään toimivampia arviointimenetelmiä.

Suomessakin on aika ajoin esitetty erilaisia laskentamalleja yliopistojen tuloksellisuuden arvioimiseksi. Viimeisimmät lienevät Osmo Kivisen, Juha Hedmanin ja Kaisa Peltoniemen (2008, 2009) käsialaa. He julkaisivat äskettäin sarjassaan jo toisen julkisuudessaakin huomioitun Turun yliopiston koulutussosiologian tutkimuskeskuksen RUSE:n raportin yliopistojen tieteellisen toiminnan tuloksellisuudesta. Raporteille on asetettu varsin kunnianhimoinen tavoite: ”Aloittain tehtävillä analyyseillä eri yliopistoissa toimivat tieteenalayksiköt saatetaan tieteellisen toiminnan tuloksellisuuden mukaiseen järjestykseen omien alojensa ehdoilla.” Raportit jättävät kuitenkin paljon kysymyksiä eivätkä tulokset ole yksiselitteisiä.

Paneudumme seuraavassa lähemmin näihin raportteihin. Ne tarjoavat esimerkkejä saatavissa olevaan aineistoon, valittavissa oleviin indikaattoreihin sekä laskentaan liittyvistä haasteista, joita suomalaisten yliopistojen tutkimustoiminnan tuloksellisuuden arvioinnissa joudutaan väistämättä miettimään. Arvioinnit eivät ole nykymaailmassa vältettävissä ja tästä syystä on toivottavaa, että keskustelu arviointimalleista jatkuu parhaan mahdollisen lopputuloksen saavuttamiseksi.

Arvioinnin kohteena koulutusala

Kivisen, Hedmanin ja Peltoniemen arviointiraporttien lähestymistapa yliopistojen tutkimustuloksellisuuteen perustuu tieteenaloittaiseen panos–tuotos-tarkasteluun. Yhtäältä tieteenaloja ei siis vertailla keskenään ja toisaalta yliopistojen panokset (resurssit) ja tuotokset suhteutetaan toisiinsa. Tekijät itse toteavat vertailuasetelmasta: ”Vaikka eri tieteenalojen tulosten vertailu on usein koettu vaikeaksi, ylivoimaista se ei kuitenkaan ole. Omissa tutkimuksissamme emme tavoittele vertailtavuutta tieteenalojen oman laadun kustannuksella, vaan opeoimme ala-

kohtaisilla tulos- ja panososuuksilla, jotka suhteutamme toisiinsa ja joiden perusteella saamme tieteenalayksiköt tieteellisen toiminnan tuloksellisuuden mukaiseen järjestykseen alojen sisällä.” (Kivinen ym. 2008, 3.)

Nykyisin on ehkä jopa itsestään selvää, että tieteenaloja ei tulisi vertailla keskenään. Tieteenalojen toimintamallit, rakenteet, rahoitusmahdollisuudet, laatuarviot ja myös julkaisukulttuurit eroavat toisistaan (esim. Becher 1989; Whitley 1984; Kekäle & Lehikoinen 2001; Kyvik 1991; Puuska & Miettinen 2008). Näin ollen toisistaan poikkeavia tieteenaloja ei tulisi arvioida ainakaan huomioimatta niiden erityispiirteitä.

Valinta on perusteltu, mutta tästä huolimatta ensimmäinen kysymys nousee juuri tästä tutkimuskohteen määrittelystä ja sen rajoista. Tutkijat käyttävät nimittäin aineistoinaan opetusministeriön ylläpitämää KOTA-tietokantaa (ks. <https://kotaplus.csc.fi/online/Etusivu.do>), jossa yliopistojen toimintaa koskevaa määrällistä aineistoa on mahdollista ryhmitellä vain koulutusaloittain. Tätä KOTA-määrittelyiden mukaista koulutusala he kutsuvat ”tieteenalaksi” ja jossakin yliopistossa edustettuna olevaa koulutusala ”tieteenalayksiköksi”. KOTA-tietokannan mukaista koulutusalan määrittelyä ei ole kuitenkaan esitetty. Lukija ei tiedä, miten se rajaa tai ei rajaa tarkastelua (mitä oppiaineita on mukana). Käytetyt ilmaisut ”tieteenala” ja ”yksikkö” saattavat olla tästä syystä jopa hämääviä, koska niillä viitataan tavanomaisesti aivan muunlaisiin kokonaisuuksiin kuin koulutusaloihin yliopistojen sisällä.

KOTA-järjestelmästä johtuen tulokset yleistyvät helposti liikaa. Vertailussa on laajoja ja tosiasiaa sisäisesti hyvin heterogeenisiä koulutusaloja, jotka pitävät sisällään useita erilaisia oppiaineita. Yliopistokohtaisista oppiaineprofileista johtuen KOTA:n koulutusala ei ole kaikin puolin vertailukelpoinen luokitus tutkimustoiminnan arviointiin. Parempi vaihtoehto KOTA-luokittelulle olisi Tilastokeskuksen tutkimus- ja kehittämistoiminnan tilastoissaan käyttämä OECD:n määritelmille perustuva 46-kohtainen tieteenalaluokittelu (ks. <http://www.stat.fi/til/tkke/index.html>). Esimerkiksi nyt arvioissa

aino

UNELMAN
KEVEÄÄ



FREDA 33
HELSINKI
MA-PE 10.30-18
LA 10.30-15
PUH. 611 611
WWW.AINO.NET

yhtenä blokkina esiintyvä tekniikan ala jakautuisi tällöin sisäisesti 12 tieteenalaan, jotka ovat erilaisin painotuksin ja volyymein edustettuina teknisissä yliopistoissa ja tiedekunnissa. Yliopistokohtaisten tilastojen tarkastelu laitoksittain osoittaa myös nopeasti, että yliopiston koulutusalan sijoitus saattaa joissakin tapauksissa olla vain muutaman laitoksen toiminnan varassa. Tällöin koko koulutusala koskevia yleistyksiä voi tuskin pitää kovinkaan tarkkoina tai oikeudenmukaisina.

Ongelmana tällä hetkellä kuitenkin on, että mainitut Tilastokeskuksen aineistot eivät sisällä KOTA:n tulosaineistoja. Lisäksi KOTA:n koulutusala luokittelu ei vastaa sellaisenaan tieteenala luokittelua, koska KOTA on luotu koulutuksen seurantaan varten. Aineistot olisi kuitenkin mahdollista saattaa vertailukelpoisiksi (vrt. Miettinen & Selovuori 2007). Tällöin voitaisiin käyttää yliopistotasoisia aineistoja. Yliopistot toimittavat tutkimuksen panostiedot Tilastokeskukselle, joten KOTA:an toimitettavan tulosaineiston luokittelu vastaavalla tavalla ei ole ongelma. Tiedot kerätään yliopistoissa laitostasoisina ja luokitellaan vasta tämän jälkeen KOTA:an koulutusaloittain tai Tilastokeskuksen tieteenaloittain. Tieteenalaluokittelu tuottaisi tarkempia kuvauksia tieteenalojen ja yliopistojen panos-tuotosuhteista tieteellistä toimintaa kuvattaessa kuin KOTA:n koulutusala luokittelu. Nyt käytössä olevaan tilastointiin perustuva tutkimustoiminnan arviointi jää liian karkealle tasolle. Toivottavasti käynnissä olevan valtakunnallisen ns. Raketti-hankkeen ja korkeakoulujen yhteisen tietovaraston myötä tarkempi tilastointi tulee mahdolliseksi nykyistä helpommin.

Samalla toivottavasti korjautuvat viimeisetkin KOTA:n julkaisutietojen käyttöön liittyvät metodiset ongelmat. Aivan viime vuosiin asti yliopistojen tiedonkeruukäytännöt julkaisujen suhteen ovat vaihdelleet siinä määrin, että KOTA-aineistoa ei ole voinut pitää kaikilta osin luotettavana (Pasanen 2007). Käsillä olevien raporttien lopussa esitetään julkaisumäärien vertailu yliopistoittain KOTA-tietokannan ja ISI Web of Science (WoS) -tietokantojen perusteella. Vaikka ei ole aivan selvää, onko tarkoitukse-

na tarkastella KOTA:n luotettavuutta, niin on hyvä pitää mielessä, että kansainväliset ISI Web of Science -tietokannat eivät kata kaikkia maailmassa julkaistavia kansainvälisiä tieteellisiä lehtiä. WoS:ssa olevien julkaisujen määrän perusteella ei siis voi arvioida, onko jonkin yliopiston KOTA-tiedoissa liikaa tai liian vähän kansainvälisiä refereer-artikkeleita.

Mittarit

Kuten tunnettua määrällisillä kuvaajilla voidaan kuvata ilmiötä vain rajallisesti. Tavanomaisesti ilmiön kuvaamiseksi tarvitaan useita mittareita ja määrällistä tarkastelua on täydennettävä laadullisella. Mittarit ovat aina valintoja, jotka rajaavat ja muokkaavat kohteesta tietynnäköisen. Näiden määrä, määrittelyt sekä datan keräämiseen liittyvät seikat saattavat rajoittaa merkittävästikin indikaattoreiden käyttömahdollisuuksia ilmiön kuvauksessa.

Harvoin rajattuihin indikaattoreihin perustuvat kuvaukset saattavat muistuttaa parhaimmillaankin norsun valaisemista kahdella taskulampulla: toisella lampulla valaistaan pimeässä kärsää ja toisella häntää. Yhdistettynä johtopäätöksenä havainnoista on, että kyseessä on pitkänomainen olio joka suippenee häntämäiseksi tupsuksi. Viime vuosina ei ole voinut välttyä vaikutelmalta, että yliopistoja on pyritty kuvailemaan erilaisissa rankinglistoissa tämänkaltaisilla metodeilla ja johtopäätökset ovat olleet vastaavalla tavoin erheellisiä. Ei ole pysähdytty kysymään, kuvaavatko valitut indikaattorit sitä, mitä niiden oletetaan kuvaavan tai, toisin sanoen, ovatko ne valideja.

Paitsi liian harvoin tai ongelmallisesti rajautuviin mittareihin perustuvia, kuvaukset voivat olla luonteeltaan myös normatiivisia: ilmiötä kuvataan sellaisilla osoittimilla, jotka antavat siitä idealisoidun kuvan, tai osoittimilla, jotka kertovat, mihin suuntaan sen tulisi kehittyä. Edellisessä tapauksessa niitä käytetään toiminnan legitimoinnin välineenä, jälkimmäisessä toiminnan kannustimina, esimerkiksi poliittisten tavoitteiden saavuttamiseksi. Reaalisen toiminnan kuvauksen kannalta tämän kaltaisten mittareiden relevanssi on kuitenkin kyseenalainen:

ne mittavat ensisijaisesti sitä, mitä haluttaisiin tapahtuvan, ei sitä, mitä tapahtuu.

Käsillä olevissa RUSE:n raporteissa suunnataan harvojen normatiivisten mittareiden ilmansuuntaan. Tuloksellisuus operationalisoidaan KOTA:an ilmoitetuiksi kansainvälisiksi referee-artikkeleiksi, tohtorin tutkinnoiksi sekä Suomen Akatemialta hankituksi rahoitukseksi. Näitä valintoja perustellaan seuraavasti: ”Kansainvälisiin julkaisuihin hyväksytyt referee-artikkelit ovat yksiselitteisesti perustutkimuksen paras tulosmittari. Tieteellisen toiminnan tuloksellisuuden tarkastelu ilman kv. referee-artikkelien huomioon ottamista olisi vailla pohjaa. Tohtorin tutkinnot ovat keskeinen tieteellisen toiminnan tulos, pohjautuahan kaiken tieteellisen työn jatkuvuus tutkijakoulutukseen. Suomen Akatemian rahoituksen saaminen on hyvä tieteen ja tieteellisen kompetenssin tulostamma, sillä siinä sekä hakemukset että hakijat alistetaan vertaisarviointiin tieteellisen toiminnan laadun verifioimiseksi.” (Kivinen ym. 2008, 9–10.) Tämän enempää valintoja ei kuitenkaan pohdita, vaikka jokainen valinta omalla tavallaan rajaa tarkasteltavaa ilmiötä.

Väite, että ”kansainvälisiin julkaisuihin hyväksytyt referee-artikkelit ovat yksiselitteisesti perustutkimuksen paras tulosmittari” on hieman outo, koska juuri hetkeä aiemmin on todettu otettavan huomioon tieteenalojen erityispiirteet. Tieteenalojen (eli tässä tarkastelussa koulutusalojen) julkaisukäytännöt vaihtelevat ja aloilla ovat varsin selvästi niiden erityispiirteisiin nojaavat omat profiilinsa. Pääasiallisesti kansainvälisiin referee-artikkeleihin keskittyviä aloja ovat Suomessa vain lääketieteelliset alat, farmasia ja psykologia, sekä perusuonnontieteet. Tästä poiketen esimerkiksi teknistieteellisellä alalla julkaisu toiminnan painopiste on ulkomaisissa kokoomateoksissa ja konferenssiartikkeleissa. Selkeästi kansallinen profiili on puolestaan yhteiskuntatieteissä, joiden julkaisu toimintaa hallitsevat monografiat ja artikkelit kokoomateoksissa. Vastaava julkaisu profiili on niin ikään humanistisella alalla, kasvatustieteissä, teologiassa sekä oikeustieteissä. Nämä profiilit ovat selvitettävissä KOTA-aineistojen avulla. (Pasanen 2007; Puuska & Miettinen 2008.)

Tätä taustaa vasten on vaikea nähdä, että arviointiraporteissa kyse olisi tuloksellisuuden arvioinnista muussa kuin varsin rajallisessa ja normatiivisessa mielessä. Voidaan toki väittää, että kansainvälinen julkaiseminen on lisääntynyt kaikilla aloilla, tai että koska koulutusaloja ei vertailla keskenään, niiden erityispiirteet tulevat huomioduiksi. Väittämät eivät poista kuitenkaan tarkastelusta sitä epäsuhtaa joka syntyy panosten ja tuotosten välille. Eikö pitäisi tutkia sitä, mikä on kokonaisuudessaan yliopistossa x alan y kaikkien tutkimuspanosten ja tieteellisten tuotosten suhde? Osa panoksistahan ei käytetty käytettävissä olevien tilastojen mukaan tuottamattomasti vaan muiden kuin kansainvälisten artikkelien tuottamiseen. Nyt aineisto tulee esittäneeksi, että osa resursseista olisi ikään kuin tehottomassa hukkakäytössä ja tehottomina näyttäytyvät erityisesti ne alat, jotka eivät pääasiallisesti julkaise tutkimuksiaan kansainvälisinä referee-artikkeleina.

Mittari kuvaakin enemmän vallitsevaa tiedepoliittista ilmastoa ja näissä keskusteluissa asettuvia tavoitteita kuin todellisuutta. Samalla arviointi hyljeksii muuta kuin tämän normin mukaista toimintaa. Yliopistojen yksiköiden välille se ei myöskään salli perustellusti valittuja tai syntyneitä eroja. Esimerkiksi Suomen historiassa on hyvinkin perusteltua keskittyä suomenkielisten artikkelien ja kirjojen kirjoittamiseen. Ongelmalliseksi erojen sivuuttaminen muuttuu tilastollisessa tarkastelussa edellä keskustelluista koulutusalojen laajuuteen ja määrittelyyn liittyvistä syistä. Hyvinkin erilaiset oppiaineet oletetaan siinä samankaltaisiksi ja niputetaan samaan koulutuslaperheeseen kuuluviksi. Tämä standardointi ohittaa yliopistojen vaihtelevista tieteenalaprofileista, jotka ovat siis eri asia kuin koulutuslaperiilit, syntyvät erot.

Tohtorin tutkintojen merkityksestä ja niiden asemasta tulosindikaattorina on helpompi olla yhtä mieltä. Toisaalta tutkijakoulut ja jatkokoulutuksen tehostamisvaateet ovat korostaneet jatkokoulutuksen koulutuksellista ulottuvuutta. Lisäksi jatkotutkintojen varsinainen tieteellinen tuotos ovat julkaisut, ei tutkinto sinänsä. Jatkotutkinnot ovat joillain tutkintoinflaation vaivaa-

milla aloilla myös ammatillisten lisäkrediittien hankkimista. Olisiko niin, että ainakin joissakin tapauksissa jatkotutkinnot kuvaavat paremmin yliopistojen koulutusfunktion toimintaa (joka toki liittyy tutkimukseen ja tutkimusjärjestelmän uusintamiseen) ja varsinaisen tutkimusfunktion toiminta tulisi katettua paremmin julkaisujen avulla (jatkotutkintojen julkaisut sisältyvät näihin)? Asiaa olisi joka tapauksessa hyvä pohtia laajemmin.

Suomen Akatemian rahoituksen nostaminen tulokseksi herättää runsaasti kysymyksiä. Ajatuksena on ollut, että Akatemian rahoitus ilmentää toiminnan laadukkuutta ja sen määrä indikoisi tällöin koulutusalan y laatua yliopistossa x. Voidaan kuitenkin kysyä, voiko perusrahoituksen ulkopuolinen projektirahoitus olla lainkaan pätevä tutkimuksen tuloksellisuuden indikaattori (Laudel 2005; Schmoch & Schubert 2009). Rahoituksen avulla tehdään tutkimusta, joka tuottaa julkaisuja ja tutkintoja. Toisin sanoen ongelmaksi nousee ilmeinen looginen ristiriita: rahoitus on syytä nähdä panostekijänä, ei tuloksena. Mikäli tämä ongelma vielä ohitetaan, keskittyminen Akatemian rahoitukseen nostaa kuitenkin jälleen esiin alojen väliset erot. Erityisesti tekniikan alalla Tekesin rahoitus on tärkeämpi tutkimusrahoituksen muoto kuin Suomen Akatemian rahoitus. Miksi ei tällöin erilaisten rahoitusprofiilien mukaisesti huomioida myös muuta kilpailtua tutkimusrahoitusta, kuten EU:n ja Tekesin rahoitusta?

Panostekijöinä Kivisen, Hedmanin ja Peltoniemen tuloksellisuusarvioinnissa toimivat professorityövuodet sekä ulkopuolinen tutkimusrahoitus (pois lukien Suomen Akatemialta hankittu tutkimusrahoitus). Näiden valintaa he perustelevat seuraavasti (Kivinen ym. 2009, 7): ”Professorityövuodet ovat relevantein panostekijä tarkasteltaessa tutkimuksen ja tieteellisen jatkokoulutuksen tuloksellisuutta; kuuluaan professoreiden työhön jo määritelmällisesti tehdä tieteellistä tutkimusta ja vastata tieteellisestä jatkokoulutuksesta. Ulkopuolisen tutkimusrahoituksen valintaa toiseksi panostekijäksi voidaan perustella sillä, että tieteellinen työ tehdään käytännössä pitkälti valtion budjetin

ulkopuolelta hankitun tutkimusrahoituksen turvin. Ulkopuolinen tutkimusrahoitus voi olla peräisin yhtäläillä julkisista kuin yksityisistäkin rahoituslähteistä.”

Valinnat herättävät jälleen kysymyksiä. Miksi panoksina on käytetty vain professorityövuosia? Tutkimusta tekevät muutkin kuin professorit. Professorien aika kuluu paljolti hallinnon, ohjauksen ja ulkopuolisen rahoituksen hankinnassa. He toki tekevät tutkimusta ja ohjaavat sitä, mutta puutteellisesta KOTA-aineistostakin on luettavissa, että henkilötyövuosien perusteella noin 75–80 % koko tutkimustyöstä olisi muiden kuin professorien tekemää. Ilmeisesti Kivinen ja kumppanit ovat ajatelleet, että ulkopuolinen tutkimusrahoitus kattaa tämän muun työpanoksen, mutta miksi sitä ei ole esitetty työvuosina, kuten professorien työpanosta? Lisäksi tässä panosvalinnassa jää huomioimatta muun budjettirahoitteisen henkilökunnan tutkimusosuus sekä budjettivaroilla tuetun tutkimuksen osuus.

Vaikka ulkopuolinen rahoitus on kasvanut huomattavasti 1990-luvulta lähtien, Tilastokeskuksen tilastojen mukaan perusrahoitus kattaa edelleen noin puolet yliopistojen tutkimusmenoista. Loogisinta olisikin ollut käyttää kokonaisuudessaan Tilastokeskuksen tilastoimia tutkimuksen henkilötyövuosia, joissa huomioidaan ulkopuolisesti rahoitetun henkilökunnan ohella myös budjettirahoitteinen osuus ns. ajankäyttökertoimien avulla (ks.http://www.stat.fi/til/yaakay/2005/yaakay_2005_2005-12-15_tie_001.html). Kertoimet perustuvat henkilöstön alakohtaiseen ajankäyttöselvitykseen, joten Tilastokeskuksen tiedot antavat KOTA-aineistoa tarkemman kuvan siitä, paljonko henkilöstön aikaa ja resursseja todellisuudessa käytetään tutkimukseen ja toisaalta opetukseen ja hallintoon.

Kaiken kaikkiaan sekä tulostekijöihin että panostekijöihin liittyvät rajaukset sulkevat hyvin paljon ulos yliopistojen todellista toimintaa. Arviointi olisi realistisemmalla pohjalla, jos tilastoidut tulokset ja panokset sisällytettäisiin arvioon mahdollisimman laajasti. Nyt osatuloksia mitataan osaresursseja vastaan ja perustelut sille, että juuri nämä (osa)resurssit tuottavat näitä (osa)tuloksia jäävät vajaiksi.

Laskenta

Varsinainen yliopistojen rankinglaskenta on toteutettu edellä esitetyillä mittareilla, mutta suhteellisten panos-tuotos-laskelmien pohjalta. Laskennassa käytetään kunkin yliopiston tieteenaloittaisia eli koulutusaloittaisia prosentiosuuksia kaikista Suomen yliopistojen panoksista ja tuotoksista kyseisellä koulutusallalla. Tämän perusteella kullekin yliopistolle lasketaan ”sijoitus” sen koulutusallalla edustaman tehokkuuden mukaan. Kuvaajien pohjalta muodostuu kuusi tulos-panos-suhdetta, jotka lasketaan tarkastelujakson jokaiselle viidelle vuodelle. Tämän jälkeen aineistoon sovelletaan tekijöiden kehittämää tuloksellisuuskynnysajatus, koska ”keskikastin ja sen alle jäävien yksiköiden tuloksellisuus ja keskinäiset erot jäsenyisivät osin epäoikeudenmukaisestikin, jos niihin ei sovellettaisi tuloksellisuuskynnystä” (Kivinen ym. 2009, 10). Tuloksellisuuskynnys on laskettu puolestaan indikaattoreista muodostuvien alan tulosaluepisteiden keskiarvoista. Tulos-panosuhteet muunnetaan loppuvaiheessa standardoiduiksi sijoituspisteiksi, joiden pohjalta lasketaan yliopistojen alakohtaiset sijoitukset viiden vuoden ajalta. Näiden pohjalta muodostuu yliopiston lopullinen sijoitus.

Käytetyillä kuvaajilla ajaututaan laskennassa osin loogisiin ongelmiin. Käsillä olevissa arvioinneissa jokainen tulostekijä jaetaan resurssitekijöillä. Näin ollen seurauksena ovat mm. seuraavat indikaattorit: Suomen Akatemian rahoitus/muu ulkopuolinen rahoitus ja tohtorintutkintojen määrä/muu ulkopuolinen rahoitus. Vaikka tuloksellisuuden voidaan väittää olevan laskennallinen abstraktio, sen muodostamien elementtien välillä tulisi olla looginen sidos, jolloin panoksella y tuotetaan puheena oleva tuotos x.

Koska muu ulkopuolinen rahoitus ei toimi resurssina Suomen Akatemian rahoituksen hankinnassa, kyseessä lienee tahaton indikaattoreiden valinnasta johtuva ongelma. Ulkopuolinen rahoitus voi toki tukea niiden kompetenssien ja meriittien syntymistä, joiden avulla Akatemian rahoitusta hankitaan, eli tukea epäsuorasti Akatemian rahoituksen hankintaa, mutta suoraa vaiku-

tussuhdetta näiden kahden asian välille on vaikea nähdä. Vastaava, joskaan ei aivan yhtä vaikeasti perusteltava, suhde syntyy tohtorintutkintojen ja ulkopuolisen rahoituksen välille. Vaikka ulkopuolisen rahoituksen avulla eittämättä tuetaan myös jatko-opiskelua, resurssitekijöistä puuttuvat nyt tutkijakoulujen rahoitus sekä Suomen Akatemian hankerahoitus, jonka avulla on työllistetty paljon jatko-opiskelijoita. Suomen Akatemian rahoitus esitetään sen sijaan tuloksena.

On myös hyvä huomata, että koska kaikki muu kuin Suomen Akatemian tutkimusrahoitus on jakajana (panostekijänä), tutkimusrahoituskilpailussa yleisesti ottaen hyvin pärjäävät yliopistot ja niiden koulutusallat jäävät häviölle verrattaessa niitä yliopistoihin, jotka eivät menesty esimerkiksi Tekesin tai EU:n tutkimusrahoituksen hankinnassa. Hyvään sijoitukseen saattaa riittää se, että koulutusallalla on muutama Suomen Akatemian hanke, mutta ei yhtään tai lähes ei yhtään muuta ulkopuolista tutkimusrahoitusta. Tilanne on vähintäänkin oudoksuttava, koska käytännössä laskenta rankaisee muun kuin Suomen Akatemian rahoituksen hankinnasta. Tämä ei rahoitusmuotojen moninaisuuden ja kilpailullisuuden huomioiden voi olla mielekäästä.

Tuottavuutta tarkasteltaessa tulisi ottaa myös huomioon viive, joka resurssien käyttöönoton ja tulosten välillä aina esiintyy. Tieteellisen työn tulokset toteutuvat harvoin samana vuonna kun rahoitus työhön on saatu. Kun tutkimuksessa on arvioitu tutkimusinvestointien ja -tuotosten välistä viivettä yliopistojen, tieteenalojen tai kokonaisten maiden tiedejärjestelmien tasolla, on päädytty julkaisujen kohdalla 2–6 vuoden viiveisiin, viittausten kohdalla pidempiinkin ajanjaksoihin (Adams & Griliches 1996; Crespi & Geuna 2008). Yhden ja saman vuoden resurssien ja tulosten vertailu antaa tilanteesta väistämättä epärealistisen kuvan, etenkin jos resurssissa tapahtuu paljon muutoksia vuodesta toiseen. Useamman vuoden keskiarvojen käyttö korjaa vuosittaisen resurssivaihtelun ongelmaa, mutta ei ratkaise täysin viiveestä syntyvää kysymystä.

Kaiken kaikkiaan Kivisen ja kumppanien laskennan toteutusta vaivaa valitettava läpinäky-

mättömyys. Ainakin allekirjoittaneille on hyvin hankala seurata saati toistaa toteutettuja operaatioita esitettyjen esimerkkitaulukoiden ja tekstin varassa. Kun kyse on epäilemättä laajaa kiinnostusta herättävästä aiheesta, olisi ollut toivottavaa, että laskenta olisi esitetty yksityiskohtaisemmin ja helpommin seurattavassa muodossa. Samalla perusteluista joillekin toteutetuille operaatioille voisi selkeyttää. Nyt vaikutelmaksi jää, että laskentaa on tarpeettomasti jouduttu monimutkaistamaan erilaisin operaatioin, koska käytetyt indikaattorit ovat ongelmallisia. Esimerkiksi liian harvojen indikaattorien käyttö saattaa johtaa siihen, että erot yksiköiden välillä muodostuvat todella merkittäviksi ja joudutaan ottamaan käyttöön senkaltaisia operaatioita kuin ”tuloksellisuuskynnys”, jotta yliopistojen väliset erot eivät kasvaisi liian suuriksi. Tätä ei kuitenkaan nähdä raporteissa kuvaajien valintaan ja laskentaan liittyvänä haasteena vaan pikemminkin yliopistojen koulutusalojen erilaisuuteen liittyvänä asiana.

Toisinaan ei ole myöskään selvää, miksi jokin asia on toteutettu tietyllä tavalla. Esimerkiksi yliopiston tietyn koulutusalan tulosten ja panosten suhteuttaminen kyseisen alan valtakunnalliseen kokonaisuuteen ja tämän pohjalta lasketut tehokkuusindeksit vaikuttavat aluksi mielekkäältä ajatukselta. Tällä tavoinhan kukin yliopisto tulee suhteutetuksi toisiinsa koulutus- alakohtaisesti. Mutta samaan lopputulokseen päästään kuitenkin yksinkertaisemmin jakamalla vain kunkin yliopiston tuotokset ja panokset koulutusaloittain. Tuloksena olevan tehokkuusindeksin avulla yliopistojen sijoitukset voidaan määritellä nopeasti. Kysehän on koko ajan tietyn yliopiston tietyn koulutusalan tulosten ja panosten suhteesta toisiinsa. Nähdäksemme tuotosten ja panosten suhteuttaminen ensin koulutusalan kokonaisuuteen ja yliopistokohtaisten indeksien laskeminen vasta tämän jälkeen monimutkaistaa laskentaa tarpeettomasti.

Tiedepoliittinen ulottuvuus

RUSE:n arviointiraporttien keskeisenä motiivina lienee ollut se, että tämänkaltaisten tuloksellisuusarvioiden avulla niukat resurssit tulisivat jaetuksi yksiköille, jotka todennäköisesti tuotta-

vat parempia tuloksia. Tekijät toteavatkin tämän varsin yksiselitteisesti vuoden 2008 raportissaan (s. 6): ”Keskeinen kysymys Suomen yliopistolaitoksen kehityksen kannalta on, miten allokoitua käytettävissä olevat panokset eri tieteenaloja edustavien yksiköiden ja eri puolilla maata sijaitsevien yliopistojen kesken siten, että yliopistolaitoksen perustehtävät saadaan optimaalisesti hoidettua ja menestys kansainvälisessä tiedekilpailussa turvattua. Korkeakoulu- ja tiedepoliitiikan päättäjien on tärkeätä tietää, mitkä yksiköt kullakin alalla pystyvät antamaan riittävän vakaat näytöt tieteellisestä tuloksellisuudesta.” Tämän perusteella tekijät ajavat itsestään selvästi ja vallalla olevien tiedepoliittisten uskomusten mukaisesti niin sanotun Matteus-efektin vahvistamista yliopistoissa: niille joilla jo on, annetaan lisää. Samalla muille ilmeisesti muodostuu kannustin tehostaa toimintaansa ja keskittyä indikaattorien osoittamiin toimintoihin.

Raporteille on luonteenomaista, ettei tehtyjä valintoja juurikaan tai lainkaan pohdita. Niinpä raportin arvolähtökohdakin esitetään pragmaattisena valintana, miettimättä lainkaan, olisiko sille olemassa vaihtoehtoja tai onko esitetty näkemys resurssien jakamisesta millään tavoin perusteltu tai tulisiko sitä sovellettaessa ottaa huomioon myös muita näkökohtia. Tuloksellisuus ei esimerkiksi aina ole positiivinen ilmiö: toisesta näkökulmasta samat kuvaajat voivat kertoa liian vähäisestä henkilökuntamäärästä, pitkistä työtunneista ja -viikoista. Pitkäjänteisen ja kestäväen henkilöstöpolitiikan näkökulmasta tällöin pitäisi pikemminkin etsiä tapoja vähentää tehokkuutta. Tätä raporteissa ei kuitenkaan pohdita eikä vaihtoehtoisia tulkintoja esitetä. Kuten monissa yliopistojen vertailuissa, mekaaninen laskenta uhkaa tällöin syrjäyttää harkinnan.

Lähes kaikessa arviointitoiminnassa keskitytään nykyisin liian yksioikoisesti panos- ja tuotosmittareihin. Mittareita on historiallisesti kehitetty ensin kuvaamaan resursseja (tiede- ja korkeakoulupolitiikan alkuaikoina oltiin lähinnä kiinnostuneita resurssien riittävydestä) ja tämän jälkeen tuloksia. Tieteellinen toiminta itsessään ja prosessit, jotka tuottavat tulokset, ovat sen sijaan jääneet vaillinaisen huomi-

on kohteeksi – ”mustaksi laatikoksi” (Godin 2002; Geisler 2005). Osin tämä johtuu siitä, että prosessi-indikaattoreita on vaikea kehittää ja prosessien seuranta ei välttämättä ole vastaavalla tavalla helppo ”automasoida”, kuten panos–tuotos-seuranta. Prosessit ovat vaihtelevia ja vaikeasti määriteltäviä. Kuitenkin esimerkiksi se, miten tutkimustyö on organisoitu ja johdettu, minkälaisin tavoittein sitä tehdään tai onko tutkimusinfrastruktuuri riittävällä tasolla, ovat niitä keskeisiä tekijöitä, joiden varassa tuloksellista tutkimustyötä tehdään (esim. Hurley 1997). Prosessi-indikaattorit ovat haasteellisia, mutta jos yliopistoista halutaan myös niiden kehittämisessä käytettävissä olevaa tietoa pelkkien mekaanisten vertailujen sijaan, olisi perusteltua lähteä etsimään uudenlaisia avauksia yliopistojen seurannan kehittämiseksi.

Kirjallisuus

- Adam, David (2002): Citation analysis: The counting house. *Nature* 415 (6873), 726–729.
- Adams, James D. & Griliches, Zvi (1996): Research productivity in a system of universities. Working paper 5833. National Bureau of Economic Research, Cambridge, MA.
- Becher, Tony (1989): *Academic tribes and territories*. Open University Press, Milton Keynes.
- Billaut, Jean-Charles, Bouyssou, Denis & Vincke, Philippe (2009): Should you believe in the Shanghai ranking? An MCDM view. 29 May 2009. Ladattu: http://hal.archives-ouvertes.fr/docs/00/40/39/93/PDF/Shanghai_JCB_DB_PV.pdf
- Crespi, Gustavo A. & Geuna, Aldo (2008): An empirical study of scientific production: A cross country analysis, 1981–2002. *Research Policy* 37 (4), 565–579.
- Florian, Rozvan V. (2007): Irreproducibility of the results of the Shanghai academic ranking of world universities. *Scientometrics* 72 (1), 25–32.
- Geisler, Eliezer (2005): The measurement of scientific activity: Research directions in linking philosophy of science and metrics of science and technology outputs. *Scientometrics* 62 (2), 269–284.
- Godin, Benoit (2002): Outline for a History of Science Measurement. *Science, Technology and Human Values* 27 (1), 3–27.
- Holmes, Richard (2006): The THES University Rankings: Are They Really World Class? *Asian Journal of University Education* 2 (1), 1–14.
- Hurley, John (1997): *Organization and scientific discovery*. John Wiley & Sons, Chichester.
- Kekäle, Jouni & Lehtikoinen, Markku (2000): *Laatu ja laadun arviointi eri tieteenaloilla*. Psykologian tutkimuksia 21. Yhteiskuntatieteiden tiedekunta, Joensuun yliopisto.
- Kivinen, Osmo, Hedman, Juha & Peltoniemi, Kaisa (2009): *Suomen yliopistojen tieteellisen toiminnan tuloksellisuus vuosina 2004–2008*. Koulutussosiologian tutkimuskeskus, RUSE, Turun yliopisto, Turku.
- Kivinen, Osmo, Hedman, Juha & Peltoniemi, Kaisa (2008): *Tieteellisen toiminnan tuloksellisuus Suomessa 2002–2006*. Koulutussosiologian tutkimuskeskus, RUSE, Turun yliopisto, Turku.
- Kyvik, Svein (1991): *Productivity in Academia. Scientific Publishing in Norwegian Universities*. Universitetsforlaget, Oslo.
- Laudel, Grit (2005): Is external research funding a valid indicator for research performance? *Research Evaluation* 14 (1), 27–34.
- Miettinen, Maija & Selovuori, Jorma (2007): NCR-tietokannan käytettävyyden yliopistojen julkaisujen laadun ja tuottavuuden arvioinnissa. Teoksessa Poropudas, Olli & Miettinen, Maija & Selovuori, Jorma & Pasanen, Hanna-Mari: *Bibliometrinen aineistojen käytettävyyden yliopistojen julkaisujen laadun ja tuottavuuden arvioinnissa*. Opetusministeriön julkaisuja 2007:2. Opetusministeriö, Helsinki.
- Neave, Guy (1998): The Evaluative State Reconsidered. *European Journal of Education* 33 (3), 265–284.
- Pasanen, Hanna-Mari (2007): Kota-tietokannan bibliometrisen aineistojen hyödynnettävyydestä rahoituskriteerinä. Teoksessa Poropudas, Olli & Miettinen, Maija & Selovuori, Jorma & Pasanen, Hanna-Mari: *Bibliometrinen aineistojen käytettävyyden yliopistojen julkaisujen laadun ja tuottavuuden arvioinnissa*. Opetusministeriön julkaisuja 2007:2. Opetusministeriö, Helsinki.
- Puuska, Hanna-Mari & Miettinen, Marita (2008): *Julkaisukäytännöt eri tieteenaloilla*. Opetusministeriön julkaisuja 2008:33. Yliopistopaino, Helsinki.
- Schmoch, Ulrich & Schubert, Torben (2009): Sustainability of incentives for excellent research – The German case. *Scientometrics* 81 (1), 195–218.
- Sipilä, Jorma & Vuorinen, Tiia (2005): Tuloksellisuudesta vailla tuloksia. *Tieteessä tapahtuu* 4/2005, 39–44.
- Van Raan, Anthony (2005): Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics* 62 (1), 133–143.
- Weingart, Peter (2005): Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics* 62 (1), 117–131.
- Whitley, Richard (1984): *The intellectual and social organization of the sciences*. Oxford University Press, New York.

Mika Nieminen on yhteiskuntatieteiden tohtori, joka on toiminut tiede-, teknologia- ja innovaatiopolitiikan tutkijana Tampereen yliopistossa sekä Valtion teknillisessä tutkimuskeskuksessa. Hän työskentelee kehitystehtävissä Tampereen yliopistossa. Otto Auranen valmistelee väitöskirjaa yliopistojen tutkimustuloksellisuuteen vaikuttavista tekijöistä ja on tutkijana tieteen-, teknologian- ja innovaatiotutkimuksen yksikössä Tampereen yliopiston yhteiskuntatutkimuksen instituutissa.