

Vertaileva evaluointi:

Havaintoja korkeakoulujen evaluoinnin alueelta

Pirkko Vartiainen

ENGLISH SUMMARY

The article discusses theoretical and methodological questions of comparative evaluation especially in the area of university evaluations. Comparative evaluation refers to the research in which the evaluation and results produced by the evaluation process are set in a comparative framework. In Finland several universities have been evaluated, mainly at the instigation of the Ministry of Education. However, the executed evaluation processes have not been studied comparatively. In spite of this the evaluation results produced by universities are used, at least partly, as the tool of the comparisons. In light of this fact two questions are deliberated in this article; a) with what kind of principles can the methods of evaluation and comparison be connected, and b) how can comparative evaluation be applied to the evaluation of universities?

JOHDANTO

»Methodologists are the preachers of science. Armed with canons of correct procedure, they have the power to castigate and exhort. They can instruct us to have clearly defined objectives and explicit frames of reference, to base our studies on good theories, and to follow the examples of Marx, Durkheim or Weber.» Tämän väittämän on esittänyt Przeworski (1987: 31), joka jatkaa »... the process of science does not work from rules to practice but from attempt to attempt. ... one good piece of research influences research practice more than many methodology textbooks.»

Tämän artikkelin tarkoituksena on tarkastella Przeworskin, hieman kitkeräksikin tulkittavan, kannanoton innoittamana vertailevan evaluoinnin lainalaisuuksia toisaalta metodologisina kannanottoina ja yleistyksinä, ja toisaalta pohdintoina näiden yleistyksien mahdollisuuksista korkeakouluevaluoinnin sovelluksessa. Vertailevalla evaluoinnilla tarkoitan tässä tutkimusasetelmaa, jossa sekä evaluointiprosessi että evaluointitulosten tarkastelu rakennetaan vertailevan metodin periaatteelle.

Suomessa on tehty useita korkeakoulujen evaluointeja esim. Jyväskylän, Oulun ja Vaasan yli-

opistoissa. Pääosiltaan kyseiset evaluoinnit ovat olleet ns. itsearviointeja, jossa tiedekunnat ja laitokset ovat olleet laajasti mukana, ainakin tuotamassa arviointimateriaalia. Laajamittainen suomalaisen korkeakouluarviointi on lähtenyt liikkeelle opetusministeriön aloitteesta. Implisiittisenä pyrkimyksenä opetusministeriöllä on nähdäkseen ollut se, että korkeakoulut evaluoinnin vakiintumisen myötä voidaan asettaa myös keskinäiseen vertailuun. Jossain määrin tämä vertailuasetelma on toteutunutkin mm. tulosoajauksen ja erityisesti resurssien allokoinnin yhteydessä. On kuitenkin syytä hieman pohtia evaluoinnin ja vertailun metodisia ulottuvuuksia kummallekin menetelmälle ominaisten ja niitä mahdollisesti yhdistävien tekijöiden valossa. Keskeisimpinä kysymyksinä olkoon tässä:

- a) millaisin periaattein evaluointi ja vertailu on yhdistettävissä, ja
- b) miten vertaileva evaluointi on sovellettavissa korkeakoulujen evaluoinnin kenttään?

Tämä artikkeli on osa meneillään olevaa tutkimushanketta, jossa tulen tarkastelemaan korkeakouluevaluoinnin periaatteita, toteutumista ja menetelmällisiä kysymyksiä.

EVALUOINNIN TEHTÄVÄ JA TARKOITUS

Cohen ja Spillane (1994: 323) aloittavat artikkelinsa lausahduksella: »Education indicators are seen as a way to improve education by improving decision-making about education». Tähän lainaukseen tiivistyy nähdäkseen yleinen oletta- ma evaluoinnin hyödyntämisestä johtamisen ja päätöksenteon instrumenttina. Eli, ajatus siitä, että evaluoinnin avulla voidaan tuottaa systemaattista ja toiminnan kehittämistä edesauttavaa informaatiota organisaatioiden toimenpiteistä ja niiden vaikutuksista. Evaluointi ja sen periaatteet ja tarkoitus saavat täten hyvinkin funktionaalisen merkityksen. Mitä nämä evaluoinnin funktionaaliset tavoitteet sitten ovat?

Organisaation tavoitteiden toteutumiseen keskeisillä evaluoinnilla on organisaation toiminnan

kannalta selvä *institutionaalinen funktio*. Institutionaalaisella funktiolla tarkoitan evaluointia, joka pyrkii herättämään ja ylläpitämään organisaation intressi- ja rahoittajaryhmien mielenkiinnon ja tuen. Näin institutionaalista tehtävää toteuttava evaluointi lisää tai vähentää organisaation legitimitettiin. Tuloksetkaksi osoittautuneella organisaatiolla on perusteltu oikeus jatkaa ja kehittää toimintaansa. Toisaalta tässä yhteydessä on syytä muistaa Meyerin & Zuckerin (1989) tutkimus, jossa he pohtivat organisaatioiden kykyä jatkaa toimintaansa entisenlaisena vaikka tulokset jatkuvasti osoittautuisivatkin huonoiksi. Miten tämä on mahdollista? Väitän, että osaksi siksi, että evaluoinnin tuottamaa informaatiota ei organisaatioissa riittävästi hyödynnetä. Kokemukse- ni mukaan organisaatioilla on myös ilmiömainen kyky poisselittää evaluoinnin tuottamat negatiiviset tulokset, mikä puolestaan oikeuttaa luopumaan kehittämistoimenpiteistä. (Vartiainen 1998a).

Organisaation toiminnan kehittämisen kannalta merkittävässä asemassa on evaluointi, jonka tehtävänä on palvella organisaation johtamista. *Johtamisen tueksi* tuotetun evaluointiprosessin tarkoituksena on lisätä informaatiota toisaalta tavoitteiden toteutumisesta mutta erityisesti toiminnan ongelmista ja ongelmien ratkaisusta sekä resurssien allokoinnista. Näin johtamista hyödyttävän evaluoinnin keskeiseksi tavoitteeksi määrittäyty rakenteiden ja prosessien kehittäminen.

Evaluoinnin *teknisellä funktiolla* pyritään organisaation tehtävien sujuvuuden evaluointiin. Teknispainotteinen evaluointi keskittyy analysoimaan organisaation toiminnan tehokkuutta ja laatua palvelujen tai tavaroiden tuottamisessa. Suure- sa osassa toteutettuja evaluointiprosesseja on korostuneet juuri evaluoinnin tekninen tehtävä, sillä ne ovat painottuneet erityisesti panos-tuotos-tyyppisiin tehokkuustarkasteluihin. Toiminnan laadullisten elementtien evaluointi on usein jäänyt, monestakin syystä, harvinaisemmaksi. (Kanter – Summers 1994: 229–231, Vartiainen 1994).

Vaikka evaluointia suoritetaan organisaation toiminnan eri osa-alueilla, on syytä kuitenkin todeta, että laaja-alainenkin arviointiprosessi tuottaa rajallisen kuvan organisaation toiminnasta. Shadish ym. (1991: 37–39) ovat ilmaisseet tämän toteamalla, että evaluointi on ainoastaan osa kompleksia, useita erilaisia riippuvuussuhteita omaavaa, ei-lineaarista toimintaa, jossa pyritään ongelmien ratkaisemiseen. Näin evaluointi saa määritelmän, jonka mukaan se hyvin usein on analysoitu kuva valitusta hetkestä tai ajanjaksosta (Vartiainen 1994). Tässä mielessä vertailu ja

evaluointi ovat lähellä toisiaan. Myös vertailijan on yleensä perustettava päätelmänsä tietyn ajankohdan rajaamaan informaatioon. Näitä analysoituja kuvia voi toki olla useita, jolloin mahdollistuu ilmiön kehitystä ja toimintaa kartoittava vertailu ja/tai evaluointi.

INTRESSIRYHMIEN ROOLI VERTAILEVASSA EVALUOINNISSA

Riippumatta evaluoinnin tehtävästä sen ominaisuuksiin kuuluu aina tietty arvosidonnaisuus. Eri- tyisesti tämä pätee tavoiteperusteisiin evaluointimalleihin, joissa verrataan saavutettuja päämääriä asetettuihin standardeihin. Tämän kaltaisten evaluointiasetelmien keskeiseksi kysymykseksi nousee, kuka määrittelee toiminnan arvot, tarkoituksen ja sisällön ja asettaa ne standardit, joihin toimintaa verrataan?

Edellä esitettyyn kysymykseen on pyritty vastaamaan mm. 'stakeholder' ja 'constituency' tyyppisten evaluointiasetelmien avulla. Esitettyjen evaluointimenetelmien kannattajat korostavat sitä, että evaluoijan on selvitettävä organisaation tärkeimmät asianosaiset: ne tahot, jotka pyrkivät kontrolloimaan organisaation toimintaa sekä ne tahot, jotka ovat tuon toiminnan kohteena. Tämän kautta arvioinnin suorittaja pystyy paremmin tuomaan esiin eri näkemykset siitä, mitä organisaation toiminnan tulisi pitää sisällään sekä suhteuttamaan evaluointitulokset asetettuihin päämääriin. (Stake 1991: 291; Chambers, Wedel & Rodwell 1992: 214; Vartiainen 1994).

Avainryhmäarviointeja edustavat monitahoarviointitutkimukset, joissa on analysoitu palveluorganisaatioiden laadullista ja määrällistä tuottavuutta. Vroeijenstijn (1995) on pohtinut tutkimuksessaan korkeakoulujen laadun sisältöä ja ilmenemismuotoja. Laadulle on vaikea löytää yhtä, täsmällistä merkityssisältöä ja Vroeijenstijn näkeekin sen etsimisen ajan hukkaamiseksi – absoluuttista laatua ei ole olemassakaan. Laatu riippuu siitä, kuka sen määrittelee ja korkeakoulujen laatu saa hänen mielestään sisältönsä eri avainryhmien (stakeholders/constituencies) vaatimusten kautta. Vaatimukset voivat olla keskenään ristiriitaisia, mutta yliopistoissa esim. tiedekunnan tehtävänä on yhteensovittaa nämä ristiriitaiset vaatimukset ja muotoilla niiden pohjalta omat tavoitteensa. Tiedekuntien tavoitteiden pohjalta yliopisto rakentaa oman toimintastrategiansa.

Monitahoarvioinnissa (Vartiainen 1994, 1998b) avainryhmät määrittelevät organisaatiolle sen

tuloksellisen toiminnan muodot. Vertailevassa evaluoinnissa mukaan tulee monia eri organisaatioita ja niiden avainryhmät. On siis oletettava, että tutkimukseen valittujen organisaatioiden tuloksellinen toiminta ja sen eri ulottuvuudet painottuvat hieman eri tavalla eri yksiköissä. Tuloksellisuuden käsite ja sen sisältö pyrkii näin laajenemaan, jolloin on tärkeää pohtia erityisesti vertailevan evaluoinnin tarkoitusta ja funktiota.

Hyödyllisin vertailevan monitahoarvioinnin evaluointifunktio lienee se, että vertailu ymmärretään oppimisena toisten kokemuksista. Lisäksi menetelmä antaa mahdollisuudet vertailuasetelmaan, jossa pohditaan sitä, poikkeako tuloksellisuuden käsitesisältö eri organisaatioissa. Näin pystytään määrittelemään kunkin organisaation tuloksekaan toiminnan sisältö ja taso, jonka perusteella onnistuneet, tuloksekkaat toimintamallit pystytään yhdistämään oman organisaation toimintaan. Viimeisessä vaiheessa tuodaan esiin tutkittujen organisaatioiden parhaimmat ja/tai huonoimmat, mutta myös organisaation kokonaistuloksellisuuteen vaikuttavat piirteet. Tällöin vertailun tuottama analyysi mahdollistaa eri korkeakouluyksiköiden tulosten keskinäisen hyödyntämisen.

Tavoiteperusteisen arvioinnin ongelmiin on pyritty vastaamaan edellä sanottujen mallien lisäksi myös ns. Behavioral objectives -mallin avulla (Cambell 1979; Kanter – Summers 1994), joka muistuttaa lähtökohdiltaan edellä kerrottua monitahoarvioinnin asetelmaa. Behavioral objectives -mallissa tuloksellisuusarvioinnin peruseräkkeet muodostetaan asiantuntijamäärittelyjen avulla. Ajatuksena mallissa on, että asiantuntijat määrittelevät arvioitavan organisaation:

- havainnoitavissa olevat konkreettiset tavoitteet,
- olosuhteet, joissa organisaation tulisi kyseiset tavoitteet saavuttaa, sekä
- asteen, jonka perusteella tavoitteiden saavuttaminen on tyydyttävää.

EVALUOINTIKOHTEN VALINTA

Se, että otetaan kantaa evaluoinnin funktioihin ja arviointisijatahoihin ei kuitenkaan ratkaise vertailevan evaluoinnin kannalta merkittävää seikkaa, eli sitä, miten arvioitavat yksiköt pitäisi valita. Evaluoinnin operationaaliset määrittelyt ovat usein jättäneet tämän kysymyksen käsittelemättä. Asia on kuitenkin ensiarvoisen tärkeä erityisesti silloin, kun pohditaan vertailevan evaluoinnin onnistumisen mahdollisuuksia.

Arvioitavan yksikön valinnassa on monia mahdollisuuksia, jotka toisaalta voivat perustua aikaisemmin tässä artikkelissa esitettyihin funktionaaliin tavoitteisiin tai muihin organisaation toiminnan osa-alueisiin. Toisaalta on todettava, että evaluointikirjallisuudesta ei edellä esitettyyn määrittelyyn löydy yksiselitteistä vastausta. Arviointitutkimusta tehtäessä on kohdespesifisti päätettävä evaluoidaanko yksilöllistä toimintaa, käytöstä tai tyytyväisyyttä (kuten Argyris 1962 ehdottaa); vai organisaatioyksiköiden toiminnan koordinoitua ja tavoitteiden saavuttamista (kuten Pennings – Goodman 1979 tai Manns – March 1978 ehdottavat); vai evaluoidaanko organisaatiota yhtenä kokonaisuutena tai sen toteuttamien ohjelmien varassa (Weiss 1972).

Edellä mainitut tekijät ovat teemoja johon vertailevan evaluoinnin on otettava kantaa. Kysymykseen palataan myöhemmin uudelleen, kun pohditaan korkeakoulujen vertailevan evaluoinnin malleja.

VERTAILEVAN EVALUOINNIN MÄÄRITTELY

Evaluointitapojen ja -menetelmien luokitteluja on esitetty kirjallisuudessa runsaasti. Pyrkimykseni on tässä tarkastella muutamia tyypittelyjä siten kuin se vertailevan evaluoinnin ominaispiirteiden tarkastelun kannalta on oleellista.

Yleisesti evaluoinnissa on eroteltavissa kolme osa-aluetta, joiden varassa useimmat evaluoinnit on toteutettu. Näitä osa-alueita ovat:

- organisaatioiden sisäisiin rakenteisiin ja funktioihin,
- organisaatioon kohdistettuihin ulkoisiin odotuksiin ja
- yhteiskunnalliseen muutokseen liittyvät tekijät.

Yhtenä keskeisistä osa-alueista on mainittava organisaation *sisäisiin rakenteisiin ja sen funktioihin* liittyvä evaluointi. Esimerkiksi palveluorganisaatioiden toiminta on luonteeltaan hyvin asiakaskeskeistä. Asiakasvuorovaikutuksen moninaisuus asettaa vaatimuksia sekä palvelutyötä tekevän henkilöstön ammattitaidolle että työyhteisön toimivuudelle. Siksi organisaation sosiaalinen ympäristö ja sen sisäiset rakenteet ja järjestelmät on huomioitava, kun organisaation toimintaa evaluoidaan vertailevasti. Vertaileva evaluointiasetelma vaatii resurssien, tulosten, sosiaalisten normien, toimeenpanon ja asiakkaiden roolien huomioon ottamista. Tarkastelemalla mainittuja tekijöitä, arvioija pystyy saavuttamaan yleisen, mahdollisesti kattavankin kuvan siitä, mihin

organisaation tavoitteet liittyvät ja miten ne muuttuvat tuotoksiksi ja tuloksiksi.

Toinen arvioinnin kannalta merkittävä osa-alue on organisaatioon kohdistuvat *ulkoiset odotukset*. Mikään organisaatio ei toimi tyhjiössä, vaan muut organisaatiot, poliittiset ohjelmat sekä sosiaaliset ja taloudelliset tekijät vaikuttavat sen toimintaan. Tästä syystä vertailevan evaluoinnin yhteydessä olisi pohdittava myös organisaation ympäristön vaikutusta toisaalta tavoitteiden ja niiden saavuttamisen kannalta ja toisaalta ympäristön esittämien vaateiden kannalta. Tämä periaate laajentaa evaluointiprosessia siten, että organisaatiota tarkastellaan avoimena järjestelmänä. Tuloksellisuuden arvioinnin kannalta esitetty oletus merkitsee sitä, että pelkkää organisaation sisäisen toiminnan evaluointia ei voida pitää riittävänä lähtökohtana. Evaluointiprosessin on huomioitava myös palveluorganisaatioiden ympäristö ja ympäristön vaikutukset, jo pelkästään asiakkaiden odotusten ja tarpeiden kartoittamisen vuoksi.

Kolmanneksi evaluointiprosessissa tulisi perehtyä siihen, miten *yhteiskunnallinen muutos* tapahtuu, kuinka organisaatioiden toiminta ja ympäristö (sekä niiden keskinäinen suhde) ovat muuttuneet sekä siihen, miten nämä muutokset mahdollisesti vaikuttavat organisaation toiminnalla aikaansaatuun vaikutukseen. Näin ollen vertailevassa evaluointiprosessissa tulisi ymmärtää organisaation rooli laajemmin osana yhteiskuntaa, eikä keskittyä pelkästään arvioitavan yksikön sisäisiin implementaatio- ja tuotosprosesseihin. (Shadish ym. 1991: 37–39; Vartiainen 1994: 14–18)

Edettäessä evaluointimenetelmien tyypittelyä käsitteelliseen luokitteluun voidaan todeta, että eri tutkijoiden käyttämä pääkäsitteistö on melko yhtenevä, mutta käsitteiden hienojakoisuuden aste vaihtelee kirjoittajalta toiselle. Esimerkiksi voidaan ottaa Ammonsin (1995: 41) monelta osin hyvin perinteinen luokittelu. Hän jakaa arvioinnin ja mittaamisen käsitteellisesti neljään luokkaan, joita ovat workload, efficiency, effectiveness ja productivity. Workload -käsitteen sisältö viittaa organisaation aktiviteetteihin ja ponnistuksiin. Workloadin mittaaminen tuottaa raakamateriaalia, jonka perusteella on mahdollista tietyin määräjain kerätä ja analysoida organisaation aktiviteettien kehitystrendit. Muilta osin Ammonsin määrittelyt noudattelevat totuttuja; tehokkuus mittaa tuotoksien määrän ja tuotannossa käytettyjen resursien määrän väliseen suhteeseen ja vaikuttavuus tarkoittaa joko palveluiden laatua taikka sitä, missä määrin toiminnan tavoitteet on saavutettu. Suomalaisen evaluointitutkimuksen käsitelmäärittely-

jen osalta kannattaneet tutustua esim. Ahosen (1985, 1991) ja Lumijärven (1994) tutkimuksiin.

Vertailevan arvioinnin kannalta mittaamisessa käytettyjen käsitteiden määrittely on tärkeää, koska vertailu lisää entisestään analyysin vaativuustasoa. Vaativuustaso kohoaa erityisesti siksi, että vertailevaan evaluointiin perustuvassa tutkimusasetelmassa käsitteitä käytetään sekä evaluointiasetelman toteuttamisen ja analyysin vaiheissa että analyysitulosten vertailun vaiheissa. Tämä pätee erityisesti silloin kun vertailuasetelmaan sisältyy organisaatio tai maa kohtaisia vertailuja, joissa kulttuuriset tekijät (esim. erot järjestelmien ja/tai organisaatioiden epävirallisissa ja virallisissa kulttuureissa) voivat antaa käytetyille käsitteille erilaisia sisältöjä.

VERTAILEVAN EVALUOINTITIEDON MERKITYS

Vertailevan metodin käytölle on olemassa monia syitä. Salminen – Lehtinen (1982: 88) määrittelevät yhteiskuntatieteellisen vertailun yhteisöjen ja inhimillisten järjestelmien erojen ja yhtäläisyyksien analysoimiseksi. Samaa korostavat myös useat muut kirjoittajat (Przeworski 1987; Ragin 1987; Mayer 1989), joiden mukaan vertailun avulla pyritään ymmärtämään, selittämään ja tulkitsemaan ilmiöitä. Kiistämättä edellä esitettyjä käsityksiä on kuitenkin todettava Sartoria (1991: 224) mukaellen, että *"comparing is controlling"*. Vertailua tehdään ymmärtämisen ja selittämisen vuoksi mutta erityisesti siksi, että voidaan kontrolloida (toteen näyttäen tai kumoten) pitävätkö yleistyksen paikkansa silloin kun niitä sovelletaan eri tapauksiin.

Kontrolloimisen, ymmärtämisen ja selittämisen lisäksi vertailu on myös oppimista, joka perustuu informaatioon muiden kokemuksista. Tämä nykyisin ns. benchmarkingiksi kutsuttu menetelmä asettaa vertailun oppimisen välineeksi siten, että sen avulla on mahdollista lisätä tietoa erilaisista toimintavaihtoehdoista. Eräällä tapaa benchmarkingia voidaan pitää jopa vertailevan evaluoinnin menetelmänä. Tutkimalla esim. muiden yhteiskuntien tai organisaatioiden toimintaa, voidaan saavuttaa laajempi ymmärrys yhteiskunnallisista, hallinnollisista ja poliittisista vaihtoehdoista, joiden avulla oman toiminnan hyviä ja huonoja puolia on mahdollista analysoida (Almond – Powell 1992: 3). Käytettäessä vertailevaa evaluointia kehittämisen välineenä saavat tässä artikkelissa aiemmin käsitellyt evaluoinnin funktiot entistä laajemman merkityksen. Erityisesti organi-

saatioiden johtamisen ja kehittämisen tueksi läpiviety vertailevan evaluoinnin prosessi voi tuottaa huomattavasti enemmän hyödynnettävää informaatiota kuin evaluointi, jossa arviointiasetelma rakennetaan esimerkiksi yhden organisaation poikkileikkauksen varaan.

Vertailevan evaluoinnin kannalta ovat huomionarvoisia myös ne mielipiteet, joita eräät tutkijat, mm. Roberts (1985: 287–289) esittävät todetessaan, että vertailu on kaiken tiedon hankkimisen perusta. Tämän näkemyksen mukaan ymmärtäminen ja tiedon kartuttaminen on ilman vertailua täysin mahdotonta. Väite on vahva, mutta on totta siinä mielessä, että analysoitavan tiedon tulee olla asetettavissa ainakin oman kontekstinsa vertailuasetelmaan, jotta tutkittavasta yksiköstä voidaan sanoa jotakin – käsittääkseni ajatus pätee erityisen hyvin evaluointitutkimusten tuottaman informaation analysoimiseen.

TUTKIMUSKOHDTE JA VERTAILUN TASO

Vertailumetodin käyttö vaatii erityisesti paneutumista tutkimuksen strategiaan valintoihin, jolloin on määriteltävä vertailun taso, yksikkö ja ajallinen dimensio. Vertailun yksiköjä voivat olla sekä poliittiset että organisatoriset systeemit kokonaisuudessaan että niiden alajärjestelmät. Vertailun taso erotellaan yleensä joko järjestelmän sisäiseksi (esim. kuntavertailu, organisaation eri osien vertailu) tai järjestelmien väliseksi (esim. maa vertailut, yksityinen – julkinen vertailu). Vertailun aikadimensiolla tarkoitetaan vertailun sitomista tiettyyn ajanjaksoon tai hetkeen, mutta myös eri ajanjaksoihin perustuvaa historiallista vertailua. Edellä sanotut periaatteet eivät kuitenkaan ole pelkästään vertailututkimukselle asetettavia vaatimuksia, vaan kaikkea tutkimusta ja erityisesti evaluointitutkimusta koskevia vaatimuksia.

Toinen vertailevan tutkimuksen strateginen valinta on se valitaanko tutkimuskohteeksi mahdollisimman samankaltaiset vai mahdollisimman erilaiset tapaukset. Tämä pohdinta on suoraan verrannollinen jo klassiseksi muodostuneeseen kysymykseen siitä, milloin eri tapaukset ovat vertailtavissa. Sartori (1991: 245–246) pohtii vertailtavuutta toteamalla, että vertailussa tulee katsoa tapausten ominaisuuksia, eli sitä, milloin tapaukset jakavat keskenään tiettyjä samoja ominaisuuksia ja milloin eivät. Hyyryläisen (1995: 4–5) mukaan saman kaltaisten ja lukumäärältään vähäisten tapausten vertailussa etsitään tapausten välisiä eroja. Sen sijaan silloin kun tapausten määrä on suuri tai ne ovat etäällä toisistaan on



Kuvio 1. Vertailtavien tapausten samankaltaisuus vs. erilaisuus.

tyypillistä, että vertailussa keskitytään yhtäläisyyksien analysoimiseen.

Vertailevan evaluoinnin kannalta kysymys tutkimuskohteiden samankaltaisuudesta tai erilaisuudesta on vähintään yhtä merkittävä kuin missä tahansa vertailututkimuksessa. Olen pyrkinyt havainnollistamaan vertailtavuuden ideaa kuvioon yksi. Kuvio hahmottaa vertailtavien tapausten sijoittumisen jatkumolle, jonka ääripäinä ovat täysin samankaltaiset tapaukset ja täysin erilaiset tapaukset.

Kuvion yksi oletuksena on, että samankaltaisten tapausten tutkimuksessa on mielekästä analysoida tapausten välisiä eroja, kiinnostavaa ei ole samankaltaisten tapausten yhtäläisyyksien kuvaaminen. Sama pätee jatkumon toisessa päässä, analyysiin mielekkyys katoaa, jos vertailtava evaluatio keskittyy erilaisten tapausten erisuuksien korostamiseen.

On kuitenkin ilmeistä, että käytännön vertailututkimukset eivät asetu kuvion osoittamiin ääripäihin, sillä yleensä täysin samankaltaisilta näytävät ilmiöt sisältävät monia erilaisia ominaisuuksia. Tällöin samankaltaisten yksiköiden valinnassa painotetaan yksiköiden samankaltaisuutta, mutta tiedostetaan valittujen yksiköiden erilaisuus tutkimukseen valittujen riippumattomien muuttujien kautta. Vertailevan evaluointitutkimuksen kannalta ongelmaksi voi kuitenkin muodostua se, että tutkimusasetelmassa ei pystytäkään tunnistamaan kaikkia yksiköiden välisiä merkittäviä eroja.

Tarkasteltaessa muutamia vertailututkimuksia voi havaita, että mm. Etzioni-Halevy (1990) on valinnut tutkimusasetelmaansa mahdollisimman samankaltaiset maat. Vertailuyksiköiden valinnan kriteereinä tutkimuksessa käytettiin maiden parlamentaaristen systeemien ja niiden rakenteiden samankaltaisuutta. Tämä takasi kirjoittajan mukaan sen, että tutkimuksessa pystyttiin paremmin havaitsemaan parlamentaarisen poliittisen eliitin toiminta ja roolit sekä havainnollistamaan eliitin käyttäytymisen piirteet. Täysin erilaisten maiden valitseminen poliittisen eliitin toiminnan

ja roolien kartoitukseen olisi oletettavasti johtanut toisaalta valittua asetelmaa huomattavasti useampien roolien havaitsemiseen, mutta estänyt samalla roolien ja käyttäytymistapojen todellisten syiden analyysin.

Hieman toisen tyyppinen esimerkki samanlaisuuden tavoittelusta vertailukohteiden valinnassa on Vuoren (1995) tutkimus. Samanlaisuus on kyseisessä tutkimuksessa pääasiassa valittuun tutkimusmateriaaliin ja organisaatio-osien tiettyihin piirteisiin liittyvää (esim. tutkitut pöytäkirjat olivat sisätautiasastoilta, jotka olivat henkilöstömäärältään ja ikärakenteeltaan mahdollisimman samankaltaisia ja jonkin hallinnollisen muutoksen kohteena). Kokonsa ja hallinnollisten järjestelyidensä puolesta valitut organisaatiot erosivat toisistaan melko paljon, joten tässä suhteessa valittujen kohteiden samankaltaisuus muodostui rajoittuneeksi.

Korkeakoulupolitiikan vertailevassa case-tutkimuksessa etsittiin samankaltaisten tapausten avulla selitettäviä tekijöitä eri maiden korkeakoulupolitiikan erilaiseen tai samanlaiseen kehitykseen. Tutkimuksen tekijä piti valitsemaansa tutkimusstrategiaa varsin hedelmällisenä. Ko. tutkimuksessa pyrittiin etsimään sellaisia tekijöitä, jotka ovat vaikuttaneet tutkimuksessa mukana olevien maiden akateemisen arvomaailman, työn ja identiteetin muotoutumiseen. (Kogan 1996)

Samankaltaisten kohteiden vertailuun liittyy myös ongelmia, joita mm. Ragin (1987: 9–10) nostaa esiin. Näistä ensimmäistä ongelmaa voidaan kuvata termillä samanlaisuuden näennäisyys. Raginin mukaan jokaisella tutkimukseen valitulla kohteella on oma historiansa ja identiteettinsä, joten tutkimuskohteiden määrittely 'samanlaisiksi tapauksiksi' on käytännössä melko vaikeaa. Tämä ongelma käy sitä todellisemmaksi, mitä tarkempaan teoreettiseen ja/tai empiiriseen erittelyyn pyritään, koska tutkimuksen kannalta relevanttien tapausten määrällä on tuolloin taipumus pienentyä. Mikäli tutkimukseen valittujen tapausten määrä käy liian pieneksi se voi johtaa olemassaolevien eroavaisuuksien yli-korostumiseen ja analyysin vääristymiseen.

Tutkimuksen vertailukohteiksi voidaan valita myös mahdollisimman erilaiset yksiköt eli tapaukset, jotka eroavat toisistaan keskeisten muuttujien tai tekijöiden suhteen. Tämänkaltaisen tutkimusotteen tarkoituksena on osoittaa, että valituissa kohteissa pätevät tietyt säännönmukaisuudet vaikka vertailuun valitut yksiköt ovat erilaisia. Lähestymistapaa pidetään hyödyllisenä silloin, kun analyysin tavoitteeksi ei aseteta sen kuvaamista miten valitut tapaukset eroavat toisistaan,

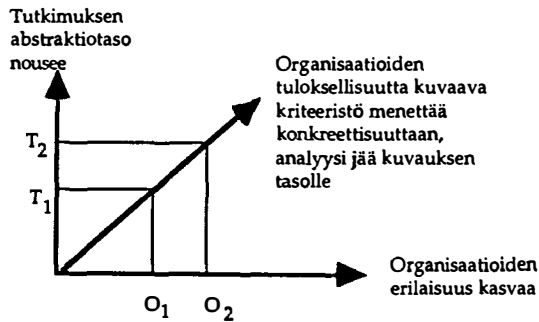
vaan pohditaan sitä, missä suhteessa ne jakavat selitettävänä olevat yhteiset muuttujat. Ongelmia tässäkin tapauksessa tuottaa se, että ei voida varmuudella tietää onko tutkimukseen löydetty eri yksiköiden kaikki yhteiset tekijät. Tutkimukseen valittujen maksimaalisesti erilaisten tapausten vertailun katsotaan olevan tarkoituksenmukaista silloin kun pyrkimyksenä on ideaalityyppien muodostaminen tai silloin kun asetelmalla tehtävät mittaukset eivät vaadi korkeaa reliabilitteettia. (Roberts 1985: 294).

KÄSITTEELLINEN YMMÄRRYS

Vertailevan evaluoinnin onnistumisen keskeisin lähtökohta lienee kumpaankin menetelmään liittyvä käsitteellinen ymmärrys. Relevantti vertailuasetelma vaatii käsitteiden selkeyttämistä siten, että eri osapuolet (tutkijat, kohderyhmät, poliitikot jne.) tulkitsevat ja ymmärtävät termit samalla tavalla. Arvioinnissa pätee samaa logiikka, tuloksellisuuden sisällön määrittäminen sekä indikaattoreiden ja kriteereiden operationalisointi näyttelevät merkittävää osaa evaluointitulosten tulkinna.

Mikäli korkeakoulujen tuloksellisuusarviointi toteutetaan vertailuasetelmana, on arvioitavia kohteita analysoitava yhteisin käsittein ja kriteerein. Kuviossa kaksi olen pohtinut vertailuasetelman vaikutusta tutkimuksen abstraktiotason ja kriteeristön kannalta tilanteessa, jossa arvioidaan samanlaisia vs. erilaisia organisaatioita. Tutkimuksen abstraktiotaso on kuvattu pystysuoralla jatkumolla, jossa T_1 kuvaa matalampaa abstraktiota-soa ja T_2 korkeampaa. Vastaavasti vaakasuoralla jatkumolla kuvataan organisaatioiden erilaisuutta, joka kasvaa origosta pois päin siirryttäessä. O_1 ja O_2 kuvaavat tutkimuksessa mukana olevien organisaatioiden erilaistumisasteita sen mukaan miten erilaisiksi ne luokitellaan tai ymmärretään. Oletuksena on, että organisaatioiden erilaisuuden kasvaessa on entistä vaikeampi löytää sellaista kriteeristöä, jonka avulla organisaatioiden tuloksellisuutta voidaan evaluoida ja vertailla kaikissa valituissa yksiköissä. Tämä johtuu pääasiassa tutkimuksen abstraktiotason nousemisesta, jolloin organisaatioiden arvioinnin konkreettisuus pienenee eikä tutkimustulosten analyysissä päästä kuvaamisen tasoa pidemmälle.

Lisäksi on oletettava, että tasolla O_2 tai sitä etäämpänä origosta olevien organisaatioiden erilaisuuden kasvaessa tuloksellisuuden mittaaminen jää osittain toissijaiseksi tehtäväksi, koska tuloksellisuutta kuvaava kriteeristö on laadittava



Kuvio 2. Vertailevan evaluoinnin vaikutus tutkimuksen abstraktiotasoon.

niin laajaksi, että se kattaa kaikki organisaatiot ja sopii niiden toimintaan. Ongelmaksi muodostuu tällöin se, että kriteeristöllä ei enää kyetä analysoimaan eri organisaatioiden tuloksellisuutta ja sen osa-alueita riittävän yksityiskohtaisesti ja tarkasti. Perustelen väitettäni myös Etzioni-Halevyn (1990: 118–119) huomioilla vertailutason merkityksestä. Vertailutasolla tarkoitetaan, kuten aikaisemmin todettiin, vertailuun valittua kokonaisuutta, joka tutkimusprosessissa jaetaan vertailtavissa oleviin "viipaleisiin". Ongelmia tuottaa yleisimmin vertailutason mikrotasoinen viipalointi, koska havainnot muodostuvat näin hyvin rajallisesti yleistettäviksi, eikä niillä enää ole kykyä tuottaa johtopäätöksiä makrotasolla. Jos tarkastelu kohdistuu tietyn systeemin osasysteemeihin on ongelmana tai vaarana taas se, että systeemi kokonaisuudessaan jää hahmottamatta, koska erilliset osasyteemit eivät pysty kuvaamaan systeemiä kokonaisuudessaan.

Mielenkiintoisen lisäolottuvuuden kuvion kaksi oletukseen antaa pohdinta vertailevan evaluoinnin toistettavuudesta. Oletus on, että esim. kohtaan T₁ asettuva evaluointimenetelmä on huomattavasti helpommin toistettavissa kuin esim. T₂ kohdan menetelmä. Otetaan esimerkiksi SERVQUAL-menetelmä, joka kuvioon kaksi sijoitettuna asettuisi melko lähelle origoa koska menetelmä on samanlaisten organisaatioiden evaluointiin rakennettu, melko pitkälle standardoitu malli. SERVQUAL:in keskeisenä teesinä on oletus, jonka mukaan palvelun laatu kulminoituu toisaalta asiakkaiden odotuksiin ja toisaalta heidän kokemuksiinsa. Evaluoinnissa mitataan (valittujen organisaatioiden) asiakkaiden odotusarvojen ja kokemusarvojen välinen ero, jolloin tutkimustulokset konkretisoituvat ns. SERVQUAL -pisteiksi. Vertailu voidaan sitten tehdä näiden dimensio-pisteiden varassa. Kokemukseni mukaan SER-

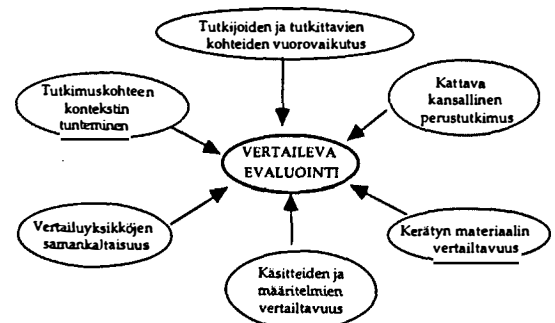
VQUAL-menetelmä on helposti toistettavissa oleva arviointiprosessi. (Vartiainen 1997)

VERTAILEVA EVALUOINTI KORKEAKOULUISSA

Koulutuksen evaluointia on tehty runsaasti, mutta merkittävä määrä tästä arvioinnista on ollut kansallisvaltioiden sisällä tapahtunutta evaluointia. Näin myös tehtyjen tutkimusten metodologiset valinnat, aineiston kokoamisen periaatteet ja käytetty käsitteistö on ollut varsin maaspesiifiä. Tästä syystä kansainväliset vertailevat koulutuksen evaluointitutkimukset on usein tuomittu epäjohtonmukaisiksi ja irrelevanteiksi. Mm. Creemers & Reynolds (1996: 258–259) toteavat, että kansallisissa tutkimuksissa löydettyjen havaintojen ja toimintatapojen vertailu ja vielä vähemmän niiden siirtäminen kulttuurista toiseen on melko kannattamatonta ja vaikeaakin.

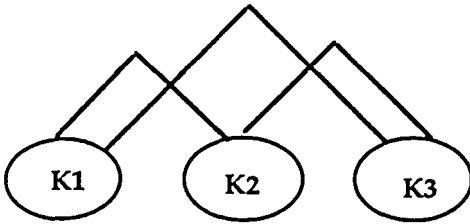
Kansainvälisessä koulutusjärjestelmien vertailussa ongelmia tuottaa se, että koulutuspolitiikan taustaolettamukset, tavoitteet, ohjausjärjestelmät yms. eroavat maittain. Kuvioon koolme olen kerännyt vertailun onnistumista parantavia tekijöitä Laukkasen (1997) ja Salmisen (1998: 203–204) kirjoitusten pohjalta. Noudattamalla näitä periaatteita on mahdollisuus päästä tutkimustulosten parempaan hyödynnettävyyteen ja luotettavuuteen. Mielenkiintoiseksi kuvion tekee se, että esitetyt kriteerit ovat sovellettavissa myös evaluointitutkimuksen luotettavuuden varmistamiseen.

Edellä sanotun perusteella on vertailevan evaluoinnin näkökulmasta todettava, että evaluointiasetelma sekä indikaattoreiden ja kriteereiden muodostaminen näyttäisi olevan mahdollista parhaiten silloin kun evaluointikohteiksi valitaan mahdollisimman samankaltaiset yksiköt. Tässä

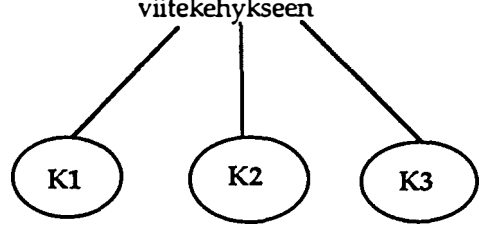


Kuvio 3. Vertailun ja evaluoinnin onnistumista parantavia tekijöitä.

Vertaileva evaluointi
kunkin korkeakoulun tuloksellisuuden
ominaispiirteet keskinäisessä vertailussa



Vertailevan evaluoinnin viitekehys
kunkin korkeakoulun tuloksellisuuden
ominaispiirteitä verrataan tutkimuksen
viitekehykseen



Kuvio 4. Korkeakoulujen vertailevan evaluoinnin mallit.

mielessä vertaileva evaluointi noudattelee pääasiassa suoraa vertailuasetelmaa hyödyntävän analyttisen vertailun (Bonnell 1960: 164–165) periaatteita. Analyttisessä vertailussa on tarkoituksena etsiä samankaltaisten yksiköiden ja ilmiöiden välillä ilmeneviä riippumattomia suhteita, muuttujia tai selittäviä tekijöitä, joiden avulla pystytään selittämään tutkimuskohteiden välillä vallitsevat erot ja/tai yhtäläisyydet. Näin myös vertaileva evaluointi analysoi esim. eri korkeakouluja keskenään, jotta pystyttäisiin löytämään korkeakoulujen yhteiset selittävät tekijät tai säännönmukaisuudet, joiden pohjalta rakennetaan selittäviä yleistyksiä korkeakoulujärjestelmästä.

Toisaalta mikäli tutkimusasetelmassa päädytään siihen, että evaluointikohteiden (organisaatioiden) keskinäinen vertailu ei ole mahdollista päädytään vertailevassa evaluoinnissa lähelle epäsuoraa vertailuasetelmaa hyödyntävän illustratiivisen vertailun periaatteita. Illustratiivisessa vertailussa organisaatioita ja ilmiöitä ei vertailla keskenään, vaan valittuun teoreettiseen viitekehyyksen tai käsitteistön kautta. Tällöin vertailulla hahmotetaan tutkittavien ilmiöiden ja teoreettisen viitekehyyksen tai käsitteistön ajatuksellista vastaavuutta, eli tarkastelussa käytetään hyväksi ideaalimalleja. Tässä mielessä illustratiivisen vertailun periaatteet ovat varteenotettavia silloin, kun pohditaan eri organisaatioyksiköiden vertailtavuutta ja sitä käsitteistöä ja mittaristoa, jolla eri yksiköitä voidaan vertailla evaluoida.

Korkeakoulujen vertailevan evaluoinnin kannalta saattaisi olla hyödyllistä noudatella Raginin (1987) mallin periaatteita, jossa vertailussa käytettävä informaatio kerätään analyysiyksikköjen tasolla ja selittäminen suoritetaan analyysiyksikköjen ominaisuuksien tasolla. Korkeakoulujen vertailevan evaluointitutkimuksen sovelluksena tämä merkitsisi sitä, että evaluointi painotetta-

siin mikrotason toimintoihin (kuvio 4). Tutkimusasetelmallisesti kyseinen ratkaisu noudattelisi monella tapaa Bonellin (1980: 165) ajatuksia vertailun periaatteista.

Kuviossa neljä ensimmäinen kokonaisuus viittaa analyttiseen evaluointiin, jossa korkeakoulu (K1, K2, K3) asetetaan suoraan keskinäiseen vertailuun. Toisin sanoen evaluointitulosten analyysin painopisteenä on muodostaa kuva (valittujen kriteereiden perusteella) kullekin korkeakoululle ominaisista tuloksellisuuden piirteistä. Tämän jälkeen arvioinnissa mukana olevat korkeakouluyksiköt asetetaan keskinäiseen vertailuun saatujen analyysitulosten valossa. Menetelmä lienee hyödyllinen erityisesti silloin, kun evaluointiprosessiin sisällytetään ainoastaan muutama korkeakoulu, jolloin arvioinnin tarkoituksena ei ole kansallisen koulutuspolitiikan kokonaisarviointi, vaan järjestelmän osien arviointi.

Kuvion toinen osa kuvaa illustratiivista asetelmaa, jossa korkeakouluja ei verrata keskenään vaan tutkimuksen viitekehyykseen ja siinä muodostettuun ideaaliseen evaluointimalliin. Tässä tapauksessa konkreettinen sovellus saattaisi toimia siten, että vertailevan evaluoinnin viitekehys muodostetaan »standarditasoksi», johon eri arvioitavien yksiköiden evaluointitulokset sitten suhteutetaan. Edelleen kehittelyä tämän mallin osalta vaatii se, millaisin periaattein tämä mahdollinen korkeakoulujen standarditaso luodaan. Evaluoinnin illustratiivinen asetelma lienee hyödyllisin silloin, kun halutaan testata eri menetelmien toimivuutta vertailevassa evaluoinnissa.

POHDINTAA

Korkeakoulujen vertailevan arvioinnin yksi keskeisimmistä ongelmista on kysymys arvioidaanko

määrää vai laatua. Käytäntö on osoittanut, että korkeakoulujen rahoituksen supistuminen 1990-luvun alussa aiheutti arviointikeskustelun painottumisen pääasiassa taloudellisen tehokkuuteen ja määrällisiin tuloksiin. Näin myös kysymys resurssien allokoinnista on ollut hallitseva trendi, KOTA-tietokanta on tästä hyvä esimerkki. Laadullisen evaluoinnin tekeminen korkeakoulujen viitekehyksessä on hankalaa ja tästä johtuen usein joudutaan tyytymään määrällisiin kriteereihin ja indikaattoreihin. Nykyisin korkeakoulujen tuloksellisuuden arviointikeskusteluissa laadullisia kriteerejä ollaan liittämässä ja institutionalisoimassa määrääjäksi valittaviin tutkimuksen ja opetuksen huippuyksiköihin. Käytännössä laadullinen mittaaminen on paljon korkeakoulujen tekemien itsearviointien varassa. Käsitykseni mukaan laadullinen evaluointi on mahdollista evaluoinnin viitekehyksessä, jossa laatu määrittyy avainryhmiltä saadun evaluointi-informaation kautta. Mielenkiintoista on se, että monitahoarvioinnin menetelmä ei sulje pois määrällisen evaluointi-informaation tuottamista. Määrälliset indikaattorit ja kriteerit voidaan yhdistää laadulliseen evaluointiprosessiin.

Määrällisen vai laadullisen arvioinnin lisäksi on ratkaistava muitakin yleisiä evaluointiperiaatteita. On pohdittava mm. sitä, miten opetusohjelmien ja opetuksen tavoitteiden tasosta päätetään. Korkeakoulujen laadun evaluoinnin yhteydessä tulisi ratkaista, mitä ovat ne konkreettiset tavoitteet, tiedot ja taidot, jotka opetuksella tulee saavuttaa? On myös pystyttävä ratkaisemaan kysymys evaluointitiedon käyttötarkoituksesta ja luonteesta; tuotetaanko luottamuksellista vai avointa informaatiota. Eli, suunnataanko evaluoinnin tulokset koulutuksen käyttäjien, vanhempien ja muiden sidosryhmien käyttöön vai ainoastaan koulun sisällä johdon ja organisaation omaan kehittämistyöhön. Kolmas tärkeä ratkaisu vertailevaan evaluointiin ryhdyttäessä on miettiä, miten evaluointitietoa voidaan hyödyntää? Periaatteet evaluoinnin tulosten hyödyntämisestä on syytä liittää osaksi evaluointiprosessia.

Evaluointitulosten hyväksikäytöstä yhtenä esimerkkinä voidaan käyttää Ammonsin (1995: 39) kokemuksia kirjastotoimen evaluointiprosessista. Hänen kokemuksensa mukaan julkisen sektorin toiminnan arvioinnit jäävät usein vaille huomiota, niistä eivät ole kiinnostuneita sen paremmin yhteiskunnan jäsenet kuin poliittiset päättäjäkään. Sen sijaan evaluointitutkimukset, joissa on vertailtu saatuja tuloksia joihinkin standardeihin, herättävät huomiota paljon enemmän. Päättäjät ja kansalaiset ovat kiinnostuneita, ja heidän on myös helpompi ymmärtää tutkimuksen johtopä-

tökset silloin, kun tutkimuksessa vertaillaan organisaatioita keskenään tai tiettyjä tekijöitä tiettyihin standardeihin tai aikaisempiin tuloksiin.

Vertailuasetelmaa hyödyntävät evaluointitutkimukset eivät kuitenkaan ole kovinkaan yleisiä. Syynä voi olla esim. se, että organisaation henkilöstö saattaa vieroksua tai jopa pelätä vertailuasetelmaa, jossa verrataan organisaatioita keskenään. Tästä syystä käytännön evaluointityö hankaloituu ja tutkijoiden on vaikea koota tarvitsemansa tutkimusmateriaali. Merkittävin ongelma on kuitenkin se, että ei ole pystytty kehittämään organisaatioiden toimintaa kuvaavia taroituksenmukaisia standardeja, joihin vertailevan evaluoinnin tuloksia voitaisiin relevantisti peilata. Tämä on kysymys, johon vertailevan evaluoinnin menetelmien edelleen kehittämisessä tulisi kiinnittää laajempaa huomiota.

Mittariston osalta vertailevassa evaluoinnissa tulisi pohtia seuraavia kysymyksiä. Ensinnäkin on ratkaistava, miten laajalla mittaristolla evaluointia kannattaa lähteä suorittamaan. Oletettavaa on, että ellei käytössä olevien mittareiden määrää rajata, voidaan päätyä tilanteeseen, jossa tulosten analyysi hajoaa pirstaletiedoksi, menettäen näin mahdollisuutensa laaja-alaisen tulosten kuvaamiseen. Toiseksi, käytettävien mittareiden taso on vertailevassa evaluointitutkimuksessa myös merkittävä. Oletettavaa on, että mittariston tulisi sisältää sekä makro- että mikrotason mittareita. Makrotason mittareita ovat laajat esim. kunnan tai yhteisön tasolla operoivat mittarit. Mikrotason mittaristo pitää sisällään esim. yksilötason tai tietyn organisaatio-osan indikaatteja. Kolmanneksi, vertailevassa evaluoinnissa tulisi pohtia käytettyjen mittareiden painotusta, joka kuvaa sen, mikä on eri mittareiden painoarvo kokonaistuloksellisuuden kannalta. Tämä kysymys liittyy läheisesti analyysin tason valintaan, koska analyysitaso määrittää myös mittariston ja eri mittareiden painoarvot (vrt. kuvio 4). Kokemukseni mukaan (Vartiainen 1994, 1997) kattaviin tuloksellisuutta kuvaaviin johtopäätöksiin voidaan päästä vasta yhdistelemällä eri mittareiden ja indikaattoreiden tuottama informaatio.

Edellä sanotut tekijät pitävät sisällään myös tiettyjä ongelmia. Vertailevan evaluoinnin osalta jää kysymykseksi esimerkiksi se, saavatko eri mittarit samanlaisen painoarvon vertailtavissa yksiköissä, vai onko yksiköiden toiminnassa sellaisia oleellisia eroavuuksia, että mittareiden painottaminen tulisi tapahtua yksikkökohtaisesti. Oletettavaa on, että vertailevan evaluoinnin tulosten yhteismittallisuus kärsii, jos mittareita painotetaan eri yksiköissä eri tavalla. Yksi ratkaisu

voisi olla se että makrotasoa analysoivat mittarit painotetaan kaikkien tutkittavien yksikköjen osalta samalla tavalla, mutta mikrotason mittareissa sallitaan jossain määrin poikkeavia painotuksia. Tämä ratkaisu taannee sen, että jokaisen tutkitavan organisaation ominaiset piirteet saadaan paremmin esiin ja oikein analysoiduksi.

LÄHTEET

- Ahonen, P. *Hallinnon arvioinnin lähestymistapoja*. Valtiovarainministeriön järjestelyosasto. Valtion painatuskeskus, Helsinki 1985.
- Ahonen, P. *Farewell to Welfare in Finland? Enter to cutthroat performance evaluation*. Hallinnon tutkimuksen päivät 13–14.11.1991.
- Almond, G. A. & Powell, G. B. (eds.). Introduction. Teoksessa: *Comparative Politics Today: A World View*. 5. painos. Harper Collins Publishers, New York 1992.
- Ammons, D. Overcoming the Inadequacies of Performance Measurement in Local Government. *Public Administration Review* 55(1995): 1, 38–47.
- Argyris, C. *Interpersonal Competence and Organizational Effectiveness*. Richard D. Irwin, Inc., Homewood, Illinois 1962.
- Bonnell, V. E. (1980). The Uses of Theory, Concepts and Comparison in Historical Sociology. *Society for Comparative Study of Society and History* 22 (1980): 2, 156–173.
- Cambell, J. P. On the nature of organizational effectiveness. Teoksessa: J.M. Pennings & P.S. Goodman (eds.), *New Perspectives on Organizational Effectiveness*, 1979, Jossey-Bass, San Francisco, 13–55.
- Chambers, D. E. & Wedel, K.R. & Rodwell, M.K. *Evaluating Social Programs*, Allyn and Bacon, Boston 1992.
- Cohen, D. K. & Spillane, J.P. National Education Indicators and Traditions of Accountability. Teoksessa: *Making Education Count: Developing and Using International Indicators*, 323–337. OECD, Paris 1994.
- Creemers, B. P. M. & Reynolds, D. Issues and Implications of International Effectiveness Research. *International Journal of Educational Research* 25 (1996): 3, 257–265.
- Etzioni-Halevy, E. Comparing Semi-corruption among Parliamentarians in Britain and Australia. Teoksessa: E. Øyen (ed.), *Comparative Methodology: Theory and Practice in International Social Research*, 1990, SAGE Publications Ltd. London, 113–133.
- Hyyryläinen, E. *Vaasan hallintotieteiden vertailuorientaatio*. Occasional Papers 1, Vaasan yliopisto, yhteiskuntatieteellinen tiedekunta, 1995.
- Kanter, R. & Summers, D. Doing well While Doing Good: dilemmas of performance measurement in nonprofit organizations and the need for a multiple-constituency approach. Teoksessa: D. McKeivitt – A. Lawton (eds.), *Public sector management: theory, critique and practice* 1994, Sage, London, 220–236.
- Kogan, M. Comparing Higher Education Systems. *Higher Education* 32 (1996): 4, 395–402.
- Laukkanen, R. Vertailevan tutkimuksen mahdollisuuksia koulutusjärjestelmän arvioinnissa. *Hallinnon tutkimus* 16 (1997): 1, 15–26.
- Lumijärvi, I. *Osittaisarvioinnista kokonaisarviointiin: multidimensionaaliset mallit erityisesti sosiaali- ja terveysalan tuloksikköiden arvioinnin välineinä*. Vaasan yliopisto, 1994.
- Manns, C.L & March, J.G. Financial adversity, internal competition, and curriculum change in a university. *Administrative Science Quarterly*, 23 (1978): December, 541–542.
- Mayer, L.C. *Redefining Comparative Politics: Promise versus Performance*, Sage, Newbury Park, 1989.
- Meyer, M.W & L.G. Zucker. *Permanently Failing Organizations*. Sage Publications, Newbury Park, 1989.
- Przeworski, A. Methods of Cross-National Research, 1970–83: An Overview. Teoksessa: M. Dierkes et al. (eds.), *Comparative Policy Research: Learning from Experience*, 1987, Aldershot: Gower.
- Pennings, J.M. & Goodman, P.S. Toward a workable framework. Teoksessa: J.M. Pennings & P.S. Goodman (eds.), *New Perspectives on Organizational Effectiveness*, 1979, Jossey-Bass, San Francisco, 146–184.
- Ragin, C. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press, Berkeley, 1987.
- Roberts, G.K. The Explanation of Politics: Comparison, Strategy and Theory. Teoksessa: P. G. Lewis & D. C. Potter & F. G. Castles (eds.), *The Practice of Comparative Politics*, 1985, The Open University Press, Avon: Longman, 285–304.
- Salminen, A – Lehtinen, S. *Hallintotiede ja vertaileva tutkimus*. Vaasan korkeakoulun julkaisuja, Tutkimuksia no 86, 1982.
- Salminen, A. *Hallintotiede: Organisaatioiden hallinnolliset perusteet*. Valtion Painatuskeskus, Helsinki, 1998.
- Sartori, G. Comparing and Miscomparing. *Journal of Theoretical Politics* 3 (1991): 3, 243–257.
- Shadish, W. R. & Cook, T.D. & Leviton, L.C. *Foundations of Program Evaluation: Theories of Practice*, Sage, Newbury Park, 1991.
- Stake, R. E. (1991). Program Evaluation, Particularly Responsive Evaluation. Teoksessa: G. T. Madaus & M. Scriven & D. L. Stufflebean (eds.), *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*, 1991, Kluwer-Nijhoff Publishing, Boston, 287–310.
- Vartiainen, P. *Palveluorganisaatioiden tuloksellisuusarviointi, teorit ja käytäntö*. Päivähoitoon, erityisesti päiväkoteihin kohdistuva monitahoarviointi, Vaasan yliopiston julkaisuja, Tutkimuksia No 40, Hallintotiede 2. Vaasa, 1994.
- Vartiainen, P. *Kokemuksia terveydenhuollon laadun arvioinnista SERVQUAL -menetelmällä*. Suomen Lääkärilehti 52 (1997): 16, 1989–1995.
- Vartiainen, P. *Työelämän vaatimuksista opiskelijoiden osaamiseen: Monitahoarvioinnilla kattavaan koulutuksen arviointiin?* Hallinnon tutkimus 1 (1998a): 66–68.
- Vartiainen, P. Local Economic Development Policy in Finland: Key Actors View to Implementation. Teoksessa: K. T. Liou (ed.), *Handbook of Economic Development*, 1998b, Marcel Dekker, Inc. New York, 401–425.
- Weiss, C.H. *Evaluation research: Methods for assessing program effectiveness*. Englewood Cliffs, NJ: Prentice Hall, 1972.
- Vroeijenstijn, A. I. *Improvement and Accountability: Navigating between Scylla and Charybdis, Guide for External Quality Assessment in Higher Education*. Jessica Kingsley Publishers, London, 1995.
- Vuori, J. *Kenen terveydenhuolto? Julkinen ja yksityisen vertailussa*. Ankkurikustannus, Helsinki, 1995.