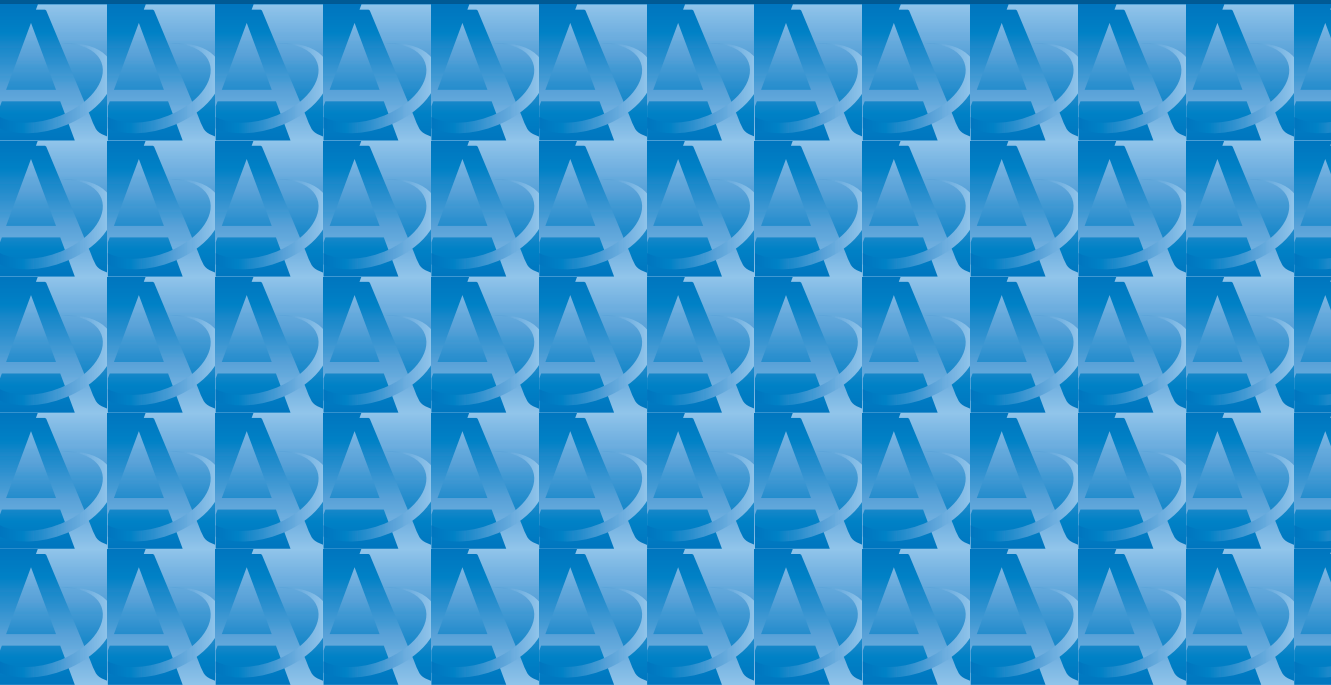


Esko Korkeakoski &
Heikki Silvennoinen
(toim.)



Avaimia koulutuksen arvioinnin kehittämiseen



Avaimia koulutuksen arvioinnin kehittämiseen

Koulutuksen arviointineuvoston
julkaisuja 31

Avaimia koulutuksen arvioinnin kehittämiseen

Esko Korkeakoski & Heikki Silvennoinen (toim.)

JULKAISUN MYYNTI:
Koulutuksen arviointisihteeristö
PL 35, 40014 Jyväskylän yliopisto
puh. (014) 260 3220
faksi (014) 260 3241
ktl-asiakaspalvelu@ktl.jyu.fi
www.edev.fi

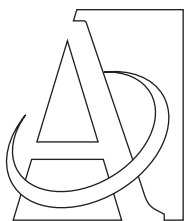
Julkaisija: Koulutuksen arviointineuvosto, Jyväskylä 2008

© Koulutuksen arviointineuvosto ja kirjoittajat

Kansi ja ulkoasu: Martti Minkkinen
Taitto: Kaija Mannström

ISSN 1795-0155 (painettu)
ISSN 1795-0163 (verkkojulkaisu, pdf)

ISBN 978-951-39-3207-7 (painettu)
ISBN 978-951-39-5347-8 (verkkojulkaisu, pdf)



Sisältö

Esipuhe.....	3
Johdanto	5
<i>Esko Korkeakoski & Heikki Silvennoinen</i>	
I Arvot ja toimintaperiaatteet arvioinnissa	
Arvotietoisuus arvioinnin perustana	11
<i>Timo Laine & Anita Malinen</i>	
Arvottaminen on arvioinnin ydintä.....	25
<i>Heikki Kalevi Lyytinen & Pentti Nikkanen</i>	
Arvioinnin eettisyys	47
<i>Pirjo Linnakylä & Päivi Atjonen</i>	
Riippumattomuus arviointitoiminnan periaatteena	65
<i>Esko Korkeakoski</i>	
Arviointi, tutkimus ja arviointitutkimus koulutuksen tietotuotannossa	79
<i>Pirjo Linnakylä & Päivi Atjonen</i>	
II Arviointitoiminnan organisoinnin tapoja ja menetelmiä	
Verkostomaisuus koulutuksen arvioinnin organisoititapana.....	99
<i>Heikki Silvennoinen</i>	
Kehittävä arviointi	117
<i>Jyri Manninen & Antti Kauppi</i>	
Itsearviointi kehittävän arvioinnin menetelmänä.....	127
<i>Mira Huusko</i>	
Utvärdering i utvecklande syfte på nationell nivå.....	139
<i>Gunnel Knubb-Manninen</i>	

III Arvioinnin kohteet menetelmällisten valintojen lähtökohtina

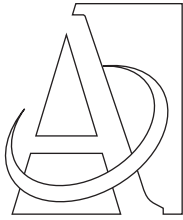
Koulutuspoliittinen arviointi	153
<i>Reijo Raivola</i>	
Koulutuksen talouden arvioinnin menetelmiä.....	169
<i>Tanja Kirjavainen</i>	
Ammatillisen osaamisen arviointi	185
<i>Marja-Leena Stenström</i>	

IV Kokoavaa tarkastelua

Arvotietoisuus, teorialähtöisyys ja vaikuttavuus arviointimenetelmien kehittämisessä.....	201
<i>Esko Korkeakoski</i>	

V Koulutuksen arvioinnin peruskäsitteitä

Koulutuksen arvioinnin peruskäsitteitä	219
Kirjoittajat.....	229



Esipuhe

Arviointi herättää aina mielenkiintoa monista eri syistä. Se on vallankäyttöä. Se on eräs laadunhallinnan keino. Arviointi ulottuu kaikkeen inhimilliseen toimintaan.

Arviointia ei liitetä kuitenkaan minkään yksittäisen tieteenalan piiriin. Arvioinnissa ei käytetä johdonmukaisesti esimerkiksi käyttäytymistieteissä vakiintunutta terminologiaa. Arvioinnin piirissä kehittynyt erityissanasto on vielä kirjavaa ja vakiintumatonta.

Arvioinnin kehittäminen on jäänyt sekä kansallisesti että kansainvälisesti osin hallinnon piirissä tehtäväksi työksi ja osin arviointia tekevien laitosten ja yksiköiden toiminnaksi. Tästä on osin seurannut, etteivät arviointi ja sen metodiikka ole olleet systemaattisen tutkimuksen tai kehittämisen kohteina. Siksi tarvitaan arvioinnille ominaisten menetelmien tutkimusta ja kehittämistä. Kun otetaan lisäksi huomioon se, että arviointitoiminnan volyymi on vajaan parin viime vuosikymmenen aikana kasvanut merkittävästi eri aloilla, on tarve arviointimetodiseen kehittämiseen entistä ajankohtaisempaa.

Arvioinnin yleistyminen hallinnossa, sen valtapoliittinen merkitys ja laadunhallinnan keinojen lisääntyminen ovat luoneet kasvualustaa myös konsulttiterminologialle, joka ei lähtökohtaisesti pyrikään täsmälliseen käsitteistöön. Tässä ympäristössä on saattanut syntyä myös kasvupohjaa ajattelulle, jossa arviointi rinnastetaan tavarantuotannon prosesseihin.

Kun puhutaan koulutuksen arvioinnin menetelmistä, niillä tarkoitetaan usein samaa kuin käyttäytymis- ja yhteiskuntatieteiden menetelmillä. Arviointitoiminnassa puhutaan menetelmistä myös laajemmassa merkityksessä. Kyse voi olla toimintatavoista, joilla arvioinnit organisoidaan tai toteutetaan, mutta myös menetelmistä, joilla arviointien aineistot hankitaan ja käsitellään. Tulosten arvottamisen menetelmät ovat kuitenkin arvioinnin ydintä. Kun tähän lisätään vielä keinot, joilla arviointitietoa levitetään päättäjille ja kehittäjille, kyse on toiminnallisesti laajasta valintojen ketjusta, millä kaikella on selvä merkitys arviointitoiminnan vaikuttavuuden kannalta.

Tässä julkaisussa tarkastellaan arvoja, arvottamista, arvioinnin metodologiaa ja menetelmiä käytännön ja teorian näkökulmista sekä eri konteksteissa. Julkaisussa käsitellään myös arvioinnin perusterminologiaa tavanomaista laajemmin ja monipuolisemmin.

Kirjoittajat ovat arvioinnin, arviointitutkimuksen ja tutkimuksen kokeneita asiantuntijoita. Kiitämme heitä sekä julkaisun valmistumista muutoin tavalla tai toisella tukeneita asiantuntijoita.

Jyväskylässä 27. helmikuuta 2008

Esko Korkeakoski

Heikki Silvennoinen



Johdanto

Koulutuksen arvioinnin menetelmällisen kehittämisen lähtökohtana voidaan pitää koulutuksen arvioinnille laissa määriteltyä tarkoitusta (32/2003), asetuksessa määriteltyjä tavoitteita (150/2003) sekä arvioinnin erityistä luonnetta koulutustiedon tuottamisen kokonaisuudessa. Menetelmällisten ratkaisujen kannalta tärkeitä arviointipoliittisia linjauksia ja toimintaperiaatteita ovat arvioinnin riippumattomuus, verkostomaisuus, monitahoisuus, osallistavuus ja kehittävyys.

Koululainsäädännössä määritellään arvioinnin tarkoitukseksi 1) turvata koululakien tarkoituksen toteuttaminen, 2) koulutuksen kehittämisen tukeminen ja 3) oppimisen edellytysten parantaminen (esim. PL/LL 2003). Arvioinnista annetun asetuksen (150/2003) mukaan koulutuksen arvioinnin tavoitteena on: 1) hankkia ja analysoida tietoa valtakunnallisen koulutuspoliittisen päätöksenteon ja koulutuksen kehittämisen pohjaksi sekä koulutuksen paikallisen kehittämistyön ja päätöksenteon pohjaksi ja 2) tukea opiskelijoiden oppimista, opetustoimen henkilöstön työtä ja oppilaitosten kehittämistä. Arvioinnin tarkoitus ja tavoitteet lähtevät koulutuksen käytännöistä ja päätyvät niiden parantamiseen.

Menetelmällisesti arvioinnin tarkoitusta tukevat myös arvioinnin toimintaperiaatteet. Näitä ovat esimerkiksi kattavuus, jatkuvuus, monipuolisuus, avoimuus, oikeudenmukaisuus sekä arviointitiedon luotettavuus ja vertailukelpoisuus. (Asetus 150/2003.)

Arvioinnin kohteina voivat olla muun muassa koulutuksen resurssit, prosessit ja tulokset sekä niiden keskinäiset yhteydet. Arvioinnin kohteet voivat ulottua koulutuspolitiikan onnistumisen ja koko koulujärjestelmän arvioinnista oppilaan tai opiskelijan oppimisen arviointiin saakka. Arviointien kohteet ovat hyvin

erilaisia, ja ne voivat olla hyvinkin hankalia arvioitaviksi. Tästä syystä ne ovat haasteellisia myös menetelmällisesti.

Käsillä olevan julkaisun tausta-ajatus on ollut selkiyttää arviointimenetelmien ”viidakkoa” ja arviointimenetelmien lähtökohtia sekä antaa aineksia arvioinnin metodologiasta ja menetelmistä käytäviin keskusteluihin ja niiden kehittämiseen. Artikkeleissa käsitellään arviointia ja sen menetelmiä myös kriittisesti. Samaan asiaan saattaa olla useita näkökulmia, eivätkä ne ole aina toisiaan tukevia. Kriittisyys luo osaltaan edellytyksiä arvioinnin menetelmällisen perustan kehittämiseksi. Toisaalta artikkeleissa pyritään ottamaan huomioon suomalaisen arviointipolitiikan painotukset: verkostomainen organisoituminen, monitahoisen asiantuntemuksen hyödyntäminen ja kehittävän arvioinnin toimintatapojen ja menetelmien vahvistaminen.

Julkaisun tavoitteena on syventää arviointimetodisen työn perusteita: arvoja arvottamisen lähtökohtana, arvottamisen keinoja sekä arvioinnin eettisiä periaatteita ja riippumattomuutta. Menetelmällisiä erityishaasteita tuovat arvioinnin organisoiminen verkostomaisesti, monitahoarviointeina, kehittävän arvioinnin menetelmin ja itsearviointeina. Julkaisussa pohditaan haasteita, joita mainitut arvioinnin organisointimuodot tuovat tullessaan, mutta tarjotaan myös ratkaisuja, joilla haasteisiin voidaan vastata. Julkaisun tavoitteena on ollut myös tarkastella menetelmällisestä näkökulmasta niitä arvioinnin kohteita, jotka eriyistä on koulutuksen piirissä koettu vaikeiksi tai poikkeuksellisiksi lähestyä arvioinnin keinoin ja toimien arvioinnin tavoitteiden mukaan. Näitä kohteita ovat esimerkiksi koulutuspoliittinen arviointi sekä koulutuksen taloudellisuuden ja tuottavuuden arviointi. Julkaisussa on myös arvioinnin peruskäsitteistöä koskeva laajempi tarkastelu. Julkaisussa ei asetuta tukemaan yhdenkään itsearviointityökalun tai laadunhallinnan työkalun ideologiaa, mutta sivutaan kyllä niiden vahvuuksia ja puutteita erilaisissa toimintaympäristöissä.

Artikkelit antavat tietoa myös seuraavista pääkysymyksistä:

- 1) Miten arvioinnin menetelmällisillä ratkaisuilla voidaan tukea kansallisen arvioinnin tarkoitusta, tavoitteita ja toimintaperiaatteita?
- 2) Miten arvioinnin prosessit ja menetelmät luovat edellytyksiä tulosten arvottamiselle ja kehittämistarpeiden esiin nostamiselle?
- 3) Miten arvioinnin menetelmällisten ratkaisujen avulla voidaan edistää yhteiskunnallisesti merkittävän arviointitiedon tuottamista?

Julkaisu tukee myös arvioinnin kehittämistyötä, jossa arvioinnin tarkoitus, arvottaminen arvioinnin pääfunktiona sekä arviointitoiminnan totuudellisuus ovat lähtökohtina. Tämä on osa arviointitoiminnan laadun parantamista.

Arvioinnin metodista kehittämistä on luontevaa jatkaa erityyppisten arviointien ja erilaisten kohteiden arvioinnissa käytettyjen menetelmien analyysillä.

Niitä tiedonhankinnan ja aineistojen analyysien menetelmiä, jotka soveltuvat erityisesti arviointitoiminnan käyttöön, on tarpeen tarkastella omana julkaisunaan. Tämän ohella arvioinnin käytäntöä palveleva funktio luo tarpeen menetelmälliselle kehittämiselle, jonka avulla entistä perustellummin voidaan sanoa, mitä tiedetyn perusteella on oikeutettua päätellä kehittämisen tarpeista.

Olemme suunnitelleet kirjan erityisesti arvioinnin paikallisten, alueellisten ja kansallisten arviointiasiantuntijoiden tarpeita varten. Toivomme kirjasta olevan hyötyä myös arviointitietoa hyödyntäville kehittäjille, kouluttajille ja päätöksentekijöille sekä muille arviointitoiminnan metodologisista ja metodisista seikoista kiinnostuneille.

I

Arvot ja toimintaperiaatteet arvioinnissa



Arvotietoisuus arvioinnin perustana

Koulutus on arvovalintoja

Perustavin kysymys arvioinnin kannalta lienee, mitä varten jokin koulutus on olemassa ja mitä varten on sen arviointi. Kaikki käytännölliset ratkaisut aina ylimmästä hallinnosta yksittäisen opiskelijan oppimisprosessiin saakka seuraavat, tai johdonmukaisuuden nimissä niiden kuuluisi seurata, vastauksesta tähän kysymykseen. ”Mitä varten?” on ihmisten toiminnan maailmassa aina arvokysymys. Linnut eivät rakenna pesiä arvovalinnan pohjalta, mutta kun ihmiset rakentavat koulutusorganisaationsa, he punnitsevat (toivottavasti), pitäisikö organisaation olla juuri tällainen vai kenties toisenlainen. Perustan pohdinnalle luo kysymys ”Mitä varten?”. Ensinnäkin on siis tarve määritellä ja valita koulutuksen tarkoitus ja tavoitteet.

Käytännössä tämä ei kuitenkaan etene kronologisesti näin, vaan jossain on jo aiemmin syntynyt *intressi* juuri tietynlaiseen koulutukseen. Koulutusorganisaatiot ovat yhteiskunnallisia instituutioita, ja intressit koulutuksen järjestämiseen syntyvät yhteiskunnan toiminnasta. Koulutuspoliittisen vallan käyttäjät ovat intressien esille nostajina keskeisessä asemassa. Heidän toimintansa on arvovalintojen tekemistä: sen suunnittelua, mitä *haluttaisiin* olevan, mitä *pitäisi* heidän mielestään olla, millaista koulutusta Suomessa nyt *tarvittaisiin*. Yhteiskunnan kehityksen ohjaaminen on tulevaisuustyötä, siis toimimista jonkin sellaisen parissa, mitä ei vielä ole. Ohjaaminen, suunnan valitseminen, on arvovalinta. Muutkin tiet ja mahdollisuudet ovat periaatteessa aina samanaikaisesti avoimina, mutta jokin tietty suunta valitaan. On tärkeää nähdä, että toiminnan perus-

tana eivät ole vain ”kylmät tosiasiat”, joista poliitikot usein puhuvat. Tosiasiat ovat jo, niiden suhteen ei tarvitse tehdä valintoja. Usein yksittäiseltä toimijalta saattaa jäädä huomaamatta ja ymmärtämättä, että jossain on jo tehty valinta, ja hän löytää itsensä uimasta virran mukana. ”Kylmien tosiasioiden” maailma on tosiasiallisesti aina arvojen kuumentama.

Arviointi liittyy kaikkeen inhimilliseen toimintaan, mutta minkä erityispiirteen arviointiin tuo se, että kyse on kasvatuksen ja koulutuksen kentästä? Kasvatuksen ja koulutuksen yhteiskunnallinen tehtävä lienee kärjistäen ja tiivistäen (ja totutusti) kahdenlainen: uusintava ja/tai uudistava. Nämä näkökulmat painottuvat eri tavoin riippuen siitä, millä koulutusjärjestelmän tasolla kulloinkin ollaan. Esiopetuksen ja peruskoulun perustehtävä ja tavoite on toisenlainen kuin esimerkiksi ammatillisen aikuiskoulutuksen tai vapaan sivistystyön koulutuksen. Ovatko nämä mainitut kaksi näkökulmaa ne, jotka viime kädessä tietoisesti tai tiedostamattomasti ovat arvioinnin kriteereinä?

Miksi esimerkiksi työssä oppimiseen käytetään nykyään niin paljon taloudellisia ja tutkimuksellisia resursseja? Miksi vapaan sivistystyön koulutuksen vaikuttavuutta on selvitetty niin tarkoin viime aikoina? Miksi peruskoulujen opetussuunnitelmiin tulee jatkuvasti lisää uusia sisältöjä? Miksi kaksoistutkintoja markkinoidaan nuorille? Miksi pitkään työssä olleille aikuisille pitää saada ammatillinen perustutkinto? Liittyvätkö nämä koulutuspoliittiset trendit joko uusintavaan tai uudistavaan näkökulmaan tai mahdollisesti samanaikaisesti molempiin? Nämä kysymykset ovat vielä aukaisematta arvojen ketjuina, mutta julkilausumattomana tietona kenties jo olemassa. Jos eri tahoilla rakentuneiden arvojen ketjua ei nähdä kokonaisuutena, on vaikea nähdä arvioinnin kriteerejä ja on vaikea tehdä kehittämisehdotuksia.

Maailma, sellaisena kuin se kulloinkin on, on perusta kaikenlaisten intressien synnylle. Se toimii todellisuuspohjana uuden suunnittelussa: uutta voidaan rakentaa vain olemassa olevan perustalle. Tulevaisuus ei kuitenkaan ole vielä olemassa, tulevaisuutta ollaan vasta haluamassa. *Intressit ja haluaminen kuvaavat arvojen luonnetta*: ne ovat asioita, joita joku haluaa ja joihin on syntynyt jokin intressi. Niitä lähdetään tavoittelemaan, ja toimintaa ohjataan niiden suuntaan – tai toiminta ohjautuu ”luonnostaan”, silloin kun valinnat tehdään tiedostamatta. Arvoja toisin sanoen arvostetaan. Ne ovat jonkin lajin hyvää. Ne ovat valitsijoille merkityksellisiä, tärkeyksiä.

Elämme maailmassa, jossa ei tyydytä siihen, mitä on. Siksi kehitämme kaikkea. Kehittämisen ilmenemismuotoina meillä on tavattoman paljon erilaista koulutusta ja sen arviointia. Tähän yhteiskuntaan on haluttu ja halutaan ihmisiä, jotka osaavat kulloisessakin koulutuksessa tärkeänä pidettyä asiaa. Maailman haluttaisiin muuttuvan näiden osaamisen suuntaisesti. Tämän lisäksi halutaan, että osaajat osaisivat asiansa *hyvin*. Siis sekin, *miten* koulutus toteutetaan, on arvoasia.

Käytämme tässä artikkelissa esimerkkinä Aikuiskouluttajan pedagogisia opintoja (APO), koulutusta, jossa itse työskentelemme. Alun perin tämä koulutus syntyi tarpeesta saada vapaan sivistystyön kentälle päteviä opettajia, mutta hyvin pian tavoitteet laajenivat koskemaan koko aikuiskoulutuksen kenttää. Nykyinen tavoite APO-koulutuksessa on saattaa yhteiskuntaan aikuisten kouluttajia – opettajia, ohjaajia –, koska aikuiskoulutusta on alettu pitää yhteiskunnassamme yhä tärkeämpänä. On olemassa jatkuvasti lisääntyviä intressejä aikuiskoulutuksen olemassaololle. Tällöin aikuisten kouluttaja on tärkeä asia, arvo. Lisäksi asetamme koulutuksellemme ikään kuin tavoitteelliseksi lisäarvoksi, että opiskelijoihastamme tulisi *hyviä* kouluttajia. Tämän arvotavoitteen kautta on rakentunut koulutuksen käytännön toteutuksen suunnitelma. Millainen on hyvä aikuisten kouluttaja? Tämä on perustava kysymys suunnittelulle: millaisilla käytännöllisillä pedagogisilla ratkaisuilla pääsisimme tavoitteeseen? Käytännön valintojen taustalla on kaikkein pienintäkin yksityiskohtaa myöten kuva Hyvästä aikuisten kouluttajasta.

Yksi perusongelma arvojen kohdalla on, että emme välttämättä ole tietoisia niistä. Kun toimimme tottumuksemme varassa itsestään selvässä maailmassa, emme ajattele. Samoin kuin ihmisyksilöt eivät useinkaan tunnista toiminnastaan, asenteistaan ja reaktioistaan omia arvojaan, samoin eivät instituutiot ole perusluonteeltaan reflektiivisiä, itseään tietoisesti arvioivia. Jopa kysymys jonkin organisaation tarkoituksesta saattaa unohtua, kun organisaatio elää ja toimii kauan, ja organisaatioon on aikojen mittaan totuttu. Tällaisena epävarmuuden aikakautena muutos on hyvin nopeaa, ja myös vanhat organisaatiot ja instituutiot joutuvat vastaamaan kysymyksen, onko niiden tarkoitus edelleen sama kuin se oli aiemmin niiden syntyessä toisenlaiseen aikakauteen. Hyvänä esimerkkinä tästä on vapaan sivistystyön koulutus, joka syntyi jo sata vuotta sitten silloisen yhteiskunnallisen tilanteen tarpeisiin.

Koulutuksen arvot voivat jäädä tiedostamattomiksi usein siitä syystä, että arvovalinnat ja koulutuksen käytännön toteutus tapahtuvat eri tahoilla, eri ajassa, eri paikassa ja eri henkilöiden toimesta. Yksittäinen opettaja ei välttämättä hahmota, mille perustalle on rakentunut se toimintaympäristö, johon hän tulee tänään töihin. Koulutuksen arvoperusta on saattanut muotoutua historiallisesti hyvin erilaisista arvolähtökohdista, eikä sillä siten ole välttämättä lainkaan yhtenäistä ja johdonmukaista linjaa. Onko esimerkkinämme toimiva APO-koulutus rakentunut johdonmukaisesti tarkoituksensa mukaisen arvolinjan suuntaisesti? Onko meillä selkeä kuva siitä, mitä tarkoittaa Hyvä aikuisten kouluttaja? Ovatko kaikki koulutuksen toimintatavat todella valittu tuon päätavoitteen suuntaisesti – ovatko arvot siis linjassa, palvelevatko kaikki välinearvot päätavoitetta? Tästä aiheesta käymme koulutuksen sisäistä jatkuvaa arvokeskustelua.

Arvioinnissakin on kyse arvovalinnoista

Miksi koulutuksia arvioidaan? Miksi ne arvioivat itse itseään, ja miksi on olemassa jo kokonainen ”arviointiosajien ammattikunta” (ks. Uusikylä 1999, 17)? Koululaki sanoo, että koulutuksen arvioinnin tarkoituksena on 1) turvata koululakien toteutuminen, 2) tukea koulutuksen kehittämistä ja 3) parantaa oppimisen edellytyksiä. Nuo lauseet näyttävät itsessään varsin neutraaleilta, mutta todellisuudessa niiden sisällä on arvoasettamuksia, jotka ovat syntyneet oman aikansa yhteiskunnallisista intresseistä. ”Toteutumisen turvaaminen”, ”kehittäminen” ja ”edellytysten parantaminen” ovat arvoihin viittaavia käsitteitä. Niissä on kyse jostain hyvästä, jota pidetään arvokkaana ja haluttavana. Näillä intresseillä on aina suunta: ei kehitetä tai paranneta miten milloinkin, vaan halutaan tietynsuuntaista kehitystä.

Arvioinnin tarkoituksena on koulutuksen ”laadun kehittäminen”. Saman tien joutuu kysymään, miksi koulutuksen laatua on kehitettävä. Onko koulutuksesta ja sen kehittämisestä tullut peräti itseisarvo? Kenen intressi ”kehittäminen” on? Yhteiskunnan tasolla kysymykseen voidaan vastata, että koulutus auttaa osaltaan kansainvälisen kilpailukyvyyn kehittymisessä. Alueellisella tasolla voidaan vastata, että koulutus palvelee alueen elinkeinoelämää, muun muassa työvoiman laatua ja saatavuutta. Koulutuksen kehittäminen voi olla koulutusorganisaation johdon intressi, jotta oma oppilaitos säilyttää asemansa koulutusorganisaatioiden kokonaisuudessa. Tähän taas kytkeytyy se, että oppilaitos näyttää haluttavalta ja kiinnostavalta potentiaalisten opiskelijoiden näkökulmasta. Koulutuksen kehittäminen voi olla myös yksittäisen opettajan tai kouluttajan intressi. Miksi? Esimerkiksi siksi, että opettajan ihmiskäsitykseen kuuluu, että hän haluaa opiskelijoidensa kasvavan ja kehittyvän tiettyyn suuntaan. Opettajan intressi voi liittyä myös oman työpaikan säilyttämiseen tai omaan ammatilliseen kehittymiseen ja kunnianhimpoon. Miten yhteiskunnalliset ja poliittiset koulutuksen kehittämisen intressit suhteutuvat yksittäisten oppilaitosten ja opettajien kehittämistyöhön? Kohtaavatko keskushallinnon, oppilaitostason ja työntekijätason intressit? Ovatko ne keskenään sopusoinnussa? Entä mikä on oppijan intressi?

Kuten koulutuksenkin kohdalla, niin myös koulutuksen arvioinnissa on kyse yhteiskuntapoliittisista arvoista: mihin suuntaan vallanpitäjät haluavat viedä suomalaista yhteiskuntaa? Mikä on kaikista tärkeintä, mikä siinä ohella hyvin tärkeää ja mikä vähemmän tärkeää, mutta suotavaa? Tällaista perusteellista arvoanalyysiä ei kaikeksi ole tehty nykyisen koulutuspolitiikan kohdalla. Miten etenee maailmanlaajuinen kehitys tiettyjen suurarvojen perässä, millaista on kehitys Euroopan unionissa, ja miten Euroopan unionin sisällä suuntaudutaan Suomessa? Onko todella niin kuin näyttäisi, että taloudelliset arvot ovat ehdottomasti johtavassa asemassa, mihin viittaa esimerkiksi puhe kansainvälisessä kilpailukyvyssä mukana pysymisestä. Koulutuksen olemassaoloa ja ”hyödylli-

syöttä” perustellaan usein nimenomaan kansainvälisellä kilpailukyvyllä. Jos tämä on yhteiskunnan perustavanlaatuisin intressi, niin miten se näkyy kaikissa alemman tason kehittämishankkeissa, joita koulutuksen maailma on tulvillaan? Miten se näkyy siinä, mitä ja miten arvioidaan?

Kysymys arvoista voidaan esittää myös kriittisesti: jos suunta on tämä, niin onko se hyvä? Arviointitutkimus voi pyrkiä olemaan tieteellistä senhetkisen toimintajärjestelmän kuvausta. Tutkimuksessa voidaan yrittää selvittää, mitä nyt on, mitä tapahtuu. Arvioinnin ja arvioitsijoiden kannalta on kuitenkin ratkaisevan tärkeää, ovatko he vain jo suunnatun prosessin toteajia, selvittäjiä, tiedonhankkijoita ja analyysin tekijöitä. Vai voidaanko myös itse kokonaisprosessin arvoja arvioida? Tällöin astutaan ulos senhetkisestä arvohakuisen toiminnan virrasta ja asetetaan aidosti kriittinen kysymys eli kyseenalaistetaan vallitsevat koulutusta ohjaavat arvot.

Koulutuksen, samoin kuin vaikkapa terveydenhuollon, arvioinnissa käytetään nykyisin tuottavuuden ja tehokkuuden termejä: kuinka monta maisteria valmistuu vuodessa, kuinka monta leikkausta tehdään vuorokaudessa? Ei ole vaikea nähdä, että sama kieli, samat metaforat ovat tulleet sisään puheeseen, jossa niitä aiemmin olisi pidetty outoina. Terveyden tai koulutuksen ymmärtäminen analogiseksi tavaratuotannon kanssa ei ole kovin vanha ilmiö. House kirjoitti 1980-luvun alkupuolella, miten USA:ssa tuotanto-metafora on tullut keskeiseksi arvioinnin kielessä (House 1983, 8–24). Koulutusta voidaan arvioida tehokkaan tuotannon näkökulmasta käytetyn ajan, kustannusten ja toisesta päästä tipahtelien tuotteiden määrän suhteina. Tällainen arviointi edellyttää, että arvioitavat kohteet määritetään mitattaviksi objekteiksi, esinemäisiksi tuotoksiksi. Millaisia mielikuvia tämän metaforan sisällä syntyy, kun puhutaan koulutuksen tuloksellisuudesta tai vaikuttavuudesta? House korostaa sitä, että kieli, jolla asiasta puhutaan, määrittää sitä, miten asia ymmärretään. Tässä tapauksessa toimii ymmärryksen isona kehyksenä tuotanto-metafora. Hän korostaa, että tällaiset metaforat eivät suinkaan ole arvoneutraaleja (emt., 20). Charles Taylor on puhunut samassa merkityksessä narratiiveista, ”joissa ihmiset elävät elämänsä” (Taylor 1989, 58). Tässä yhteydessä hän tarkoittaa kokonaistarinaa, kehystä, jonka sisällä koulutuksen mielletään toimivan.

Arvioinnin menetelmiä ajatellen on keskeistä hahmottaa arvioitavan asian luonne. Filosofian termein sanottuna: tutkittavan kohteen ontologinen määrittäminen määrittää tutkimusmetodien valintaa. Jos koulutuksen tulokset mielletään tuotoksiksi, se synnyttää esinemäisen mielikuvan ja samalla mitattavuuden idean. Huomio kiinnittyy erityisesti mitattaviin ilmiöihin: gradujen ja väitöskirjojen määrään, valmistuneiden ja keskeyttäneiden määrään, valmistumiseen kuluneeseen aikaan, sen nielemisiin rahasummiin jne. Sen sijaan oppiminen, oppineisuus on laadullinen ilmiö, jonka arvioiminen vaatii toisenlaisia lähestymistapoja.

Autonomiseen arviointiin?

Kysymys *arvioinnin autonomisuudesta* on kaikilla arvioinnin tasoilla arvovalintojen kannalta keskeinen. Juuri tätä kysymystä ei useinkaan esitetä, vaan arvot ja valinnat otetaan annettuina, vaikkapa vain toteamalla, että tätä rataa yhteiskunnan juna nyt vain kulkee, tällainen näyttää olevan maailman kehityssuunta. Jos arvioija esittää kriittisen kysymyksen, ovatko nämä arvot päteviä, joiden suunnassa koulutusta halutaan kehittää, hän esittää autonomisen kysymyksen. Eli hän ei arvioi vain vallitsevan suuntauksen kehyksissä, vaan astuu askeleen tai pari pois päin, ottaa etäisyyttä ja katsoo kehityssuuntaa sieltä. Ei-autonominen arviointi voi myös olla kriittistä, mutta vain hallitsevan suunnan sisäisissä kysymyksissä (esimerkiksi ”Mikä jarruttaa gradujen tuotannon tehostamista?”), ei suunnan itsensä suhteen. Tällä hetkellä kenties epäautonomisinta eli vallitsevan arvomaailman virrassa uimista edustaa arviointi, jonka ainoana tarkoituksena on arvioitavan imagon parantaminen.

Autonomisuus tarkoittaisi arvioijan riippumattomuutta häntä korkeampien tahojen, joilla on valta määrittää suuret arvolinjaukset, koulutuspolitiikasta. Tämä koskee yhtä lailla arviointineuvostojen, yksittäisten instituutioiden ja koulutusten, kuin myös yksittäisen opettajan arviointitoimintaa. Mihin esimerkiksi yliopistot ovat menossa? Millaisia ovat ne arvot, joita toteutetaan ja tavoitellaan? Sen sijaan, että arviointi olisi vain jo meneillään olevan prosessin toteamista, sen pitäisi kyetä tekemään arvoanalyysi vallitsevasta tilanteesta ja sen kehityssuunnasta, jotta arviointi olisi autonomista.

Arvosuunnat, joissa arvioija ja arvioitava ovat sisällä, täytyy tiedostaa, jotta voisi nähdä ja toimia toisin, vapaampana niistä. Kutsumme tätä *arvotietoisuudeksi*. Tavallisesti emme ole kovinkaan tietoisia meitä ohjaavista arvoista, vaan toimimme enimmäkseen tavan ja totumuksen mukaan: sen mukaan, mikä on normaalia, miten asiat näyttävät enimmäkseen menevän. Me totumme nopeasti uusiin arvosuuntauksiin ja alamme pitää niitä luonnollisina. Yksilöillä, samoin kuin instituutioillakin, on taipumus mieltää vallitseva kehityssuunta ”normaaliksi” tai ”luonnolliseksi”, jolloin muita mahdollisia suuntia ei enää pidetä realistisina. Tilanteita ei ymmärretä useinkaan valintatilanteiksi, koska muita valittavia mahdollisuuksia ei nähdä.

APO-koulutuksessa olemme pohtineet hyvän aikuiskouluttajan määrittelyä eri suunnista. Meillä yksi ulkoapäin voimakkaasti asetettu arvo vaatimus on ”työelämälähtöisyys”, kuten kaikilla muillakin koulutuksen aloilla. Ammattiin valmistavan koulutuksen tarkoitus on toki palvella työelämää, mutta millaiseen työelämää me opettajiamme kouluttaisimme? Siihenkö, mikä on nyt, vai siihenkö, minkä oletetaan olevan lähitulevaisuudessa, jos työn maailma etenee tässä annetussa suunnassa? Onko muita mahdollisuuksia?

Nykyiseen suuntaan sopeuttavan koulutuksen vastakohta voisi olla koulutus, jossa kasvaisi mahdollisimman autonomisia, arvotietoisia ja kriittisiä opettajia, joiden päätavoite olisikin työelämän muuttaminen paremmaksi riippumatta siitä, mikä on ympäröivän arvomaailman suunta. He voisivat arvioida koulutusta ja yksittäisten oppijoiden oppimista arvokehyksissä, joissa keskeistä olisi vaikkapa ihmisen kokonaisvaltainen, mahdollisimman moniulotteinen kasvu, yhteisöllisyys ja ekologisuus. Tällainen arvojen asettaminen olisi siis mahdollista, mutta ei välttämättä helposti toteutettavissa. Ne eivät ole niitä arvoja, jotka parhaiten palvelevat nykyisenkaltaista taloudellista kasvua, kuten sen sijaan tekevät tehokas suorittaminen, kilpailu ja specialisoituminen. Autonomisuus edellyttää tässäkin arvotietoisuutta: millaisissa arvokehyksissä opettaja elää, millaisiin arvoihin häntä yhteiskunnassa yritetään sopeuttaa? Toisaalta, mitä hän itse pitäisi hyvänä opettamisena, millaisiksi ihmisiksi hän haluaisi oppilaittensa kasvavan?

Nuo edellä esitetyt esimerkit ovat tarkoituksellisesti yksiulotteisia. Todellisuudessa ei vallitse vain yksi arvosuuntaus, joka hallitsisi kaikkea elämää, vaikka taloudellisilla arvoilla onkin nyt ylivalta lähes kaikilla elämän aloilla. Koulutuksenkin arvomaailma on moninainen, kun menemme lähemmäksi yksittäisiä koulutuksia. Vielä moninaisemmiksi koulutukset muuttuvat yksittäisten kouluttajien kohdalla. Eivätkä arvot suinkaan ole aina yhdensuuntaisia edes saman oppilaitoksen sisällä; koulutusmaailmassa saattaa olla samanaikaisesti hyvinkin ristiriitaisia arvosuuntauksia.

Miksi konstruktivismiksi nimetty oppimisteoriarypäs on tämän aikakauden suosikki? Teoreettisen ja tieteellisen pätevyytensäkö vuoksi? Eivätkö kaikki aikaisemmatkin hallitsevat teoriat ole olleet päteviä? Monille opettajille konstruktivismi merkitsee arvona ihmisläheistä, sekä yksilöä että yhteisöllisyyttä kunnioittavaa toimintatapaa. Opettajat eivät ehkä näkisi konstruktivismin kytkeytyvän taloudellisen kasvun ja globaalin talouden kehityksen arvomaailmaan – tehokkuuteen, tuottavuuteen, kilpailuun. Onko niillä kuitenkin jokin yhteinen arvo-perusta, koska ne ovat saman yhteiskunnallisen aikakauden kaksi yhtäaikaista toimintalinjaa? Arvotietoisuuden kasvattaminen tarkoittaa tällaisten kysymysten selvittämistä. Näin autetaan arvioijaa, yhtä lailla kuin koulutuksen toteuttajia, paikantamaan itsensä ja näkemään, missä ja mihin he ovat menossa ja miksi.

Yksi arvioinnin ja arvottamisen filosofinen ulottuvuus on *ihmiskuvan ja arvojen välinen kytkentä*. Millaisia ihmisiä haluaisimme koulutuksissamme kasvavan, eli naiivimmalta kuulostaen: millainen olisi hyvä ihminen tässä kyseisen koulutuksen kontekstissa? Nämä ovat vanhoja kasvatusfilosofisia kysymyksiä, jotka eivät nykyisin ole erityisen muodikkaita, eivät siis ehkä ”nykyaikaisia” kysymyksiä. Kenties ei kannata pohtia kysymystä inhimillisen kasvun moniulotteisuudesta, ellei sillä ole jotain tekemistä taloudellisen tuotannon kasvun kanssa. Taideopetuksen puoltajien kuuleekin toisinaan perustelevan taideopetuksen tarpeellisuutta sillä, että luovuus lisää tuottavuutta ja kilpailukykyä.

Koulutus ja sen arviointi tapahtuu aina jonkin ihmiskuvan, samoin kuin hallitsevien arvojen, kehyksissä, usein tiedostamatta. Nämä kaksi taustakehikkoa, paradigmaa, joiden kautta yleensäkin maailmaa arvioimme, liittyvät kiinteästi toisiinsa. Ihmiskäsitystä ja arvoja yhdistävä iso kysymys kuuluu: millaista olisi ihmiskuvani mukaisen ihmisen hyvä elämä? Ihmiskuvan ydintä on, ”mihin ihmiset ovat menossa” (Taylor 1989, 47). Ihmisten päämäärät, mitä he pitävät hyvänä ja tavoiteltavana, kertovat heistä olennaisimman. Ihminen tuottajana ja kuluttajana on varmasti oman aikamme yhteiskunnassa yksi keskeisimmistä ”hyvän elämän kertomuksista”.

Mihin arviointi perustuu?

Arvioinnin ydinkysymys koskee arvioinnin kriteerejä: mihin arviointi perustuu? Todellisuudessa arvioinnin kriteerit tulevat arvioitsijoille ulkoapäin annettuina, ja sen vuoksi kriteerien kysyminen saattaa jäädä näennäiseksi. Joku ulkopuolinen taho on jo päättänyt ne arvioitsijan puolesta. Kysymys arvioinnin perustasta on kuitenkin syytä esittää silloin, kun arvioitsija ei ole kokonaan ulkopuolisen määräysvallan alainen, kun hänellä on vähänkin vapautta määrittää itse arviointityönsä kriteerejä. Samalla kun hän kysyy tuon kysymyksen aidosti ja avoimesti, hän huomaa, että koulutuksen maailmassa käytetyt kriteerit eivät ole lainkaan itsestään selviä, niiden perusta on vaikeasti hahmoteltavissa. Mihin perustuu esimerkiksi kaikki määrällinen arviointi, jota pidetään varsin yleisesti selkeimpänä arvioinnin muotona? Voiko kaikkea, esimerkiksi laadullisia saavutuksia, arvioida määrällisesti?

Jos arvioinnin kriteerejä ei oteta annettuina ja hyväksyttyä sellaisenaan, joudutaan niitä ajattelemaan itse. Tämä on yksi autonomisen arvioinnin ulottuvuus. Ajatteluprosessi alkaa luontevasti kysymyksestä, voiko arvioinnille löytää objektiivisia perusteita vai tekeekö jokainen sitä vain oman tuntumansa mukaisesti, täysin subjektiivisesti. Arvioinnin täydellinen objektiivisuus lienee mahdotonta siksi, että arviointi on aina arvosidonnaista. Arvot, intressit, hyvinä pidetyt asiat puolestaan syntyvät, nousevat ja kuolevat historiallisten tilanteiden, yhteiskunnallisten intressien mukaan. Ne ovat tässä mielessä aina jossain määrin suhteellisia. Arvioitsija on myös omine arvoineen ja ajatuksineen osa yhteiskuntaansa ja kulttuuriaan, ja näin ollen hänenkin näkökulmansa voi olla vain suhteellinen (arvojen luonteesta ks. Kekes 1993).

Kuten edellä on jo kuvattu, rakentuvat myös koulutukset erilaisissa arvokehysissä erilaisten intressien tuloksina. Niille asetetaan tavoitteet eli määritetään, mitä on se Hyvä, jota tässä koulutuksessa pyritään toteuttamaan. Mihin siis voisi koulutuksen arviointi perustua, jos kriteerejä eivät määrää jotkin koulutukselle ulkoiset tahot? Autonominen arviointi voinee perustua vain sen selvittämislle,

kuinka määritetyt tavoitteet ovat toteutuneet koulutuksen käytännössä. Toisin sanoen *arvopohjaiset tavoitteet määräävät arvioinnin kriteerit*.

APO-koulutuksen tavoitteeksi on määritetty hyväksi aikuiskouluttajaksi kasvaminen. Mainittu hyvyys on ollut jatkuvien keskustelujen ja tutkimisen kohde. Se on täydentynyt ja tarkentunut ajan mittaan analyttisissä keskusteluissa, mutta varmasti se tulee vieläkin muuttumaan, kun koulutuksen järjestäjien ymmärrys asiasta muuttuu. Hyvän aikuiskouluttajan tuntomerkeiksi on tällä hetkellä määritetty autonomisemmaksi kasvaminen ja pedagoginen taitavuus. Nämä ovat siis perusta opiskelijoidemme arvioinnille.

Millaisin kriteerein siis teemme arviointia? Kysymyksen voisi muotoilla myös näin: mitä asettamamme tavoitteet tarkoittavat todellisuudessa, toteutuneina, miten tunnistamme ne opiskelijoiden toiminnassa? Eli *arvioinnin ensimmäinen perusta on koulutuksen tavoitteen luonteen ymmärtäminen*. Meidän tapauksessamme tämä on osoittautunut erittäin paljon aikaa, keskustelua ja tutkimusta vaativaksi tehtäväksi, joka näyttää myös suhteellisen loputtomalta. Hyvyyksien määrittely ei liene koskaan niin yksinkertaista, että sen voisi kirjoittaa vaikkapa itsestäänselvyytenä opetussuunnitelmaan (millainen on hyvä pedagogi, hyvä lääkäri, hyvä muurari, hyvä tutkija jne.).

Jos arvioitsija on tiedostanut ja ymmärtänyt koulutuksen tavoitteet – aina toki jollain suhteellisella tasolla –, on hänellä edessään toinen tehtävä: hänen tulee selvittää, tutkia, mitä koulutuksessa tapahtuu, jotta hän voisi punnita tavoitteiden ja toteutuman välistä suhdetta. Tämän selvitystyön tulosten esittäminen ei itsessään ole vielä arviointia. Arviointia on vasta edellä mainittu punnitseminen. Ainakin meidän tapauksessamme se on osoittautunut hyvin prosessiluonteiseksi sen sijaan, että kyseessä olisi suoraviivainen yhtälö, jossa verrattaisiin vain tavoitetta ja toteutumaa toisiinsa. Koulutuksen toteutuman ja kokemusten tutkiminen nimittäin antaa uutta ymmärrystä itse tavoitteesta. Kun arvotavoitteet puolestaan tarkentuvat tai muuttuvat, muuttuvat myös arviointiperusteet. Arvioinnista tulee näin perusteiltaan *dynaamista*.

Kun arvioinnissa on päästy vaiheeseen, jossa on punnittu koulutuksen nykyinen tila, on jäljellä vielä arvioinnin viimeinen askel, joka kylläkin palautuu kehittävässä arvioinnissa uudella tasolla takaisin alkupisteeseen. Se on jälleen selkeästi arvopitoinen kysymys: mihin suuntaan tästä pitäisi edetä, miten tätä koulutusta pitäisi kehittää?

Itsearviointi ja ulkoinen arviointi arvokeskusteluna

Lähdemme ajatuksesta, että koulutuksella on suhteellinen autonomia omien arvojen suhteen ja arviointi lähtee niiden perustalta. Arviointi on siten suhteellisen autonomista. Kummallekaan taholle arvot eivät siis tule ulkopäin valmiiksi

annettuina. Olennaiseksi nousee kysymys, ovatko arvioitavan koulutuksen arvot hyvin perusteltuja? *Pystyykö koulutus itse perustelemaan arvonsa*, onko se edes miettinyt niitä, tunnistaako se arvolähtökohtansa ja tavoitteensa? Toisen sanoen: onko koulutuksen toteuttajilla kehittynyt *arvotietoisuus*? Jos ei ole, niin edessä on kaksi mahdollisuutta: 1) Koulutuksen pitää aloittaa itsearviointityö arvojen tiedostamiseksi ja mahdollisesti muuttamiseksi, jos niille ei löydy perusteita; arvothan usein omaksutaan annettuina ilman kriittistä arviointia. 2) Ulkoinen arvioija auttaa koulutusta tuossa tiedostamistyössä, toimii peilinä. Tämä edellyttää ulkoisen ja sisäisen arvioinnin dialogia. Voisiko oppilaitoskulttuurien arvotietoisuuden kehittäminen tällaisessa dialogissa siis olla olennainen osa arviointisijoiden ammattikunnan työtä?

Jos arviointi toteutetaan kaikkien osapuolten näkemyksiä kuunnellen ja huomioiden, ongelmaksi muodostuu arviointiin osallistuvien tahojen kyky kriittiseen itsearviointiin, joka on yksi autonomisen yksilön tai yhteisön tunnusomainen piirre. Itsearviointi ei ole itsestään selvästi taito, johon kaikki kykenevät. Itsearviointiin täytyy oppia, ja siksi ei voida automaattisesti olettaa, että itsearvioinnin kautta ”tuotetaan” totuudellista tietoa arvioinnin kohteesta. Itsearvioinnin edellytyksenä on siis ainakin autonomisuus ja arvotietoisuus. Itsearviointi on mahdotonta, ellei itseään arvioiva voi asettaa arvioinnilleen kriteerejä. Mutta miten onnistuu kriteerien asettaminen oman toiminnan arvioinnille?

Ainakin yksi mahdollinen vastaus ovat keskustelut, joissa käsitystä omasta toiminnastaan voi suhteuttaa toisten näkemyksiin. Jos itsearvioinnin kohteena on yhteisö tai organisaatio, keskustelut jäsentävät ja tekevät näkyväksi oman toiminnan arvoja, toimintaperiaatteita ja pedagogisia ratkaisuja. Tämä edellyttää kuitenkin, että kaikki organisaation jäsenet ovat mukana keskustelussa. Tällöin kaikki organisaation sisällä olevat – erilaisetkin – intressit tulevat mukaan keskusteluun.

Itsearviointikeskustelut ovat osallistujille haastavia: kukaan ei pääse helpolla. Itsearviointiin liittyy olennaisesti tapojen ja tottumusten tekeminen näkyväksi. Yksilötasolla tämä tarkoittaa oman sanattoman tiedon ”aukipuhumista” (miten osaaminen ja taito jäsennetään arvioinnin kohteeksi). Yksilöllisten kuvausten ”summa” kuvaa puolestaan arvioitavan organisaation osaamista, koulutuksen konkretiaa. Näissä keskusteluissa lienee tärkeintä, että ”aukipuhuminen” jatkuu niin kauan, että kaikki koulutuksen laatua määrittävät tärkeät asiat tulevat käsitellyiksi ja yhteisesti arvioituiksi.

Oman kokemuksemme mukaan APO-koulutuksessa tällainen prosessi ei pääty johonkin tiettyyn ajankohtaan, vaan itsearviointikeskustelu on jatkuvaa. Itsearviointia toteuttava yhteisö altistuu kaikissa yhteisissä tapaamisissaan oman toiminnan kriittiselle arvioinnille (~ yhteisöllinen reflektointi). Oleellinen piirre itsearviointia toteuttavassa yhteisössä on myös, että arviointi ei jää vain oman organisaation sisälle. Kriittinen arviointi ulotetaan siis koskemaan myös ympäri-

vää yhteiskunnallista todellisuutta, esimerkiksi sieltä tulevia arvopaineita. Itse asetetut kriteerit ja yhteiskunnan määrittämät kriteerit tulevat arvioiduksi.

Miten tähän itsearvioinnin kuvioon sitten suhteutuu niin sanottu ulkopuolinen arviointi? Mitä ”ulkopuolinen” voi sanoa ja millä perusteella? Mitä on niin sanottu ”riippumaton tieto”, josta arvioinnin kehittämiseen liittyvissä keskusteluissa puhutaan? Minkälainen on ulkopuolisen tai riippumattoman arvioijan arvotietoisuus? Ulkopuolinenkin arvioija tarvitsee aina arvioinnilleen *kriteerin*. Kehittäminen itsessään ei voi olla kriteeri. Kehittämisellä täytyy aina olla suunta, ja suunta täytyy tietoisesti valita.

Kehittävissä arviointimenetelmissä (esim. Räisänen 2005) korostetaan yhteisöllisyyttä, monitahoisuutta ja itse ”subjektin” oman käsityksen arvoa ja merkitystä. Siis ainakin teoriassa on olemassa mahdollisuus, että arvioinnin kriteerit asetetaan ”demokraattisesti”, kaikkien osapuolten näkemystä kuunnellen ja huomioiden yhteisessä dialogissa.

Opettajat ja koulujen johtajat ovat jo pitkään puhuneet siitä, että itse opettamiseen on yhä vähemmän ja vähemmän aikaa. Opettajien työn luonne on muuttunut vähitellen tähän suuntaan, ja syynä muutokseen ovat ennen kaikkea ulkoapäin tulevat paineet, muun muassa koulutuksen markkinointiin ja rahoituksen hankkimiseen liittyen. Usein on niin, että koulutuksen toteutumiseen vaikuttaa ensisijaisesti taloudellinen kannattavuus eikä esimerkiksi yhteiskunnallinen tilanne. Jos arvioinnin kautta saadaan tällaisetkin vinoumat ja epätasapainot näkyviin, kehittämisehdotukset eivät välttämättä kohdistukaan vain itse oppilaitoksen toimintaan, vaan koulutuspoliittiselle tasolle asti. *Arvojen ketju pitäisi kokonaisuudessaan tehdä läpinäkyväksi suurista yhteiskunnallisista kehyksistä aina yksittäisiin oppilaitoksiin ja niiden opiskelijoihin saakka, ja siten kehittämisehdotuksetkin voisivat ulottua arvoketjun kaikille tasoille.*

Arvioinnin yhteydessä puhutaan sidosryhmistä, joiden mukanaolo nähdään välttämättömäksi. Tärkein sidosryhmä koulutuksen arvioinnissa ovat itse oppijat. Miten oppijat arvioivat koulutusta omasta näkökulmastaan ja omista tarpeistaan käsin? Oppijat ovat toki olleet mukana arviointiprosesseissa, mutta ongelma heidän suhteensa on pikemminkin arvioinnin ajankohta ja opiskelijoiden kyky arviointiin. APO-koulutuksessa olemme panneet merkille, että opiskelijat ovat tottuneet olemaan arvioitavina ja he ovat tottuneet myös siihen, että tavoitteet heidän oppimiselleen määritellään ulkoapäin. Kun HOPS:n kirjoittaminen vaatiikin aidosti kykyä itsearviointiin ja oman osaamisen arviointiin, tilanne koetaan vaikeaksi, joskus jopa ahdistavaksi. Jos ja kun arviointiin sisällytetään oppijoiden tekemää arviointia, heitä täytyy opettaa ja ohjata tekemään sitä. Oppija liepee kuitenkin yksi arvioinnin niin sanotuista ”subjekteista”.

Oppimisen arvioinnin täytyy ulottua riittävän pitkälle tulevaisuuteen, jotta oppimiseksi määritellyt muutokset ovat ylipäättään ehtineet tapahtua. Oppilaitokset kehittävät usein toimintaansa juuri opiskelijapalautteen perusteella, mutta

tuo palaute on kerätty yleensä välittömästi koulutuksen yhteydessä tai hyvin pian sen jälkeen. Arvioivaa palautetta kyetään kuitenkin antamaan vasta, kun opittua on päästy testaamaan. Arvoketjun tämäkin linkki vaatii kriittistä arviointia ja oikeanlaista kysymistä. Tyydyemmekö mittaamaan helposti mitattavia asioita?

Arvotietoisemmaksi arvioitsijaksi

Tiivistämme edellä sanotun seuraavalla tavalla: Koulutusten toteuttaminen, arviointi ja kehittäminen on arvoperusteista. Koska ihmiset ja heidän organisaationsa eivät yleensä ole toiminnassaan itsetietoisia, jäävät toiminnan perustana olevat arvot suurelta osin tiedostamattomiksi. Tämän vuoksi koulutusten arvioinnissa ja kehittämisessä mukana olevilta henkilöiltä voidaan edellyttää huomion kiinnittämistä arvoihin, pyrkimystä kasvattaa omaa arvotietoisuuttaan. Tämä on erityisen tärkeää silloin, kun pyritään kriittiseen ja autonomiseen arviointiin, kun arvoja ei oteta annettuina, kun ei toimita vain ulkoisen ohjauksen varassa. Itsearviointi tulee erityisen keskeiseksi niissä tapauksissa, joissa koulutuksella ja opiskelijoilla on vapautta tehdä autonomisia arvovalintoja.

Kun arvioija haluaa kehittää omaa ammatillista osaamistaan, on lähtökohdiana *oman* arvotietoisuuden tutkiminen. Se tarkoittaa oman merkitys- ja arvoperspektiivin tutkimista, toisin sanoen omien perustavaa laatua olevien arvojen ja niiden perustelujen läpinäkyväksi tekemistä. Arvioijan oma arvoperspektiivi on väistämättä rajaamassa hänen työtään arvioijana. Oman arvoperspektiivin tiedostaminen estää osaltaan vinoutumien tai vääristymien muodostumista.

Oman arvoperspektiivin tutkiminen ulottuu lopulta aivan perustavimpiin oletuksiin ihmisestä ja maailmasta – omaan ihmiskäsitykseen ja siihen kytkeytyvään tietokäsitykseen. Arvioija voi katsoa arvioinnin kohdetta esimerkiksi humanistisen tai naturalistisen ihmiskäsityksen läpi: mihin hän arvioinnin kohteessa kiinnittää huomiota (jopa kriteereistä huolimatta ja niiden ohii)?

Eettinen ulottuvuus on olennainen osa arvioitsijan arvotietoisuutta. Keskeinen eettiseen ulottuvuuteen liittyvä *kysymys itselle* olisi: ”Mitä tästä arvioinnistani seuraa arvioinnin kohteelle?” ja ”Seuraako siitä jotakin hyvää?”.

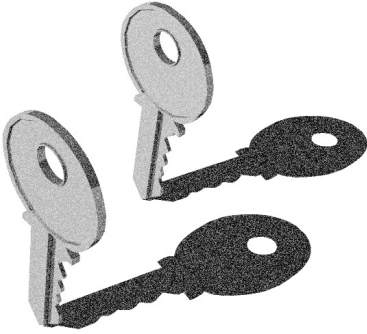
Arvioijalta voidaan odottaa myös aitoa kiinnostusta ja pyrkimystä *arvioinnin kohteen parempaan ymmärtämiseen*. Tällä tarkoitetaan arvioitavan kohteen perusolemuksen ja -tehtävän ymmärtämistä (nämä ovat pikemminkin kysymyksiä ”Mitä varten?” kuin ”Mitä?”).

Kokoavasti voi todeta, että arvioijana kehittyminen on autonomisoitumisprosessi, jonka seurauksena arvioija voi tehdä havaintoja, johtopäätöksiä ja kehittämis ehdotuksia *autonomisena arvioijana*, toisin sanoen omat ja ulkopuolelta tulevat rajoitukset tiedostavana ja siten niistä niin vapaana kuin se inhimillisessä

elämässä on mahdollista. Arvioijan työ vaatii erityisesti herkkyyttä ja avoimuutta siihen liittyvän vallan ja vaikuttavuuden vuoksi.

Lähteet

- House, E. R. 1983. How we think about evaluation. Teoksessa E. R. House (toim.) *Philosophy of evaluation*. San Francisco: Jossey-Bass.
- Kekes, J. 1993. *The morality of pluralism*. Princeton: Princeton University Press.
- Räisänen, A. 2005. Kehittävään arviointiin. Teoksessa K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnista*. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä, 109–129.
- Taylor, C. 1989. *Sources of the self*. Cambridge: Harvard University Press.
- Uusikylä, P. 1999. Poliitiikan ja hallinnon arviointi. Teoksessa R. Eräsaari, T. Lindqvist, M. Mäntysaari & M. Rajavaara (toim.) *Arviointi ja asiantuntijuus*. Helsinki: Gaudeamus, 17–30.



Arvottaminen on arvioinnin ydintä

Arvot, arvoteoriat ja arviointi kytkeytyvät toisiinsa. Arvioinnissa on aina mukana arvoarviointia. Arviointi on arvotonta, ellei siihen sisälly arviointikohteen arvottamista. Arviointitieto ei ole arvoneutraalia. Arvioinnin ydintehtävä on arvottaminen – milloin kriteerien suhteen, milloin muita lähestymistapoja soveltaen. Artikkelin kirjoittajat pyrkivät selkiyttämään ja vahvistamaan arvioitsijoiden arvotietoisuutta: jokaisen arvioitsijan tulisi selvästi tiedostaa arvioinnin suhde kohteen arvotaustaan. Arvot määräävät, mitä yksilö priorisoi valintoja tehdessään. Arviointi on arvon antamista tai määrittämistä. Kasvavaksi ongelmaksi on koettu, että arvioinnin taustalla olevista arvoista ei keskustella, niitä ei pohdita eikä perustella, ei kyseenalaisteta eikä aina edes tiedosteta. Arvoarviointi mahdollistaa erilaisten näkökulmien ja vaihtoehtojen tuomisen esille sekä syventää ja selkiyttää arviointitietoisuutta. Artikkelissa käydään läpi arvo-käsite yksilön, yhteisön ja yhteiskunnan näkökulmasta tarkasteltuna. Samalla esitetään joitakin yleisimpiä arvoluokituksia sekä kuvataan lyhyesti tunnetuimmat arvoteoriat. Teoreettis-filosofisista tarkasteluista käytetään esimerkkinä tasa-arvon olemusta; tasa-arvohan liittyy perusjuonteena suomalaiseen koulukulttuuriin ja sen arviointiin. Lopuksi tarkastellaan arvioinnin arviointia eli meta-arviointia ja erityisesti perusarvoja arvioinnin laadun arvottamisessa.

Asiasanat: Arvo, arvostaminen, arvottaminen, arviointi, arvoarviointi, meta-arviointi

Arvoista

Näyttää siltä, että arvojen ja arvioinnin yhteyttä pidetään itsestään selvänä. Liian usein tyydytään vain toteamaan, että arviointi on arvon määrittämistä arviointikohteelle. Arvojen ja arvioinnin yhteyttä ei ole juuri analysoitu sen enempää arvojen kuin arvoteorioidenkaan kannalta. Myös itse arviointitoiminnassa arvojen tarkastelu saatetaan sivuuttaa kokonaan. Arvot eivät kohoa arviointitietoisuuteen. Näistä lähtökohdista on tässä yhteydessä tarkoitus palata perusteisiin eli arvoihin ja arvottamiseen eräänlaisena arvioinnin kivijalkana tai syvärakenteena.

Arvo on tieteen ja arkielämän kielenkäytössä hyvin monimerkityksinen ja merkityssisällöltään väljä riippuen eri yhteyksistä ja tieteenaloista, joissa käsitettä käytetään. Käsitteen merkityssisältö vaihtelee sen mukaan, käytetäänkö sitä taloustieteessä, matematiikassa, estetiikassa vai moraalifilosofiassa. Terminologisesti käsitettä arvo on käytetty ensimmäisenä juuri talouden piirissä ja vasta myöhemmin filosofiassa, jossa sen merkitys palautuu hyvyyden ja totuuden väliin pohdintoihin (Joas 2000, 21). Tässä yhteydessä arvoja tarkastellaan lähinnä moraalifilosofisena kysymyksenä ja käytännölliseen arviointiin liittyen.

Arvo-käsite liitetään kirjallisuuden lukuisissa määritelmissä yksilön tai ryhmien haluihin, tarpeisiin, mieltymyksiin, toivomuksiin, odotuksiin, koettuihin tunteisiin, sisäistettyihin preferensseihin, toiminnan päämääriin, käyttäytymiseen, valintatapumuksiin tai toiminnan tulosten arviointikriteereihin (Ahlman 1976; Allardt 1964; Rainio & Helkama 1974; Rokeach 1973; Entonen & Lyytinen 1979; Pietarinen 2000). Arvo on abstraktio, eräänlainen standardi ja arvostuksen tuloksena syntynyt vakiintunut suhtautumisohjain eli arviointikriteeri, joka ohjaa henkilön tai ihmisryhmän valintoja ja ilmenee sanoissa ja teoissa sekä yleensä käyttäytymisessä sen perustana (vrt. Peltonen & Ruohotie 1992; Hirsjärvi 1983; Pietarinen 2000). Arvot ovat ihmisen ajattelussa ja yhteiskunnan kulttuurissa vallitsevia käsityksiä yksilöiden, yhteiskunnan ja ihmiskunnan keskeisistä päämääristä. Ne ovat siis käsityksiä hyvästä elämästä, hyvästä yhteiskunnasta ja hyvästä maailmasta (Suhonen 1988; vrt. myös Schein 1987). Pihlström (2000) tuo arvoihin edellisistä poikkeavan vivahteen todetessaan, että arvoja voidaan pitää hypoteeseina, joita jatkuvasti testataan ja tarvittaessa korjataan jonkinlaisessa ”elämän laboratoriossa”.

Yksilön kannalta yhteiskunnan tai yhteisön käyttäytymisessä ilmeneviä tavoitteita voidaan kutsua yhteiskunnan tai yhteisön piirissä vallitseviksi arvoiksi. Entonen ja Lyytinen (1979) mainitsevat yhteisön arvo-ullottuvuuden esimerkkinä tasa-arvoisuuden, joka heidän mukaansa on katsottava lähinnä yhteiskunnan rakenteen ominaisuudeksi. Esimerkiksi suomalaisen koulutuspolitiikan toimivuutta arvioitaessa hyvin keskeinen arvo on tasa-arvo eettisenä ihanteena ja koulutuspolitiikan lähtökohdana – tavallaan ”yhteisenä hyvänä”. Tätä tarkastellaan lähemmin artikkelin luvuissa 4 ja 5. Arvojen suhde yhteiskuntaan näkyy

myös yhteiskunnan instituutioissa. Yhteiskunnassa arvoilla on siis myös organisatorisia ilmentymiä. Juridisia oikeusarvoja edustaa esimerkiksi oikeuslaitos ja uskonnollisia arvoja kirkko. Yliopistot ja koululaitos edustavat teoreettisia ja sosiaalisia arvoja (Ahokallio 1987). Yksilöllisen orientaation lisäksi arvoilla on siten myös yhteiskunnallinen luonne. Esimerkiksi yksilön arvo työmarkkinoilla ei määräydy vain subjektiivisesti, vaan keskeisesti myös yhteiskunnallisesti (vrt. Engeström 2005).

Hirsjärven (1983) ja Niiniluodon (1993) mukaan arvot voidaan jakaa itseisarvoihin ja välinearvoihin. Jaottelu itseisarvoihin ja välinearvoihin on peräisin jo Aristoteleelta, joka totesi, että toiset asiat ovat itsessään arvokkaita, toiset niiden kautta (Aristoteles, suom. 1989). Itseisarvoja eli päätearvoja Hirsjärvi kuvaa lopullisina päämäärinä ja Niiniluoto sinänsä tai sellaisenaan arvokkaina riippuen siitä, voidaanko niiden avulla saavuttaa joitakin muita hyötyjä. Itseisarvoina on pidetty muun muassa moraalista hyvää, totuutta, kauneutta, oikeudenmukaisuutta ja pyhää. Välinearvot eli instrumentaaliset arvot ovat Hirsjärven mukaan nimensä mukaisesti keinoihin verrattavia pyrittäessä itseisarvoihin. Niiniluoto kuvaa välinearvot asioiksi, joiden uskotaan olevan hyödyllisiä jonkin muun arvokkaana pidetyn asian tavoittelemisessa.

Ahlman (1976) on luokitellut arvotyyppjä seuraavasti: a) hedoniset arvot (onni, mielihyvä, ilo, nautinto, aistillisuus), b) vitaaliset arvot (elämä, terveys, tahto, kuntoisuus), c) esteettiset arvot (kauneus, ylevyys, suloisuus, taide), d) tiedolliset arvot (totuus, tieto, oppi, koulutus, viisaus, tiede), uskonnolliset arvot (usko, toivo, pyhyys, laupeus), e) sosiaaliset arvot (altruismi, ystävyys, rakkaus, uskollisuus, vapaus, veljeys, kunnia, isänmaallisuus, turvallisuus), f) mahtiarvot (voima, valta, sota, rikkaus, raha, voitto), g) oikeusarvot (oikeudenmukaisuus, ihmisoikeudet, tasa-arvo, laillisuus) ja h) eettiset arvot (hyvyys, moraalinen oikeus). Niiniluodolta (1993, 85) saadaan Ahlmanin luetteloa täydentämään i) ekologiset arvot (luonnonkauneus ja terveys, eläinten oikeudet) sekä j) egologiset arvot (omanarvontunne, itsekkyyys, oma etu). ”Sen mukaan mitä tarpeita arvot tyydyttävät, ne voidaan jakaa arvo- eli sisältöalueisiin eli arvomaailmoihin, esim. intellektuaaliseen, uskonnolliseen, taloudelliseen ja vitaaliseen arvoalueeseen” (Hirsjärvi 1983, 17).

Turunen (1993, 22) identifioi omaksi arvokategoriakseen niin sanotut platoniset arvot, joita hän nimittää varsinaisiksi arvoiksi. Ne ovat edellä mainitut totuus, hyvyys ja kauneus. Näitä arvoja tarkastellaan myöhemmin arvioinnin laadun arvoina, joita voidaan soveltaa yleisellä tasolla meta-arvioinnissa, kun arvioidaan arvioinnin laatua.

Arvohierarkialla tarkoitetaan arvojen järjestystä. Esimerkiksi itseisarvot sijoitetaan arvojärjestyksessä korkeammalle kuin välinearvot. Hirsjärven (1983) mukaan yleensä katsotaan, että arvot, joilla on ajallista kesto, ovat korkeammalla kuin arvot, joiden kesto on ajallisesti vähäinen. Esimerkiksi niin sanotuilla pla-

tonisilla perusarvoilla on ollut historiallista kestävyyttä. Arvojen hierarkioita on yritetty laatia muun muassa sen mukaan, mikä on arvojen esiintymisfrekvenssi yhteisön jokapäiväisessä elämässä. Jos arvot määritellään valinnan kriteereiksi, kuten usein tehdään, merkitsee hierarkkisuus tällöin sitä, että A valitaan ennen B:tä, B ennen C:tä ja myös A ennen C:tä. (Hirsjärvi 1983, 17). Täten arvoja priorisoidaan eli asetetaan keskinäiseen arvojärjestykseen.

Arvojen osalta voidaan tehdä jako arvosubjektiin ja arvo-objektiin. Arvojen taustalla on Niiniluotoa (1993; 1994, 117) lainaten aina ”arvostamiseksi” kututtava inhimillinen toiminta. Arvostaminen on suhtautumistapa tai akti, jossa joku arvosubjekti X pitää jotain kohdetta eli arvo-objektia A arvokkaana. Arvoasetelmien perusrakenne noudattaa kaavaa ”X pitää A:ta arvokkaana” tai ”A on arvokas X:lle”. X siis arvostaa A:ta, jolloin X:llä on uskomus, että A on hyvä tai X:n kannalta on hyvä, että A vallitsee. Arvosubjekti X voi olla yksilö, ryhmä tai kokonainen kulttuuri. Arvo-objektista A voidaan erottaa a) arvon kantaja (konkreettinen olio, esine tai tapahtuma) ja b) arvo-ominaisuudet (arvostuksen perustana olevat arvo-objektin piirteet (ks. emt. 85).

Arviointi perustuu siten prosessina arvostuksiin. Arvostuserot näkyvät esimerkiksi valituissa arviointikriteereissä ja -standardeissa (Linnakylä & Välijärvi 2005; vrt. Lyytinen 2006). Joskus kriteerit ja standardit otetaan käyttöön annettuina ja arviointimalliin kuuluvina, ilman että arvoanalyysissä niitä millään tavoin preferoidaan tai sovitetaan arviointiobjektiin. Tällöin arviointia määrää enemmän annettu arviointimalli kuin sen taustalla olevat arvot, arvioitsijan omat arvot tai soveltuvuus arviointiobjektiin.

Arvoteoriat ja arvottaminen

Tässä yhteydessä pyritään arvoteorioiden osalta hyvin pelkistettyyn tarkasteluun luonnehtimalla lyhyesti arvojen perusteorioita ja tarkastelemalla niiden mahdollista yhteyttä arviointiin. Filosofiasia esitetyt arvoteoriat pyrkivät vastaamaan seuraaviin kysymyksiin: Ovatko arvot olemassa ihmisestä riippumatta? Ovatko arvot vain henkilökohtaisia kannanottoja vai onko niillä objektiivisia perusteita? Onko olemassa ”oikeita” arvoja? (Niiniluoto 1993, 87.)

Arvotoreettiset eli aksiologiset (vrt. antiikin kreikan aksia ’arvo, hinta’) kysymykset ovat liittyneet keskeisesti arvojen subjektiivisuuteen ja objektiivisuuteen. Subjektiivinen viittaa siihen, että arvot ovat puhtaasti subjektiivisia ominaisuuksia, joilla ei ole mitään ontologista sijaa ihmismielen ulkopuolella. Arvot tarvitsevat siis arvottajansa. Subjektiivisuudella viitataan jonkinlaiseen puolueelliseen tai näkökulmaltaan rajoittuneeseen ja objektiivisella vastaavasta näkökulmasta riippumattomaan, todelliseen. Kaikki arvo on suhteessa subjektiin.

Hyvyys on näin ollen hyvyyttä tietyllä subjektilla, ja hyvyys tietyllä subjektilla on taas hyvänä pitämistä.

Aarnio (2007) erottaa kahdenlaista objektiivisuutta: a) epistemologisen objektiivisuuden ja b) metafyyssisen objektiivisuuden. Epistemologinen objektiivisuus tarkoittaa sellaista tosiseikkaa, jonka totuus tai luonne on tasapuolisesti useiden tai ihannetapauksessa kaikkien tarkkailijoiden (arvioitsijoiden) havaittavissa – eli intersubjektiivista tai muuten tavoitettavissa. Metafyyssinen objektiivisuus puolestaan on sellainen propositio, joka on tosi, totuus ei-ymmärrettynä epistemisesti. Totuus ei riipu siitä, tiedetäänkö tuota totuutta.

Subjektiivisuuden ja objektiivisuuden lisäksi aksiologisten pohdintojen kohteena on ollut arvojen luonnollisuus eli arvojen sidostuminen arvo-objektin ominaisuuksiin: arvo-ominaisuuden ja luonnollisen ominaisuuden välillä olisi siis välttämätön yhteys. Tällöin arvot palautuvat esimerkiksi yksilöiden tarpeisiin tai motivaatioon. Arvoihin suhtaudutaan siis realistisesti. Luonnollinen on taas kovin moniselitteinen tarkasteltaessa esimerkiksi luonnollisen ja ei-luonnollisen välisiä eroja, jolloin joudutaan helposti semanttiseen kiistaan. Yksi tapa luonnehtia luonnollista on yhtäläistä se todelliseksi.

Arvoteoriat voidaankin jakaa kolmeen luokkaan: a) objektivistinen, b) subjektivistinen ja c) naturalistinen tai naturalistis-realistinen arvoteoria. Mooren objektivistisen arvoteorian (Moore 1903) mukaan arvot ovat olemassa ihmisestä riippumatta ja arvot edeltävät haluja. Arvo on aistein havaitsematon ei-luonnollinen ominaisuus. Järkevä keskustelu arvokysymyksistä on mahdotonta, koska jokainen ihminen näkee intuitionsa avulla, mitkä asiat ovat objektiivisesti hyviä. Subjektivistisen arvoteorian mukaan jokaisella yksilöllä on oma hyvänsä. Arvosubjektivismiin mukaan ”arvot ovat pelkkiä subjektiivisia arvostuksia, mieltymyksiä tai asenteita” (Puolimatka 1999, 265). Toisin kuin objektivistisessä teoriassa halu edeltää arvoa ja luo arvon. Russellin (1947, 237) subjektivismiin mukaan arvoväitteitä ei voida perustella rationaalisesti, koska ne ovat mielettömiä ja arvokeskustelu on siten tyhjänpäiväistä kinastelua – arvokiistat ovat makuasioita (ks. esim. Russell 1947, 237–238). Naturalistisessa arvoteoriassa arvoa ei määrää yksilön halu, vaan lajityypillinen halu tai tarve. Aristotelisen naturalismin mukaan arvo-ominaisuudet voidaan määrittellä tosiasiaominaisuuksien eli luonnollisten ominaisuuksien avulla. Naturalismin mukaan arvojen ja tosiasioiden välille on mahdollista rakentaa silta: naturalismi tekee rationaalisen keskustelun arvokysymyksistä mahdolliseksi. (Niiniluoto 1993, 87.) Arvokeskustelua voitaisiin siten käydä kuvailemalla ja analysoimalla arvo-objektin ominaisuuksia. Siinä tukena voidaan käyttää esimerkiksi tutkimustietoa. Naturalistisessa arvoteoriassa voidaan Kotkavirran ja Nyysösen (1996, 35–36) mukaan kääntää arvolauseet tieteelliseen tosiasiatietoon perustuviksi lauseiksi. Tämä teoria antaa tavallaan perustellun luvan arvo-arvioinnille.

Arvorteorioita tiivistäen voidaan todeta, että 1) objektivistisen suuntauksen ongelmana on, kuinka arvoista saadaan tietoa, 2) subjektivistisen suuntauksen ongelmana on arvon kaiken yleispätevyyden häviäminen ja 3) naturalistinen suuntaus mahdollistaa arvorkorjauksen perustelemisen tosiasioiden avulla.

Virtanen (2004, 18–20) on tarkastellut kriittisesti nykyisten tarjolla olevien arviointistandardien soveltamiseen liittyviä rajoituksia. Hän esittelee Shadishin ja hänen tutkijakollegoidensa (1991, 46–64) tekemän kolmijaon arvottamisen lähestymistavoista, joita ovat deskriptiivinen, preskriptiivinen ja metateoreettinen. Virtasen mukaan kyseinen jako on avuksi, kun yritetään ymmärtää nykyisten arviointistandardien johdonmukaisuutta. *Deskriptiivinen arvottaminen* on yksinkertaisesti kuvaus asianomaisten arvoista. Sen vuoksi tieto asianomaisten arvoista on yhteydessä arvioinnin poliittiseen kontekstiin. Deskriptiivisiä lähestymistapoja rajoittaa se, millaiseksi arvioitsija kuvittelee asianomistajan moraalien ja eettiset sitoumukset. Deskriptiiviset teoriat kuvailevat arvoja väittämättä, että yksi arvoista on paras muihin arvoihin verrattuna. *Preskriptiivinen arvottaminen* sen sijaan puolustaa tiettyjen arvojen etusijaa tai johtoasemaa. Esimerkiksi Rawlsin (1972) oikeudenmukaisuusteoriassa oikeudenmukaisuuden arvo on tärkeä, mutta silti se on vain yksi monien muiden tarjolla olevien arvojen joukossa kuten esimerkiksi ihmisoikeudet, vapaus, tasa-arvoisuus, etuisuus jne. Virtanen toteaaakin, että jokainen arvioitsija, joka on tekemässä tutkimus- tai arviointitoimenpiteitä ja joka laittaa etusijalle preskriptiiviset eli ohjailevat eettiset teoriat, on vaikean tehtävän edessä yrittäessään löytää toiminnalleen oikeutuksen ja selityksen. Scrivenin (1980) alkujaan aloittama *metateoreettinen arvottaminen* on yhteydessä arvottamisen luonteen tutkimiseen sekä arvottamisen oikeutuksen analysointiin. Se kuvailee, miten ja miksi arvoilmaukset on konstruoitu, korostaa arvottamisen rakennetta tai johdonmukaisuutta sekä yrittää paljastaa arvojen oikeuttamisen luonnetta. (Virtanen 2004.)

Tarkastelunsa loppupäätelminä Virtanen esittää, että nykyisillä arviointistandardeilla on vain vähän tarjottavanaan arvo- ja etiikkakeskusteluille. Ongelmana on, että standardit eivät välttämättä ole yhteydessä arvoteoriaan. Seuraamalla erilaisten standardien käyttöohjeita ei arvioitsija voi vakuuttua siitä, että arviointi on automaattisesti laadultaan hyvää ja eettisesti korkealaatuista. Soveltamalla teknisesti esimerkiksi ISO-standardeja tai Euroopan laatumallin (EFQM tai CAF) mukaisia kriteerejä arviointiohjeineen saatetaan päätyä kyseisten mallien oikeoppiseen toteutukseen, mutta samalla jättää huomioimatta edellä mainittuihin lähestymistapoihin liittyvä arvottaminen. Malleilla on taipumus myös jakaa arviointikohde osiin, jolloin uudelleen kokonaisuutta hahmotettaessa on vaara, ettei kokonaisuudelle ja arviointiobjektille tehdä täyttä oikeutta (Atjonen 2007, 231).

Shadishin ja hänen tutkijakollegoidensa (1991, 47–48) sekä Virtasen (2004, 19) mukaan ei mikään yllä mainituista eettisistä lähestymistavoista ole riittä-

vä. Kyseisten tutkijoiden mukaan hyvä arvoteoria olisikin mainittujen kolmen teorian yhdistelmä, joka ottaisi mukaan niiden parhaat elementit ja jättäisi pois heikoimmat osat.

Arvojen arviointi

Arviointia ei voi koskaan olla ilman arvoja (Scriven 1997; Virtanen 2004, 20; Vuori 2004, 1–3). Yleensä arvoista todetaan, että ne ovat syvästi tajunnallisia. Ne ovat myös näkymättömiä. Toisaalta ne ilmenevät mielipiteissämme, käsityksissämme, motiiveissamme, pyrkimyksissämme, ajankäytössämme, huolenaiheissamme, kokemissamme ongelmissa, suhteissamme aineellisiin ja aineettomiin asioihin, ihanteissamme ja päiväunelmissamme jne. (Lyytinen & Nikkanen 1996, 126). Edellä mainitut ovat siten tavallaan ”arvoindikaattoreita”. Ne ilmentävät edellä mainittujen käyttäytymispiirteiden taustalla olevia arvoja, jotka eivät näy, eivät kuulu eivätkä suoraan ole kosketeltavissa. Ne ilmaisevat, mitä pidämme arvokkaana tai hyvänä. Arvoindikaattorit auttavat tunnistamaan usein suljetun verhon takana olevia arvoja, jotka eivät ole esineellisiä ja konkreettisia. Jos arvot ovat subjektiivisia, ne eivät ole ainakaan puhtaasti kohteen ominaisuuksia, vaan arvojen kokijasta ja kantajasta johtuvia.

Jos arvot ovat subjektiivisia ja tajunnallisia ja siten kokijasta riippuvia, miten ne sitten ovat tunnistettavissa, arvioitavissa ja kenties saatettavissa hierarkkiseen järjestykseen? Suoraa oikotietä arvojen tunnistamiseen ei liene. Subjektiiviset arvot ovat tunnistettavissa ihmisten ajattelun ja toiminnan myötä sikäli kuin niissä on johdonmukaisuutta. Välttämättä ei ole niin, että ihmiset pitäytyisivät arvokkaina ja oikeina pitämässään toimintatavoissa (Häyry & Häyry 1988, 33). Sen havainnointi ja analyysi, mitä ja miten ihmiset ajattelevat ja mitä ja miten vastaavasti tekevät, on joka tapauksessa arvojen tunnistamisen keino ja lähde. Periaatteessa se voi tapahtua systemaattisesti tutkimalla tai työyhteisöissä ja ryhmissä (arviointiryhmissä) käytävänä arvoarviointina. Tällaisen arvoarvioinnin lopputuloksena voi syventyneen arvotietoisuuden lisäksi olla konkreettisenä tuotoksena dokumentoitu listaus arviointitoimintaa ohjaavista arvoista tai niiden tärkeysjärjestyksestä (Lyytinen & Nikkanen 1996, 129).

Arvokeskustelu, jota voi kutsua myös arvoarvioinniksi, on arvojen tunnistamisen ohella myös keino arvojen arviointiin. Mikä tahansa ”ilmava” arvoihin liittyvä keskustelu ei ole arvokeskustelua. Arvokeskustelun tulisi olla rationaalista. Täyttääkseen rationaalisuuden vaatimuksen keskustelun tulisi täyttää seuraavat ehdot: a) arvokeskusteluun osallistuvien omat uskomukset maailmasta ovat keskenään ristiriidattomia, b) kunkin arvoalinnat perustuvat näille uskomuksille, c) kyseiset uskomukset perustuvat tosiasioihin, d) arvojen asettaminen tärkeysjärjestykseen tapahtuu itse kunkin oman vapaan tahdon mukaan eikä

ulkopuolisesta pakosta tai henkisestä paineesta sekä e) itse kukin voi hyväksyä arvojärjestelmänsä kokonaisuudessaan, mikä tuo rationaalisuuteen moraalisen näkökulman (Häyry & Häyry 1988, 38).

Mainitut ehdot ovat arkikeskusteluun verrattuna erityisen vaativia ja edellyttävät vahvaa arvotietoisuutta. Häyry ja Häyry (1998, 38) tarjoavat myös filosofiassa käytetyn toisen tien arvojen yhteiseen tarkasteluun, jolloin tietoiset ihanteet otetaan tarkastelun kohteeksi, vaikka siihen sisältyykin epärealistisuutta. Kysymys kuuluu, millaisten ihanteiden mukaan tulisi toimia, jos toimittaisiin rationaalisesti ja vallitsevaa tietoa soveltaen. Vastaavasti voidaan kysyä, mitkä teoreettiset ideaalit voitaisiin ottaa toiminnan arvottamisessa lähtökohdiksi.

Erityisesti verkostomaisessa arvioinnissa, jossa arviointikenttä on moniarvoinen, arvojen yhteinen arviointi on tärkeää. Arviointiryhmien koostuminen tutkijoista, hallintoviranomaisista ja monia eri tahoja edustavista toimijoista tuo arviointiin moniäänisyyttä ja -arvoisuutta. Tätä ei välttämättä kuitenkaan tiedosteta eikä oteta dialogiseen käsittelyyn. Siksi arvolauseiden muotoilu suhteessa arviointikohteeseen voi arviointiryhmässä olla hyvin problemaattista, mutta erityisen tarpeellista yhteisten arviointikriteerien ja yhteisen arviointitotuuden määrittämiseksi. Mitkä ovat arviointikriteerit, jotka täyttävät hyvyyden vaatimukset ja millaisiin arvoihin ne sidostuvat? Korostuuko hyvyyden kriteerinä esimerkiksi toiminnan hyöty tuottavuuden kannalta vai koulutuksen tasa-arvo opiskelijoiden näkökulmasta? Jos itse kullekin arviointiryhmän jäsenelle yhteisen arvokeskustelun tuloksena syntyvät ”implisiittiset standardit” omaan arviointitietoisuuteen, voidaan olettaa, että arviointikohteen arvo suhteessa sovittuihin standardeihin on pätevämmän määritettävissä kuin tilanteessa, jossa standardista ei ole muodostunut selkeää käsitteähahmoa.

Niin sanottu arviointitotuus ei ole koskaan yksiselitteinen. Hyvyyden määrittely on aina kontekstisidonnaista. Barnes, Bloor ja Henry (1996) katsovat, että havaintojen tekeminen jonkin toiminnan hyvyydestä tai huonoudesta on aina kontekstuaalista ja indeksikaalista (vrt. Harrinvirta, Uusikylä & Virtanen 1998). Arviointikohteesta tehtäviin havaintoihin on yhteydessä ja mahdollisesti vaikuttamassa aina runsas joukko eri tekijöitä. Havaintoja tekevä arvioitsija omine taustasitoumuksineen, arvoineen, asenteineen, ihanteineen, käsityksineen ja mielipiteineen liittyy keskeisesti tulkintojen konstruoimiseen. Jokaisella arviointisijalla ja tutkijalla on oma yksilöllinen totuutensa, joka nousee yksilön itse konstruoimasta tiedosta, ajatusrakennelmista, perusolettamuksiin liittyvästä tiedosta. Muun muassa siksi arviointitulokset eivät ole aina yksiselitteisiä.

Arviointi on aina joissakin suhteissa rajallista, subjektiivista (Harrinvirta ym. 1998). Tästä rajallisuudesta taas seuraa, että arvioinnin arvotaustaan tulee kiinnittää erityistä huomiota. Arvioinnin eettiset kysymykset nousevat keskiöön. Arvioijan on kyettävä ottamaan etäisyyttä omiin arviointeihinsa ja arviointikriteerien taustalla oleviin arvoihin. Tähän liittyy Vuori (2004) toteakin: ”Nykyinen

arviointibuumi tuottaa enemmän tietoa, jolla hoidetaan oireita, mutta ei kyetä ja uskalleta arvottaa mahdollisia syitä.” Hän puhuu lisäksi arvioinnin ”pyhästä lehmästä”. Sellaiseksi Vuori näkee ”...arviointistandardit, joiden laaja levinneisyys on saanut niiden luojat, tutkijoita myöten, unohtamaan tieteen ydinongelman: arviointitotuus on aina hieman relativistista, ja standardit eivät ole arvovapaita”. Vuori on epäilevä myös sen suhteen, että monitoimijainen tai -tahoinen arviointi voisi paljastaa arvoja: ”...voivathan kaikki toimijat olla samalla tavoin indoktrinoituja: tilattu arviointi on ihan hyvä, jos tilaaja saa, mitä haluaa”. Vuoren (2004, 2) mukaan ”todellinen kasvava ongelma siis on, että arviointi perustuu tai peilaa arvoja, mutta näitä arvoja ei perustella ja kyseenalaisteta arvioinnin lähtökohdissa tai edes päätelmissä. Arviointitutkimuksen suurin haaste on edelleen arvojen analyysin sivuutus”. Tutkijoiden ja arvioitsijoiden tulisi rehellisyyden nimissä tiedostaa ja ilmaista, minkä taustasitoumuksen varassa he toimivat. Jo tieteen tekemisen periaatteet (pelisäännöt) edellyttävät sitä (ks. esim. Hirsjärvi, Liikanen, Remes & Sajavaara 1986; Karma 1983; Ketonen 1976; Niiniluoto 1980).

Voidaan toisaalta kysyä, missä määrin arvojen näkyväksi tekeminen, niistä puhuminen, niiden kyseenalaistaminen tai analysointi on omiaan tuottamaan jotain konkreettista ja käytännöllistä. Saattaahan käydä niinkin, että arvoarviointi jää vain jaloksi, kohottavaksi ja arviointitietoisuutta syventäväksi, mutta lopuksi kaikki konkreettisuus ikään kuin haihtuu ilmaan. Tuottaessaan eri näkökulmia, ohjatessaan yhteistä ajattelua ja tuodessaan esille erilaisia vaihtoehtoja sekä luodessaan perustaa valintojen tekemiselle arvoarviointi on sinänsä kuitenkin jo tehnyt tehtävänsä.

Arvottaminen arvioinnin ydinfunktiona

Arvioinnin arvottavuus

Arviointi määritellään arvon antamiseksi arviointikohteelle. Arvioinnin ydintehävä on siten arvottaminen kulloinkin valittujen kriteerien suhteen. Arvottaminen on arvioinnin kivijalka. Kun arviointikohteelle annetaan arvo, se merkitsee asiantilan suhteuttamista johonkin arvokkaana tai parempana pidettävään asiantilaan.

Peruslogiikka on siis seuraava: todetaan, että x on hyvä tai että x on parempi kuin y, sekä perustellaan esitetyt väitteet. Tällöin arvottaminen on tosiasioiden ja arvostusten analysointia ja tulkintaa, jota ohjaavat aina myös arvioijan arvot. Siihen voi sisällyttää myös arvojen problematisoinnin ja priorisoinnin. Arviointikohteen tilaa verrataan siten johonkin toivottavaan tai tavoitteeksi asetettuun asiantilaan. Arvottaminen merkitsee, että mukaan tulee hierarkkista ainesta, jolla ilmaistaan suosimista, arvonantoa, hyväksyvää myöntämistä tai puoltamista

tai vastaavasti edellä mainittujen vastakohtia (Ricoeur 2005). Arvottamisessa ei vielä asemoida arviointikohdetta. Tässä suhteessa arvostelussa mennään pitemmälle eli omaksutaan positio (vrt. Ricoeur 2005), jonka jälkeen epävarmuus tavallaan loppuu.

Mäntysaari (1999, 64) toteaa Scriveniä (1997, 8–9) lainaten, että arviointi sidostuu neljään peruspredikaattiin, joista vähintään yhden on oltava mukana arviointia koskevassa johtopäätöksessä. Nämä predikaatit ovat: luokittelu/arvostelu (grading), keskinäiseen arvojärjestykseen asettaminen (ranking), pisteytys (scoring) sekä arvottaminen ansion mukaan (apportioning). Viimeksi mainittu tarkoittaa käytännössä esimerkiksi sitä, että tiettyjen kriteerien perusteella määritetään arviointikohteen vahvuudet ja kehittämistarpeet. Pisteytys ja arvojärjestykseen paneminen liittyvät kiinteästi toisiinsa. Päätarkoitus on antaa arvioinnissa arvo arviointikohteelle noudattamalla arvioinnin asianmukaisia pelisääntöjä

Vierasperäinen arvioinnin termi evaluatio on arvottamisen merkitysisällön kannalta mielenkiintoinen. Sen kantasana on sanskritin bala, jonka merkitys on olla voimakas, voida hyvin, olla jonkin arvoinen, merkitä jotakin taloudellisessa merkityksessä eli esimerkiksi rahassa. Siitä tulee vastaavasti latinan kielen samaa merkitsevä valere-verbi, joka esiintyy myös muodossa evalere, johon esimerkiksi englannin kielen verbi to evaluate ja ranskan valoir pohjautuvat. Samakantainen on myös verbi evalescere, joka merkitsee voimaantumista (vrt. voimaannuttava arviointi; Lyytinen & Nikkanen 2005). Kantasanan merkitys 'merkitä jotakin rahassa' viittaa absoluuttiseen arvottamiseen, jolloin arvo saa merkityksen 'hinta' (vrt. viron hindamine). Arviointi-käsitteen merkitysisällössä ei ole siis mitään uutta suhteessa sen etymologiseen taustaan. Merkitä jotakin, vaikka rahassa, merkitsee juuri arvottamista. Liiketaloudessa arvo on tunnusluku, joka kuvaa taloudellista merkitystä. Tässä yhteydessä arvon ontologinen status on merkityksellinen. Myyjän näkökulmasta tuotteen tai palvelun arvo on hinta, jonka hän saa tuotteista ja palveluista markkinoilla (markkina-arvo), kun taas ostajan näkökulmasta tuotteen ja palvelun arvo on sen käyttöarvo. Hyötyä kuvataan käyttöarvolla ja ostovoimaa taas vaihtoarvolla (Oksanen 2000; vrt. Smith 1776). Hyötyä tuottavalle on vakiintunut taloudelliseen sanastoon suomennos hyödyke ('a good').

Pohjimmiltaan arvottamisessa on kysymys valintojen tekemisestä tiettyjen arvojen perusteella, kun punnitaan esimerkiksi eri toimintavaihtoehtojen perusteltavuutta tai hyvyyttä. Arvot liittyvät muun muassa valintoihin, vakaumukseen, päämääriin, yhteiseen tahtotilaan, tarpeisiin – siis lähes mihin tahansa haluttavaan, tavoiteltavaan tai hyvään. Arvoilla on selkeä status arvottamisessa, kun sitä ei yleensä objektiivisuuteen pyrkivässä tieteessä ole (vrt. Ketonen 1976). Atjonen (2007) toteaa Heinosen (2002) viitaten, ettei arvioija yksinkertaisesti arvioi mitään, ellei hän päädy erilaisista arvoista johtamiensa kriteerien ja/tai

standardien perusteella arvottaviin johtopäätöksiin kohteen hyödystä, ansioista ja merkittävydestä. Asioiden arvottaminen on välttämätön osa arviointiprosessia, jossa tuotetaan perusta merkityksille (Guba & Lincoln 1989, 109).

Arvioinnin erottaa tutkimuksesta sen arvottava luonne. Suhteutettaessa arvon määrittäminen erilaisiin kriteereihin kriteerien valinta on sinänsä arvoperusteinen. Jotta arvottaminen olisi mahdollista, tarvitaan arvot ja arvostukset, joiden suhteen ilmiötä arvotetaan (Patton 1997; Owen 1993, 30; vrt. Viinamäki 2004). Arvoperusteisuus näkyy esimerkiksi laatukriteereissä ja siinä, mitä niissä kulloinkin painotetaan ja mitä niistä priorisoidaan.

Arviointikriteerien valinta sinänsä on jo arvottamista. Ovatko kriteerit tai niiden suuri määrä esimerkiksi arviointikohteen kannalta tukevia, tukahduttavia tai kenties jopa tuhoavia, jos niiden taustalla ovat toiminnan tehokkuutta korostavat arvot? Mitkä ovat kriteerien valinnan seuraukset esimerkiksi arviointikohteen tulevaisuuden kannalta? Arviointikohteen vangitseminen ahtailla kriteereillä ja toiminnan malliversioilla voi olla kehittämismotivaation kannalta kyseenalaista, varsinkin jos ei tiedetä tai ymmärretä, mihin ne perustuvat tai mitä arvoja ne edustavat. Niiden pelkkä auktoriteettiasema ei ole riittävä perusta arvottamiselle. Kriteereillä on joka tapauksessa myös legitimoiva vaikutus, jos kriteerit hyväksytään yleisesti. Uuden todellisuuskonstruktion taustalla ovat siten arvioija tai arviointiryhmän arvot. Arvotettaessa käytetään siten myös valtaa, varsinkin jos kilpailevia kriteerejä ei ole.

Kriteerien avulla todellisuus konstruoidaan arvottamisessa uudelleen ja luodaan siten uutta todellisuutta, jonka taustalla ovat arvioijan arvot. Arviointikriteerien avulla saadaan esille esimerkiksi niin sanottuja hyviä käytäntöjä, joiden ajatellaan toimivan todellisuutta uudistavina keinoina. Käytäntöjen hyvyys siidostuu arvioijan arvoihin.

Suhteessa kriteereihin voidaan toiminta todeta onnistuneeksi, epäonnistuneeksi, erinomaiseksi, hyväksi, tyydyttäväksi jne. Käytettäessä useita kriteerejä voidaan päätyä niiden perusteella toiminnan kokonaisuuden arvottamiseen useasta näkökulmasta. Tällöin arviointikohde voi saada tietystä näkökulmasta erinomaisen ja toisesta näkökulmasta huonon arviointituloksen tai eri näkökulmat yhdistäen kokonaisuutena hyvän arviointituloksen.

Shadishin (1991, 47–50) mukaan arvottaminen edellyttää teoriaperustaa, josta vallitsevat arvot ja arvostukset yksittäisessä arvioinnissa ovat johdettavissa ja johon ne ovat palautettavissa arviointitulosten ja kehittämissuosituksen yhteydessä. Niin ikään tulisi luoda teoriaperusta, jonka mukaan arvottaminen ja arvostukset priorisoidaan (ks. myös Raunio 1987, 1; Viinamäki 2004). Housen (1993; vrt. Viinamäki 2004) mukaan arvottamisen sitominen filosofiassa analysoituihin arvostuksiin, joita ovat esimerkiksi teoreettis-filosofiset tarkastelut tasa-arvosta tai oikeudenmukaisuudesta (esim. Rawls, 1972, teoria oikeudenmukaisuudesta) on yksi tapa arvostusten priorisoinnissa. Raunio (1987, 58) esittää

edellä mainittua Rawlsia lainaten oikeudenmukaisuuden yleisperiaatteen, jonka mukaan kaikki sosiaaliset primaariset hyvät – vapaudet, mahdollisuudet, tulot ja varallisuus sekä itsekunnioituksen perusta – tulee jakaa tasan, ellei epätasainen jakautuminen ole huonoimmassa asemassa olevien eduksi. Rawls (1972) pitää tasa-arvoa tärkeänä sosiaalisena periaatteena, joka on oleellisempi kuin tehokkuus ja hyvinvoinnin kokonaismäärän maksimointi. Tässä yhteydessä arvioitavasta ilmiöstä riippuu, miten hyvin filosofian jäsentämät periaatteet kyseessä olevaan arviointikohteeseen sopivat. Joissakin tapauksissa esimerkiksi oikeudenmukaisuus tai tasa-arvo ovat sinänsä arvoja, joiden suhteen arvioitavaa ilmiötä voi olla tarkoituskin arvottaa.

Edellä on viitattu esimerkkinä koulutuksen tasa-arvoon arvona. Tasa-arvon käsite on arvioinnin kannalta siinä mielessä pulmallinen, että se on monimerkityksinen. Tällä hetkellä koulutuspolitiikassamme korostetaan mahdollisuuksien tasa-arvoa, jolla tarkoitetaan tasa-arvoisia mahdollisuuksia koulutukseen ja tuen saamiseen. Tasa-arvo voi tarkoittaa myös ihmisten toiminnan samanlaisia tuloksia (esimerkiksi oppimistuloksia) tai samanlaista kohtelua johonkin etuun, haittaan tai normiin nähden, jolloin puhutaan relationaalisesta tasa-arvosta. Vastaa- vasti voidaan puhua juridisesta tasa-arvosta eli ihmisten samanlaisesta asemasta suhteessa lain sisältöön tai vaikkapa ihmisten samanlaisista ominaisuuksista jollakin rajatulla alueella eli ontologisesta tasa-arvosta (Launonen 2007).

Viime aikoina tasa-arvoa on tarkasteltu myös tehokkuuden näkökulmasta. Suppeasti nähtynä tasa-arvo voi näyttää tehokkuuden esteeltä edellyttäessään voimavarojen sijoittamista myös ei-lahjakkaisiin. Laajemmasta näkökulmasta tasa-arvo kuitenkin voi lisätä yhteiskunnan suorituskykyä. Se mahdollistaa yhteiskunnan voimavarojen käyttöönoton ja estää kalliiksi koituvaa syrjäytymistä sekä syrjäytymisen edellyttämiä palveluja ja tulonsiirtoja (Lampinen 2006). Tällöin tasa-arvo voi olla myös välinearvo, jolloin varsinaisena päämääränä onkin kilpailukyvyyn vahvistaminen.

Lisäksi arvottaminen Shadishin (1991) mukaan edellyttää intressiryhmien arvostusten huomioimista sekä arviointitulosten suhteuttamista intressiryhmien arvostuksiin (ks. myös Vartiainen 2003; Viinamäki 2004; Temmes 2004). Näin on perusteltua erityisesti kehittävässä arvioinnissa, jolloin esimerkiksi arviointikohteen piirissä tapahtuvan itsearviointin taustalla olevat arvot ja ulkopuolisen arvioinnin arvot eivät synnytä sellaista arvokonfliktia, joka lamaannuttaa kehittämisen.

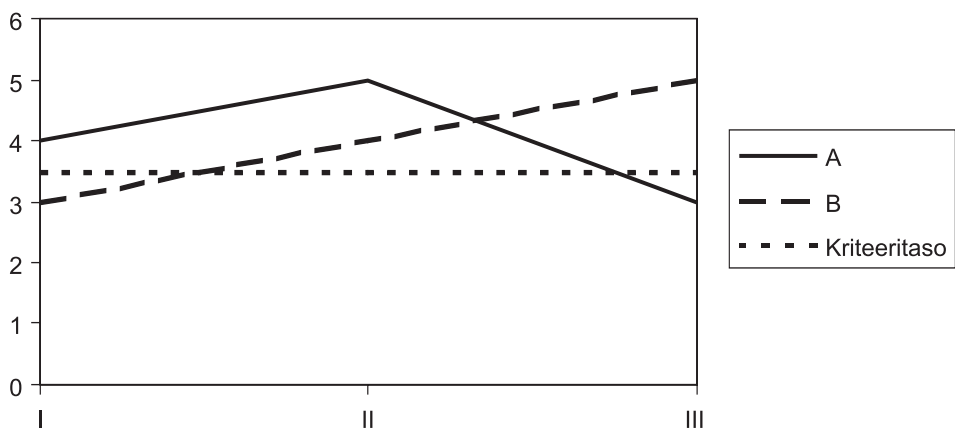
Arviointitieto ei ole koskaan arvoneutraalia, vaan arviointitiedon käyttö on suhteutettava vallitseviin kulttuurisiin, sosiaalisiin tai taloudellisiin arvostuksiin (Linnakylä & Välijärvi 2005, 16). Arviointi ei siten ole arvottamisen lähtökohdiltaan itsetarkoituksellista, vaan se liittyy yhteiskunnallisiin päämääriin, jotka ilmenevät koulutuksen osalta koulutuspoliittisissa tavoitteissa, kuten esimerkiksi koulutuksen tasa-arvon tai elinikäisen oppimisen edistäminen. Arvioija sankari-

na, joka tietää kaiken, unohtaa dialogisuuden ja objektivoi ja mallintaa arviointikohteensa sitä kunnioittamatta, ei edistä kehittämistä eikä yhteiskunnallisia päämääriä (Eräsaari 1999).

Arvottamisen absoluuttisuus ja suhteellisuus

Periaatteessa arvottaminen voi olla absoluuttista tai suhteellista. Absoluuttisessa arvottamisessa arvotettavaan kohteeseen liitetään tai sille johdetaan jokin arvon mitta. Esimerkiksi voimavarojen arvottamisessa mittana käytetään rahaa. Tällöin voidaan määrittää, kuinka paljon parempi tai huonompi asiantila on. Koulutuksen arvioinnissa ei absoluuttinen arviointi useinkaan ole mahdollista, koska arviointikohteen mittauksen taso ei ole riittävä. Koulutusilmiöt ovat hyvin usein mittaustasoltaan järjestysasteikkollisia. Arvioitaessa koulutuksen tuottavuutta tai taloudellisuutta tietyistä näkökulmista voidaan päästä absoluuttiseenkin arvottamiseen.

Seuraavassa havainnollistetaan näitä arvottamistapoja Sintosen (1987, 109; vrt. Sinkkonen & Kinnunen 1994, 75) esittämällä kuviolla ja tämän artikkelin kirjoittajien esimerkillä. Kuvatkoon yhtenäinen viiva A oppilaitoksen aluevaikutuksen tilaa kolmen kriteerin (I–III) ja niitä koskevien asteikkojen suhteen. Kriteerejä voisivat olla esimerkiksi työhön sijoittuminen, koulutuksen taso ja aluekehitysinnovaatioiden määrä. Vastaavasti katkoviiva B kuvaa toisen oppilaitoksen aluevaikutuksen tilaa.



Kuvio 1. Kahden oppilaitoksen aluevaikutuksen tila tiettyjen kriteerien suhteen

Suhteellisessa arvottamisessa määritetään, onko oppilaitoksen A aluevaikutus parempi tai huonompi kuin B:n aluevaikutus. Absoluuttisessa arvottamisessa taas määritetään, paljonko A:n tuottama asiantila on B:n tuottamaa parempi tai

huonompi. Arvottaminen voidaan tehdä kummallakin tavalla riippuen siitä, millä mittauksen tasolla arvottamisen kohde on mitattavissa.

Kriteereille voidaan määrittää myös ko. kriteeriominaisuuden määrää osoittavat standardit, jotka voivat olla joko aineisto- tai teorialähtöisiä tai niiden yhdistelmiä ja joiden avulla arvioinnin tasot määritetään. Aineistolähtöisessä arvioinnissa on mahdollista määrittää esimerkiksi tyydyttävä, hyvä tai erinomainen taso käyttämällä hyväksi havaintoaineiston tulosten jakaumaa. Hyvä taso standardina voi olla arviointiryhmän sopimus, johon ko. ryhmä on ”arviointiviisautensa” päättänyt. Usein vaadittava taso – esimerkiksi riittävän hyvä – määritetään sen jälkeen, kun arviointikohde on tutkittu. Hyväksi voidaan sopia esimerkiksi havaintoaineiston keskiarvo tai pistemäärä, jonka tietty suhteellinen osuus arvioitavista on saavuttanut. Jos käytössä on useita kriteerejä, tiettyjä kriteerejä on mahdollista painottaa niiden tärkeyden mukaan.

Suhteellinen arvottaminen tulee varsin usein kysymykseen koulutusta koskevissa arvioinneissa. Arviointikohteiden keskinäinen vertailu on suhteellista, koska mittauksen taso on useimmiten järjestysasteikollinen.

Smith ja Glass (1987, 29–32) ovat esittäneet yleisimpiä kriteerejä tai arvoulottuvuuksia, joiden avulla voidaan määritellä arvopäätöksiä. Niitä ovat:

- **Vaikuttavuus** (effectiveness): Vaikuttavuus on (uudistus-) ohjelman tai tuotoksen ensimmäinen ja ilmeisin kriteeri. Arvioitsija kysyy: kuinka hyvin (uudistus-) ohjelma saa aikaan sen, mitä oli aiottukin? Hän saa vastauksen katsomalla tutkimusdataa, joka havainnollistaa tavoitteiden saavuttamista. Esimerkiksi sellaisissa (uudistus-) ohjelmissa, joissa on asetetut päämäärät ja tavoitteet, arvioitsija saattaisi määritellä ne osallistujat, jotka saavuttivat tavoitteet.
- **Tehokkuus** (efficiency): Arvioitsija kerää dataa vastatakseen kysymykseen, kuinka paljon (uudistus-) ohjelma maksaa suhteessa siihen, mitä sillä saavutetaan. Tehokkuuden standardit voivat olla joko absoluuttisia tai suhteellisia.
- **Arvojen loukkaamattomuus tai oikeudenmukaisuus** (fairness or justice): Arvioitsija kerää dataa vastatakseen kysymykseen, miten yhteiskunta yleensä hyötyy tällaisten (uudistus-) ohjelmien olemassaolosta. Voiko joillekin yhteiskunnan alueille koitua vahinkoa samalla, kun toiset alueet hyötyvät? Siinä määrin kuin vahinkoa aiheutuu, siinä määrin (uudistus-) ohjelma epäonnistuu tyydyttämään loukkaamattomuuden kriteerin.
- **Hyväksyttävyyys** (acceptability): Arvioitsija kerää dataa, joka ilmentää osallistujien tyytyväisyyttä (uudistus-) ohjelmaan tai tuotokseen.
- **Esteettiset seikat** (aesthetics) liittyvät (uudistus-) ohjelman tai tuotteen laadullisiin ominaisuuksiin kuten eheys, sopusointuisuus ja yhtenäisyys.

Smith ja Glass (1987) kysyvät vielä, kuka tekee päätökset. Kuka tekee arvioinnin? Arvioinnin teoreetikot eroavat toisistaan näiden kysymysten suhteen. Heidän mukaansa jotkut teoreetikot vakuuttavat, että itse arvioitsija, joka tuntee parhaiten todistusaineiston, on parhaassa asemassa päätöksen tekoa ajatellen. Toiset vuorostaan vakuuttavat, että arvioitsija on yksinkertaisesti objektiivinen, neutraali ja teknisesti taitava henkilö, joka antaa todistusaineiston ”päätöksentekijälle”, joka sitten tekee päätöksen. Edelleen, muut teoreetikot ovat sitä mieltä, että todistusaineisto pitää olla kaikkien niiden osapuolten saatavilla, joita ohjelma koskee. He tähdentävät, että monimuotoiset ryhmät tekevät arvopäätelmät sen mukaisesti, mikä on heidän oma intressinsä sekä mitkä ovat heille kaikkein tärkeimpiä standardeja. (Emt., 31–32.)

Arvioinnin meta-arviointi

Arvioinnin laadun kriteereistä

Arvioinnin arvottaminen on arvioinnin laadun arviointia eli meta-arviointia. Se merkitsee arvioinnin vahvuuksien ja heikkouksien arviointia käyttäen itsearviointia tai ulkopuolista arviointia. Käytännössä siinä tutkitaan, miten arviointi on toteutettu ja mikä on sen lopputuloksen laatu. Virtanen (2007, 221) määrittelee meta-arvioinnin toimenpiteeksi, jolla varmistutaan arvioinnin laadusta valitun kriteerin tai kriteeristön suhteen. Virtanen on taulukoinut esimerkkejä meta-arvioinnin käsitteen määrittelemisestä. Hänen mukaansa erilaiset meta-arvioinnin variaatiot määrittyvät sen perusteella, mikä on meta-arvioinnin pääasiallinen sisältö ja tavoite, kuka sen tekee ja millaisia kriteerejä siinä hyödynnetään (vrt. edellä Smith & Glass 1987, 29–33). Sisältönä voi Virtasen mukaan olla a) arvioinnin laadunvarmistamisen ja hyödynnettävyyden varmentamisprosessi, b) arvioinnin laadunvarmistaminen standardilähtöisesti, c) arviointitulosten validiteetin ja reliabiliteetin varmistaminen, d) arviointiprosessin ja arvioinnin tulosten auditointi sekä e) arviointien laadun ja yhdenmukaisuuden analyysi. Virtasen edellä mainitusta taulukosta ilmenee useita tahoja, jotka voivat tehdä meta-arvioinnin. ”Pääsääntöisesti meta-evaluaation tekee jokin varsinaisesta arvioinnista riippumaton taho, jolloin kyse on tavallaan kolmannen tahon tekemästä laatuvarmistuksesta. Kun arvioinnin tilaaja tilaa arvioinnin ja arvioinnin tekijä tekee sen, lisäksi tarvitaan kolmas taho, joka varmistaa viime kädessä riippumattomasti tehdyn arviointiprosessin ja sen myötä tuotetun arviointiraportin laadun. Käytännössä ammattitaitoiset arvioinnin tekijät tekevät jatkuvaa meta-evaluaatiota koko arviointiprosessin ajan, joskus tiedostamatta ja joskus hyvin tietoisesti. Varsin usein arvioinnin tekijä tekee meta-evaluaation itsearviointina, kun arviointiraportti on valmistunut.” (Virtanen 2007, 223.)

Arvioinnin arvottamisessa on kyse asioiden priorisoinnista, joka perustuu paitsi valittuihin kriteereihin, myös usein siihen, mitä henkilö arvostaa omista lähtökohdistaan, paradigmoistaan, lähtien. Vaihtoehtoiset paradigmat ohjaavat niin arvioitsijoita kuin tutkijoitakin. Paradigma on kuin järjestelevä viitekehys, joka sisältää tarkastelijan (arvioitsijan) taustalla vaikuttavien uskomusten ja olettamusten asetelman. Se on kuin linssi, jonka läpi tarkastelija (arvioitsija) tutkailee maailmaa (ks. Smith & Glass 1987, 38). Arvioinnin metodologit eivät ole yksimielisiä siitä, mikä olisi oikea tapa arvioida. On olemassa useita erilaisia esityksiä ja analyyseja arvioinnin lähestymistavoista ja siitä, mitä arviointi on tai mitä sen pitäisi olla (ks. esim. Aaltonen 1989; House 1980; Laukkanen 1994; Smith & Glass 1987; Vaherva 1983; Nikkanen & Lyytinen 1996, 97–98).

Paradigmojen moninaisuudesta muistuttaen Smith ja Glass (1987, 49) ovat antaneet meta-arvioinnille joukon kriteerejä. Seuraavat kriteerit on sovitettu mukaelleen arviointitoimintaan:

- **Tekninen riittävyys:** Onko arviointimalli osoittanut vakuuttavasti, että havaitut vaikutukset aiheutuivat esimerkiksi toteutetusta uudistusohjelmasta? Oliko mittaaminen tai havaintoaineiston koonti validia ja reliabelia?
- **Kuvailun riittävyys:** Kuinka hyvin arviointisuunnitelma kuvattiin? Kuvattiinko konteksti, arviointiasetelma ja otos tai näyte hyvin?
- **Kuvauksen piiri:** Tarkistettiin kaikki arviointikohteen näkökulmat? Otettiin ohjelman kaikki kustannukset ja vaikutukset, kaikki näkökulmat ja ryhmien intressit huomioon arvioinnissa?
- **Ajotus:** Raportoitiinko määrääjassa niin, että raporttia voitiin käyttää päätöksenteon apuna?
- **Kustannusten vaikuttavuus:** Saavutettiin enimmäismäärä tähdellistä tietoa kohtuullisilla kustannuksilla?
- **Kommunikointi:** Olivatko raportin ominaisuudet sellaisia, että kaikki osapuolet kokivat raportin ymmärrettäväksi, uskottavaksi ja käyttökelpoiseksi?
- **Hyödyllisyys:** Missä määrin rahoittajat valistuivat raportista niin, että he voisivat vielä tehokkaammin olla poliittisen toiminnan avuksi? Missä määrin raportti hyödytti eri toimijatahoja käytännössä? (Emt., 1987.)

Listaa voisi täydentää vielä arvioitsijoiden **asiantuntijuuden arvioinnilla**. Käytettiinkö arvioitsijoiden pätevyyskriteereitä ja miten relevantteja kriteerit olivat suhteessa ko. arviointiin? Miten kattava arvioitsijoiden asiantuntijuus oli arviointikohteen kannalta?

Arvioinnissa, kuten myös tieteellisessä tutkimuksessa, voidaan varmistua saadusta tuloksesta käyttämällä triangulaatio-menetelmää. Siinä valitun me-

netelmän ohella tai jälkeinpäin käytetään toista menetelmää. Jos molemmilla menetelmillä päädytään samanlaiseen lopputulokseen, saa arvioitsija tai tutkija suuremman varmuuden tuloksen oikeellisuudesta.

Perusarvot arvioinnin laadun arvottamisessa

Kuten jo todettiin, arvottamisessa on perusteltua soveltaa edellä mainittuja teoreettisia arvopohdintoja. Tässä yhteydessä analysoidaan meta-arvioinnin näkökulmasta esimerkkinä platonisia perusarvoja arvioinnin laadun arvoina. Niiden voidaan katsoa olevan myös useiden edellä todettujen kriteerien taustalla olevia perusarvoja.

Klassiset perusarvot totuus, kauneus ja hyvyys edustavat tiedollisia, eettisiä ja esteettisiä arvoja. Totuus (ant. kr. *aletheia* 'totuus, unelman toteutuminen, vilpittömyys, rehellisyys') arvona liittyy erityisesti tieteen ihanteisiin. Totuus ilmentää pyrkimystämme tietopääoman ja ajattelun kehittämiseen. Tietopääoman kartuttamiseen liittyy kriittisyys ja valikoivuus. On pystyttävä analysoimaan tietoa ja havaitsemaan siinä loogisia ristiriitoja, testaamaan tiedon totuusarvoa ja valikoimaan oleellinen tieto epäolennaisesta. Kysymys on todellisuuden hahmottamisesta ja jäsentämisestä, jossa on tiedostettava myös oman ajan katsomuksellinen moniarvoisuus ja oman maailmankuvan ja ihmiskäsityksen olemassaolo. Totuuteen liittyy olennaisesti perusteltavuus. Oma tiedon valinta on kyettävä perustelemaan, mutta samalla on oltava avoin tiedon (totuuden) itseohjautuvuudelle. Edellä mainitut totuusarvon luonnehdinnat ovat käytännön arvioinnin arkea. Totuus voidaan katsoa näin määritellen myös arvioinnin arvoksi, joka on mukana jokaisessa arvioinnissa. Arvioinnissa tavoitellaan totuutta havaintotietoja kokoamalla, tekemällä näkyväksi toiminnan sisältöjä ja niiden ominaispiirteitä, vertailemalla arviointikohdetta koskevia tietoja keskenään ja toiminnalle asetettuihin tavoitteisiin, etsimällä tiedolle tulkitsemalla merkityksiä, ymmärtämällä ja oppimalla kulloinkin arvioitavan ilmiön olemusta jne. Edistäessään oppimista arviointi palvelee tavallaan uuden totuuden löytämistä.

Kauneus (ant. kr. *kalon*, 'moraalinen kauneus, hyve, kunnioitettavuus') esteettisenä arvona liittyy kokonaisuuksien hahmottamiseen (vrt. ant. kr. *aisthanomai*, 'huomata, havaita, aistia'). Siihen kuuluu kyky järjestää havaintoaineisto niin, että siinä tunnustetaan sekä itse kokonaisuus, mukaan lukien "pienen" ja "salatun" todellisuuden havaitseminen, että sen rakenne – osien suhteet. Tähän kuuluvat myös teoreettiset hahmotukset. Platonin (suom. 2007) mukaan tieto ja totuus ovat kauniita, ja on olemassa kaunis sinänsä.

Kauneuteen arvona liittyy niin ikään laadun arvostaminen ja valmius laatuvalintojen tekemiseen (Tuomi 1995). Kauneuteen sisältyy myös kauneudesta nauttiminen ja kyky itsenäiseen, esteettiseen tuottamiseen eli luovuuteen. Kau-

neudesta nauttimiseen liittyy myös ymmärtävä suhde ”rumaan” ja disharmoniseen.

Kauneus arvioinnissa merkitsee ehjää ja eleganttia kokonaisuutta, joka on enemmän kuin osiensa summa. Siinä toteutuu arvioinnin tavoiteltava laadukkuus. Arvioinnin filosofia, tavoitteet, toteutus ja lopputulos muodostavat harmonisen kokonaisuuden. Arvioitava todellisuus on tehty ikään kuin taiteen tavoin ja luovalla tavalla ymmärrettäväksi. Arvioinnin yleisvaikutelma vetoaa emotionaalisesti.

Hyvyys (ant. kr. *agathosyne*, ’hyvyys moraalisisessa merkityksessä, hyväntahtoisuus, jalous’) on eettinen arvo, joka ilmentää näkemystämme hyvästä elämästä ja sen perusteista. Hyvyys on erittäin monimerkityksellinen ja vaikeasti määriteltävissä oleva käsite. Siihen liittyy terve moraalitajunta: taju oikeasta ja väärästä. Niin ikään siihen kuuluvat vastuullisuus, rehellinen ja oikeudenmukainen käyttäytyminen sekä moraalinen rohkeus. Viimeksi mainittu tarkoittaa valmiutta oman vakaumuksen ilmaisemiseen ja toimintaa sen puolesta myös riskialttiissa tilanteissa sekä heikossa asemassa olevien puolustamista (Tuomi 1995).

Hyvän erilaisten merkitysten tarkasteluissa on todettu, että hyvä on subjektiivinen käsite. Toisaalta on todettu, että hyvä ei ole vain subjektiivista eikä vain objektiivista, vaan sitä kolmatta, jossa nämä kohtaavat (Ojanen 2004, 73). Aristoteles ja von Wright (2001) näkevät molemmat, että erilaisten ”hyvien” taustalla ei ole mitään yhteistä merkitystä, hyvän ideaa. von Wright on korostanut jatkuvasti, että kunkin yksilön hyvä määräytyy hänen omien rehellisten arvostustensa perusteella, eikä yksilöiden arvostuksista riippumatta päteviä arvoja ole olemassa. Nämä ristiriitaisilta vaikuttavat näkemykset voidaan ehkä nivoa yhteen, jos von Wright tulkitaan jonkinlaiseksi ”uniformaalin subjektivismin” kannattajaksi. Tällöin ajateltaisiin, että arvostusten kesken on saavutettavissa laaja yksimielisyyttä sillä edellytyksellä, että yksilöiden arvostukset ovat yleistettävissä olevia, puolueettomia, pyyteettömiä sekä hyväntahtoisesti toisten ihmisen hyvään suhtautuvia. Mutta miten nämä kriteerit ovat puolestaan oikeutettavissa (ks. Salme-la 1999)? Tällöin yksilöiden hyväksymät hyvät eheytyvät ryhmän yhteisesti hyväksymäksi kollektiiviseksi hyväksi, jolloin hyvä saa aiempaa objektiivisemmän luonteen. Tässä hyvä on merkityksessä ’hyvyys’. Voisi sanoa, että hyvä on itse asiassa jotakin arvokasta. Niin sanottu nominalistinen kanta lähtee siitä, ettei sellaisia käsitteitä kuin hyvä sinänsä ole olemassa, vaan ainoastaan yksittäisiä hyviä asioita.

Platonin (suom. 2007) mukaan hyvän idea on tiedon ja todellisuuden ylin selityspenuste. Hyvä on siis tiedon ja totuuden alkusyy. Hyvä on todellisuuden perustana oleva luova voima. Kun Platon toteaa, ettei hyvä itse ole olevaa, vaan olevaisen takana, ”yläpuolella”, (epekeina), hän viittaa siihen, että hyvä olisi korkeammassa mielessä kaiken muun luovana voimana oleva ainoa todellinen oleva. Platonin mukaan hyvä on jotakin, johon kaikki sielut pyrkivät. Hyvä siis

luo tahtoa ja vahvistaa motivaatiota. Russellin (1954) mukaan hyvyys taas liittyy tarpeiden tyydyttämiseen. Se, mikä on hyvää, tyydyttää ihmisen tarpeita.

Ahokallio (1987) puhuu teknisestä, inhimillisestä ja arvotetusta hyvästä. Ihminen voi olla teknisesti hyvä ollessaan hyvässä fyysisessä kunnossa ja inhimillisesti hyvä suhtautuessaan toisiin ihmisiin hyvyydellä ja ymmärtämyksellä. Arvostettu hyvä taas tarkoittaa hyvää tietyn tarkoituksen kannalta.

Arviointi voi olla hyvää tarjotessaan oikeaan aikaan osuvia kehittämissuosituksia ja luodessaan halua kehittämiseen eli hyvän edistämiseen. Hyvään liittyy siis tulevaisuusorientaatio. Edellä todetun perusteella hyvään arviointiin kuuluvat myös totuuden kriteerit sekä arvioinnin eettinen korkeatasoisuus. Hahmottaessaan kokonaisuuksia ja täyttäessään laatutavoitteet arvioinnin hyvyyteen sisältyvät lisäksi kauneuden kriteerit. Arvioinnin hyvydessä voidaan katsoa kaikkien perusarvojen tavalla tai toisella toteutuvan.

Lopuksi

Arviointi toteutuu arjessa usein käytännöllisenä harkintana. Arvot jäävät tällöin implisiittisiksi. Emme siis analysoi sen enempää esimerkiksi totuutta, hyvyyttä kuin kauneuttakaan tietoisesti. Siitä voi taas seurata, että arviointi – ammattimainenkin – teknistyy, ja arvot sekoittuvat kenties intresseihin ja valtaan. Vallinnat tehdään intressien perusteella eikä niinkään yhteisesti arviointiryhmässä pohdittujen arvojen perusteella. Arviointi muuttuu siten suuremmassa määrin vallankäytöksi kuin mitä se luonnostaan jo on.

Tarvitaan siis moniulotteista arvokeskustelua ja -analyysiä. Arvot ovat myös ajassa muuttuvia, yksilöllisesti ja yhteisöllisesti. Siksi tarvitsemme arvojen ajantasaistavaa läpivalaisua arviointitoiminnan perustaksi. Itse asiassa jokaisen arvioinnin keskeisin haaste on onnistunut arvottaminen arviointiobjektin hyvyyden määrittämiseksi, olkoon se sitten subjektiivista, objektiivista tai jotain kolmatta. Samalla täytyy muistaa, että tyytymättömyys tarpeeksi hyvään on kaiken kehittämisen perusta. Laatu ei ole koskaan valmis arvioinnin aikana eikä sen jälkeen.

Lähteet

- Aaltonen, R. 1989. Naturalistinen paradigma evaluaatiotutkimuksessa. Teoksessa Kriittinen ajattelu aikuiskasvatuksessa. Vapaan sivistystyön 31. vuosikirja. Helsinki: Kirjastopalvelu, 145–162.
- Aaltonen, T. & Junkkari, L. 1999. Yrityksen arvot ja etiikka. Porvoo: WSOY.
- Aarnio, V. 2007. Onko arvoja olemassa. Realistisen arvo-ontologian puolustus. Helsingin yliopisto.

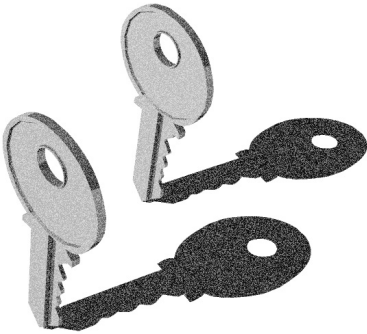


- Ahlman, E. 1976. Kulttuurin perustekijöitä. Alkuteos julk. 1939. Jyväskylä: Gummerus.
- Ahokallio, T. 1987. Elämys – maailmankatsomuksen perustus. Helsinki: Kirjapaja.
- Allaire, Y. & Firsirotu, M. 1984. Theories of organizational culture. *Organization Studies* 5 (3), 193–226.
- Allardt, E. 1964. Sosiaaliset arvot. Yhteiskuntatieteiden käsikirja II. Keuruu: Otava.
- Aristoteles. 1989. Nikomakhoksen etiikka. Suom. S. Knuutila. Helsinki: Gaudeamus.
- Atjonen, P. 2007. Hyvä, paha arviointi. Helsinki: Tammi.
- Barnes, B. & Bloor, D. & Henry, J. 1996. Scientific knowledge. A sociological analysis. Cambridge: Athlone Press.
- Engeström, Y. 2005. Mistä arvot tulevat ja mihin ne menevät: toiminnan näkökulma. Julkaisussa A.-M. Pirttilä-Backman, M. Ahokas, L. Myyry & S. Lähteenoja (toim.) Arvot, moraalit ja yhteiskunta. Sosiaalipsykologisia näkökulmia yhteiskunnan muutokseen. Helsinki: Gaudeamus, 325–344.
- Entonen, T. & Lyytinen, H. K. 1979. Avoimen yliopiston toimivuus kasvatustieteen opetuksessa. Osaraportti I. Jyväskylän yliopisto. Kasvatustieteen laitoksen julkaisuja 85.
- Eräsaari, R. 1999. Menestysuniversumia merkitsemässä. Julkaisussa R. Eräsaari, T. Lindquist, M. Mäntysaari & M. Rajavaara 1999. Arviointi ja asiantuntijuus. Helsinki: Gaudeamus.
- Guba, E. & Lincoln, Y. 1989. Fourth generation evaluation. Newberry park, London: Sage, 202–234.
- Harrinvirta, M., Uusikylä, P. & Virtanen, P. 1998. Arvioinnin tila valtionhallinnossa. Tutkimukset ja selvitykset 7/98. Helsinki: Valtionvarainministeriö.
- Heinonen, S. 2002. Korkeakoulujen työttömien omaehtoisen täydennyskoulutuksen sidoryhmäperusteinen arviointitutkimus. Helsingin yliopisto. Kasvatustieteen laitoksen tutkimuksia 180.
- Hirsjärvi, S. (toim.) 1983. Kasvatustieteen käsitteistö. Helsinki: Otava.
- Hirsjärvi, S., Liikanen, P., Remes, P. & Sajavaara, P. 1986. Tutkimus ja sen raportointi. Helsinki: Kirjayhtymä.
- House, E. 1980. Evaluating with validity. Beverly Hills, CA: Sage.
- House, E. 1993. Professional evaluation: Social impact and political consequences. Newberry Park: Sage.
- Häyry, H. & Häyry, M. 1988. Arvot ja niiden arviointi. Julkaisussa A. Kajanto (toim.) Rakennamme tulevaisuutta. Utopiasta vai uhkakuvista. Kokkola: Vapaan sivistystyön yhteistyöjärjestö, 18–36.
- Joas, H. 2000. The genesis of values. Cambridge: Polity Press.
- Karma, K. 1983. Käyttäytymistieteiden metodologiset perusteet. Helsinki: Otava.
- Ketonen, O. 1976. Se pyörii sittenkin. Porvoo: WSOY.
- Kotkavirta, J. & Nyssönen, S. 1996. Ajatus etiikka. 1.–2. painos. Helsinki: Weilin & Göös.
- Lampinen, O. 2006. Tasa-arvo utopiaa vai realismia. Julkaisussa P. Niemelä & T. Purssiainen. Hyvinvointi yhteiskuntapolitiikan tavoitteena. Juhlakirja professori Juhani Laurinkarin täyttäessä 60 vuotta 8.10.2006. Sosiaalipoliittisen yhdistyksen julkaisuja 62. Kuopion yliopisto.
- Laukkanen, R. 1994. Koulutuksen tuloksellisuuden arvioinnin keskeiset kysymykset ja trendit. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitoksen julkaisusarja A. Tutkimuksia 59.
- Launonen, L. 2007. Koulun arvotietoisuus. Julkaisussa O. Ikonen & P. Virtanen (toim.) Erilainen oppija – yhteiseen kouluun. Jyväskylä: PS-kustannus.
- Levomäki, I. 1998. Arvojen moninaisuus tietoyhteiskunnassa. Sitra 178. Helsinki.
- Linnakylä, P. & Välijärvi, J. 2005. Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulutuksen kehittämiseksi. Jyväskylä: PS-kustannus.
- Lyytinen, H. K. 2006. Koulutuksen arviointiosaaminen verkostossa – mahdollisuuksia ja kehittämishaasteita. Julkaisussa H. Silvennoinen (toim.) Koulutuksen arviointi verkostoituu. Koulutuksen arviointineuvoston julkaisuja 18. Jyväskylä, 45–65.

- Lyytinen, H. K. & Nikkanen, P. 1996. Arvot ja arvoarviointi oppivassa koulussa. Julkaisussa P. Nikkanen & H. K. Lyytinen, *Oppiva koulu ja itsearviointi*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 124–129.
- Lyytinen, H. K. & Nikkanen, P. 2005. Kehittämiskoivomaa arvioinnista. Teoksessa: H. K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnista*. Koulutuksen arviointisihteeristön julkaisuja 6. Jyväskylä, 129–141.
- Moore, G. E. 1903. *Principia ethica*. Revised edition. Toim. T. Baldwin. Cambridge: University Press 1993.
- Mäntysaari, M. 1999. Arviointitutkimuksen taustaoletukset. Julkaisussa R. Eräsaari, T. Lindquist, M. Mäntysaari & M. Rajavaara (toim.) *Arviointi ja asiantuntijuus*. Helsinki: Gaudeamus, 54–68.
- Niiniluoto, I. 1980. Johdatus tieteenfilosofiaan. Käsitteen- ja teorianmuodostus. Keuruu: Otava.
- Niiniluoto, I. 1993. Arvojen muutos ja moraalikasvatus. *Aikuiskasvatus* 13 (2), 84–93.
- Niiniluoto, I. 1994. Järki, arvot ja välineet. Keuruu: Otava.
- Nikkanen, P. & Lyytinen, H.K. 1996. *Oppiva koulu ja itsearviointi*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Ojanen, E. 2004. Hyvä työntekijä, hyvä ammattilainen, hyvä ihminen. Julkaisussa *Ammattikorkeakouluetiikka*. Opetusministeriön julkaisuja 2004: 30. Helsinki, 72–75.
- Oksanen, M. 2000. Yhteiskunnan moniarvoisuus ja luonnon arvottaminen. Teoksessa A. Haapala & M. Oksanen (toim.) *Arvot ja luonnon arvottaminen*. Helsinki: Gaudeamus.
- Owen, J. M. 1993. *Program evaluation: Forms and approaches*. St.Leonards: Allen & Unwin.
- Patton, M. Q. 1997. *Utilization-focused evaluation*. Thousand Oaks: Sage.
- Peltonen, M. & Ruohotie, P. 1992. *Oppimismotivaatio*. Aavaranta-sarja. Keuruu: Otava.
- Pietarinen, J. 2000. Ihmislähtöiset ja luontoarvot ja luonnon omat arvot. Teoksessa A. Haapala & M. Oksanen (toim.) *Arvot ja luonnon arvottaminen*. Helsinki: Gaudeamus.
- Pihlström, S. 2000. Luonnon arvo, inhimillinen toiminta ja ihmettely. Teoksessa: A. Haapala & M. Oksanen (toim.) *Arvot ja luonnon arvottaminen*. Helsinki: Gaudeamus.
- Platon. 2007. *Valtio*. Suom. K.-M. Itkonen. Helsinki: Otava.
- Rainio, K. & Helkama, K. 1974. *Sosiaalipsykologian oppikirja*. Porvoo: WSOY.
- Raunio, K. 1987. Arvot ja käytäntö. Sosiaalipolitiikan arvolähtökohtien ja toimintastrategioiden tarkastelua. Turun yliopiston julkaisuja. Sarja C. Osa 63.
- Rawls, J. 1972. *A theory of justice*. Oxford: Oxford University Press.
- Rescher, N. 1969. *The introduction to value theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Ricoeur, P. 2005. Arvostelemisesta tekona (suom. Jarkko S. Ruusuvuori). *Niin & näin, filosofinen aikakauslehti* 2, 71–73.
- Rokeach, M. 1973. *The nature of human values*. New York: Free Press.
- Russell, B. 1947. *Religion and science*. London: Oxford University Press.
- Russel, B. 1954. *Human society in ethics and politics*. London: Allen & Unwin.
- Salmela, M. 1999. Realismi ja antirealismi 1900-luvun suomalaisessa moraalifilosofiassa. *Niin & näin, filosofinen aikakauslehti* 1, 1–14. Saatavilla www.netn.fi/199/netn_199_salm.html. (Luettu 3.9.2007).
- Schein, E. 1987. *Organisaatiokulttuuri ja johtaminen*. Espoo: Weilin + Göös.
- Scriven, M. 1997. *The new science of evaluation*. Evaluation as a tool in the development of social work discourse. Conference in Stockholm, Sweden.
- Shadish, W. R. Jr., Cook, T. & Leviton, L. 1991. *Foundation of program evaluation: Theories and practice*. Newberry Park: Sage Publications.
- Sinkkonen, S. & Kinnunen, J. 1994. Arviointi ja seuranta julkisella sektorilla. Kuopion yliopiston julkaisuja E. *Yhteiskuntatieteet* 22.



- Sintonen, H. 1987. Sosiaalihuollon vaikuttavuusprojekti. Sosiaalipolitiikan ja erityisesti sosiaalihuollon arviointi metodisten kysymysten mm. kustannus-hyötyanalyysin näkökulmasta. Teoksessa T. Purola, K. Urponen & H. Sintonen. Vaikuttavuuden arvioinnin mahdollisuuksista ja menetelmistä sosiaalihuollossa. Sosiaalihuollon julkaisuja 1987: 18. Helsinki: Valtion painatuskeskus, 77-119.
- Smith, A. 1776. An inquiry into the nature and causes of the wealth of nations. Oxford: Clarendon Press.
- Smith, M. & Glass, G. 1987. Research and evaluation in education and in the social sciences. N.J.: Englewood Cliffs, Prentice-Hall.
- Strömmer, O. 1999. Henkilöstöjohtaminen. Helsinki: Edita.
- Suhonen, P. 1988. Suomalaisten arvot ja politiikka. Helsinki: WSOY.
- Temmes, M. 2004. Arviointityön autuus ja salahaudat arvioijan näkökulmasta. Hallinnon tutkimus 23 (2), 86-93.
- Tuomi, A. 1995. Eettiset, esteettiset ja totuusarvot. Julkaisematon opetusmoniste. Jyväskylä.
- Turunen, K. E. 1993. Arvojen todellisuus. Johdatus arvokasvatukseen. Jyväskylä: Atena Kustannus.
- Vaherva, T. 1983. Koulutuksen vaikuttavuus. Käsiteanalyttistä tarkastelua ja viitekehyyksen hahmottelua. Jyväskylän yliopisto. Kasvatustieteen laitoksen julkaisuja A 1.
- Vartiainen, P. 2003. The substance of stakeholder evaluation: Methodological discussion. International Journal of Public Administration 26, 1-18.
- Viinamäki, O.-P. 2004. Seurantatiedon hyödyntäminen ohjelma-arvioinnissa: Esimerkkinä Euroopan sosiaalirahaston ohjelma-arvioinnit ja seurantajärjestelmä. Hallinnon tutkimus 23 (2), 30-40.
- Virtanen, P. 2004. The rise and fall of evaluation standards. Hallinnon tutkimus 23 (2), 16-29.
- Virtanen, P. 2007. Arviointi. Arviointitiedon luonne, tuottaminen ja hyödyntäminen. Helsinki: Edita.
- von Wright, G. H. 2001. Hyvän muunnelmat (suom. Vesa Oittinen) Helsinki: Otava.
- Vuori, J. 2004. Arviointi(tutkimus) on arvotonta, ellei se arvota arviointiaan. Hallinnon tutkimus 23 (2), 1-3.



Arvioinnin eettisyys

Arviointi on määrittelynsä mukaisesti asioiden arvottamista eli suhteuttamista yhteisesti hyväksytyihin arvoihin. Siksi eettisyys on sen avainelementti. Eettiset arvostukset määrittävät arvioinnin tehtävää ja päämääriä, mutta yhtä lailla sen metodologiaa ja johtopäätösten tekoa. Artikkelissa tarkastellaan arviointietiikkaa neljästä näkökulmasta: arvioinnin suhdetta arvoihin, arvioinnin ohjaavuuden ja arvostelevuuden ristiriitaa, arviointia vallankäyttönä sekä vastuuna ja tilivelvollisuutena. Lisäksi pohditaan arviointieettisten ongelmien lisääntymistä ja eettisten standardien merkitystä sekä esitellään arviointieettisiä periaatteita ja hyveitä.

Asiasanat: etiikka, arviointi, arvot, oikeudenmukaisuus

Miksi arviointi on eettistä toimintaa?

Koulutuksen arvioinnissa tulisi kiteytyä koulutuksen etiikan keskeisiin sisältö, joka näkyy myös koulutuspolitiikan jaloimmista pyrkimyksissä (House 1993; Weeden ym. 2002). Siksi arvioinnin tulisi edistää yhteiskunnassamme demokraattisesti sovittuja pyrkimyksiä tasa-arvoiseen ja oikeudenmukaiseen koulutukseen, joka takaa jokaiselle sukupuolesta, iästä, asuinpaikasta, kielestä, etnisestä taustasta tai taloudellisesta asemasta riippumatta yhtä hyvät koulutus- ja oppimismahdollisuudet ja joka tukee erityisesti niitä, joiden edellytykset ovat eri syistä muita heikommät.

Jotta näin tapahtuisi, arvioijien tulisi tiedostaa arviointiperusteidensa eettisyys. Vaikkapa pelkkä vaikuttavuus, tehokkuus tai taloudellisuus ei riitä, vaan olennaista on, mitä arvoja edistetään tehokkuudella ja taloudellisuudella. Vaikuttavuus voi ilmetä tuloksina, joita ei loppujen lopuksi haluta jonkin sellaisen kustannuksella, joka on kasvatuksessa ja koulutuksessa vielä arvokkaampaa kuin taloudellisuus (House & Howe 1999). Esimerkiksi monet oppilaat yltyvät erinomaisiin oppimistuloksiin, mutta samalla toisten itseluottamusta saatetaan murentaa, jos odotukset ovat liian suuret. Taloudellisuuden kriteerein voidaan lakkauttaa monet maaseudun pienet koulut, mutta tuhota samalla syrjäseutujen elinvoimaisuus sekä asukkaiden elämänusko ja tasa-arvoiset koulutusmahdollisuudet.

Arviointitiedon avulla voidaan kuitenkin edistää yhteisesti sovittujen arvojen toteutumista koulutuksessa, jos arviointitietoa hankitaan eettisten periaatteiden mukaisesti (Castells 1999; House 1993). Arvioinnin etiikka ilmenee moraalisisina valintoina ja päätöksiinä, jotka kattavat koko arviointiprosessin kohteen valinnasta arviointimenetelmien ja mittavälineiden laadun varmistamiseen, arvioitavien kohteluun sekä arviointitulosten tulkintaan, johtopäätöksiin ja niistä seuraaviin kehittämistoimiin (House 1980; 1993). Tässä mielessä arvioinnin eettiset periaatteet ovat paljolti samat kuin tutkimuksen (Hallamaa ym. 2006; Kuula 2006; Morris 2005). Arvioinnin eettinen vastuu korostuu tutkimusta enemmän siinä, että tutkija voi pidättäytyä käytännöllisten seuraamusten pohdinnasta ja suositusten tekemisestä ja keskittyä empiirisen ja teoreettisen vastaavuuden arviointiin, mutta arviointi velvoittaa aina tarkastelemaan eksplisiittisesti arvottavia päätelmiä.

Eettisten periaatteiden luonne ja tiukkuus arviointitiedon hankinnassa liittyy myös valittuun arviointiparadigmaan (House & Howe 1999). Positivistinen, psykometrinen paradigma vaatii ensisijaisesti tiedon ja arvojen erottelua, metodista ja etenkin psykometristä ehdottomuutta tiedon luotettavuuden ja pätevyyden takaamiseksi sekä tiedonkeruun riippumattomuutta. Pragmaattiset opetussuunnitelma- ja ohjelma-arvioinnit painottavat puolestaan arvioitavien osallistamista ja heidän arvostustensa huomioonottoa sekä arviointien hyödyllisyyden, tarkkuuden ja toteutettavuuden korostusta (vrt. Patton 1997; 2002).

Konstruktivistinen oppijakeskeinen arviointi tähdentää arvioitavien omia arvoja ja arviointikriteereitä, oppijan äänen kuulumista sekä arvioinnin pedagogista hyödyllisyyttä ja välitöntä sovellettavuutta. Postmoderni arviointi on melko relativistista, ja se hyväksyy monenlaiset näkemykset arvoista ja eettisistä periaatteista. Se pyrkii pikemminkin murentamaan kuin rakentamaan arvioinnin yhtenäistä arvoperustaa ja siihen perustuvaa asiantuntijuutta (House & Howe 1999; Mabry 1997).

Arvioinnin paradigmat ja niiden mukaiset metodologiat sekä etenkin joidenkin arviointihankkeiden keveys ja epätieteellisyys on johtanut joskus arviointien pätevyyden ja tieteellisyyden kyseenalaistamiseen. Erityisesti arviointien epäil-

lään ajavan joidenkin ryhmien etuja (Stufflebeam 1997). Tehokkuuden, markkinavoimien vaatiman hyödyn tai relativistisen ja yksilökeskeisen arvioinnin vastavoimaksi on vaadittu sosiaalisesti ja kulttuurisesti vastuullista arviointiparadigmaa (House & Howe 1999; Paechter 2003), jossa eettisyys ohjaa vahvasti sekä arviointien suunnittelua, toteutusta että tulosten tulkintaa.

Näin eettiseksi ihanteeksi muodostuu sosiaalinen oikeudenmukaisuus ja kansalaisvastuullisuus, joita edistetään kaikkien arviointitahojen tasa-arvoisella osallistamisella ja dialogisuudella, jossa valta-asetelmia avataan ja ristiriitoja ratkaistaan demokraattisesti pohtien ja neuvotellen (House & Howe 1999; House 2005). Arvioija ei ole tällöin passiivinen sivustakatsoja eikä kaikkietävä filosofi, vaan neuvotteleva asiantuntija, joka edistää eri tahojen osallistamista ja vuoropuhelua. Hän varmistaa, että kaikki asianosaiset ovat edustettuina ja saavat myös tuoda esille omat intressinsä arvioinnin tavoitteista ja toteutuksesta. Myös aineiston analyysistä ja tulosten esittelystä kannattaa keskustella, sillä metodologiset valinnat vaativat menetelmäasiantuntemuksen lisäksi myös menetelmien valintaperusteiden pohdintaa. Osallistaminen takaa yleensä sen, etteivät epäsuotuisat tulokset tule yllätyksenä eikä niitä pyritä vähättelemaan tai liioittelemaan (House 2005).

Miten eettisyys näkyy arvioinnissa?

Parhaimmillaan koulutuksen arviointi sekä heijastaa että edistää keskeisimpiä eettisiä arvoja, kuten totuudellisuutta, oikeudenmukaisuutta, tasa-arvoa, vastuullisuutta, rehellisyyttä ja suvaitsevuutta (Fitzpatrick 2005; Morris 2005; Weeden ym. 2002). Arvioijan odotetaan etsivän totuutta, oikeaa ja luotettavaa tietoa, jota hankitaan systemaattisesti, tavoitteisiin sopivin ja pätevin menetelmin ottaen huomioon ennen kaikkea arvioitavien etu (House 1980; 1995). Arvioitavien etuihin kuuluu muun muassa se, että arvioinnin tarkoitus, keskeiset arviointikriteerit ja seuraukset paljastetaan avoimesti ennakkoon (House 1995). Arvioitavien oikeuksiin sisältyy myös mahdollisimman monipuolinen palautetieto, jota voi hyödyntää edelleen oppimisen tai kehittämisen välineenä (Weeden ym. 2002).

Osuimmillaan arviointi tavoittaa koulutuksen ja osaamisen todellisen laadun sekä kattaa arviointikohteen tärkeimmät sisältöalueet mahdollisimman aidoissa ja elämänläheisissä opiskelutilanteissa. Luotettavan ja pätevän tiedon hankkimiseksi vastuullinen arvioija käyttää menetelmiä, jotka hän hallitsee, jotka sopivat arviointitehtävään ja jotka eivät missään vaiheessa aiheuta haittaa arvioitaville (House 1980; 1993; Häyry 1999). Hyödyllisimmillään arviointi tuottaa tuloksia, jotka ovat luotettavia, edistävät oppimista ja koulutuksen kehittämistä sekä lisäävät ymmärrystä tuloksien taustalla vaikuttavista tekijöistä.

Totuuteen pyrkiminen ei kuitenkaan aina ole helppoa, sillä monta kertaa arviointi on ”totuuden kertomista ihmisille, jotka eivät halua sitä tietää” (Chelmsky 1995, 53). Joskus arvioinnilta vaaditaan totuuden sijasta arvioitavien aseman puolustamista tai vahvistamista, koulutuspoliittisten tai hallinnollisten ratkaisujen legitimoitua tai resurssien lisäämistä. Toisinaan vaaditaan uusien opetusmenetelmien, välineiden tai materiaalien puolesta tai vastaan puhumista tai johtopäätöksiä, jotka lisäävät resursseja tai vahvistaisivat joidenkin eturyhmien valtaa (House 1995). Niin sanotun totuuden etsiminen on filosofiselta kannalta katsoen problemaattinen kysymys, jossa vaativan korrespondenssijattelun sijasta arviointiarjen kannalta toimivia tulkintoja ovat koherenssiteoreettinen ja pragmatistinen tulkinta (Heikkinen ym. 1999). Arvioinnin kohteiden kannalta todellisuudessa on tietysti olemassa monenlaisia totuuksia, jotka ovat yhtä tärkeitä ja tosia, mutta erilaisia esimerkiksi eri henkilöille tai intressiryhmille. Monien totuuksien yhtäaikainen olemassaolo on tietysti arvioinnin ilmeinen haaste.

Liian usein arviointeja tehdään kevyesti, nopeasti ja vähin resurssein niin, etteivät mitkään pätevyys- tai luotettavuuskriteerit täyty eikä hankittu tieto riitä totuuden kertomiseen. Metodologinen luotettavuus korvataan silloin uskottavuudella, anekdooteilla tai vakuuttelulla, riippumattomuus eturyhmien välisillä sopimuksilla ja yleistettävyyys harkinnalla ja tapauskuvauksilla. Uskottavuus kuitenkin riippuu siitä, onko olemassa vahva ammattikunta, joka ymmärtää arvioinnin eettisyyden ja arviointityön ja -tutkimuksen laadun merkityksen. Arviointitutkijatkin voivat joutua poliittisten tai hallinnollisten voimien vietäväksi, mikä saattaa koko ammattikunnan epäilyksen alaiseksi (House 1995; Simons 2000; Newman & Brown 1996). Ikävätkin tulokset on silti uskallettava paljastaa, samoin tiedon luotettavuuden tai pätevyuden rajoitukset.

Arvioinnilla on vahva eettinen sisältö, koska arviointi ei luonteensa mukaisesti ole intresseistä riippumatonta toimintaa, vaan ihmisten työn ja sen tulosten arvon vertaamista toisten ihmisten arvostukseen. Coombesin (1992) mukaan arviointi on vähintään neljästä syystä eettistä toimintaa:

1. Arviointi on tekemisissä ihmisten arvojen soveltamisen kanssa, joten arvot voivat olla ristiriidassa keskenään.
2. Arviointi palvelee mahdollisesti ristiriidassa olevia päämääriä, jotka synnyvät ohjaamisen ja arvostelun yhdistämisestä.
3. Arviointi on tekemisissä vallankäytön kanssa, eli joillakin on enemmän valtaa kuin toisilla ja joillakin on mahdollisuus vallan väärinkäyttöön.
4. Arviointi liittyy vastuuseen ja tilivelvollisuuteen.

Lisäksi koulutuksen arvioinnin vahva eettisyys liittyy siihen, että arvioitavana on useimmiten lapsia, nuoria tai aikuisia, niin opiskelijoita kuin opettajiakin, joille voi koitua arvioinnista haittaa, jos esimerkiksi tiedon luottamuksellisuut-

ta ei kunnioiteta. Keskenkasvuiset eivät osaa puolustaa oikeuksiaan aikuisten lailla, jolloin opettajan pedagogisen vallankäytön kannalta eettisellä sensitiivisyydellä on suuri merkitys (Atjonen 2004; Farrell 2005; Nuutinen 2000; Räsänen 2002). Lisäksi henkilöllisyyden suojasta ja henkilötietojen käsittelystä on säädetty lailla, mikä osaltaan vaikuttaa arviointimenetelmien valintaan, arvioinnin käytännön toteutukseen sekä tulosten julkistamiseen, hyvin samaan tapaan kuin tutkimuksessakin (vrt. Kuula 2006; The student evaluation standards 2003). Erityisen haasteelliseksi arvioitavien henkilöiden tai yhteisöjen eettisesti kunnioitava kohtelu tulee kehittävässä arvioinnissa, jossa informantit eivät voi todellisuudessa suojautua anonyymisyyden taakse.

Tarkastelemme seuraavassa Coombesin jäsentelyn avulla eräitä keskeisimpiä arviointieettisiä näkökohtia.

Miten eettiset periaatteet voivat olla monitulkintaisia, jopa ristiriitaisia?

Arviointieettisten periaatteiden välille syntyy käytännössä jännitteitä, jopa ristiriitoja, ja ne ovat joskus jo käsitteellisesti monitulkintaisia. Esimerkiksi tasa-arvo voidaan ymmärtää koulutukseen pääsyn, resurssien, oppimismahdollisuuksien, valintojen tai pedagogisen, taloudellisen ja sosiaalisen tuen tasa-arvona tai edellytyksiltään heikoimpien erilaisena kohteluna tulosten tasaisuuden tavoittelemiseksi oikeudenmukaisuuden hengessä (Grisay 1984; Husén 1986; Weeden ym. 2002). Oikeudenmukaisuutta ei toisaalta voi olla ilman tasa-arvon oletusta (House 1997; Rawls 1971). Mutta onko oikeudenmukaista tukea ainoastaan heikkoja oppilaita ja jättää keskitasoiset tai lahjakkaat erityistä tukea vaille? Onko parempi opettaa maahanmuuttajaoppilaita oman äidinkieltensä mukaisissa ryhmissä, joissa oppilaat tuntevat olonsa kotoiseksi ja suomen kieltä voidaan opettaa fokuksidusti toisena kielenä, vai suomenkielisissä ryhmissä, joissa oppilaat omaksuvat toista kieltä arkiopiskelussa ja saavat samalla suomenkielisiä ystäviä?

Oppilasarvioinnissa epäoikeudenmukaisena pidetään esimerkiksi sitä, että arvioidaan sellaisia tietoja tai taitoja, joita ei ole koulussa opetettu. Toisaalta koulussa tulisi arvostaa koulun ulkopuolista oppimista, harrastuksissa opittua tai etenkin täysin omatoimisesti etsittyä ja hankittua osaamista. Entä voidaanko asenteita, itsetuntoa ja motivaatiota arvioida, jos opettaja ei ole näitä erityisesti kehittänyt tai jos koulun opiskeluympäristö ei ole tukenut näiden vahvistumista? Epäreilua on myös kohdentaa arviointia kapea-alaiseen kognitiiviseen osaamiseen ja jättää muut alueet, kuten asenteet, aloitekyky ja rohkeus, arvioimatta. Reilua ei myöskään ole arvioida laaja-alaisia taitoja yksipuolisin keinoin, kuten kielitaitoa pelkästään sanaston tai kielen rakenteiden tuntemuksena tai suosimalla luonnontieteiden arvioinnissa pelkästään monivalintatehtäviä omin sanoin tai toin ilmaisemisen sijasta.

Totuudellisuuden ja avoimuuden arvostaminen arviointitulosten välittämises-
sä voi olla joskus ongelmallista. Esimerkiksi OECD:n oppimistulosten arviointi-
ohjelmassa PISA:ssa pohdittiin, voiko maahanmuuttajanuorten oppimistuloksia
julkistaa avoimesti ja rehellisesti. Näillä nuorillahan ei ole ollut samanlaisia op-
pimismahdollisuuksia kuin valtaväestöllä. Oikeudenmukaisuuden ja reilun pelin
hengessä tulosten julkistaminen on ollut arveluttavaa, jos siitä saattaisi seurata
maahanmuuttajanuorten leimautumista. Toisaalta vain julkistamisen kautta näi-
den nuorten koulutuksen ja oppimisen ongelmiin voitiin toivoa jatkossa tosis-
saan puututtavan. Kansainvälisen vertailuaineiston mukaanhan joissakin mais-
sa, esimerkiksi Ruotsissa ja Norjassa, maahanmuuttajanuorilla oli huomattavasti
paremmat oppimismahdollisuudet perusopetuksessa kuin Saksan ja Tanskan
kouluissa (Hvistendahl & Roe 2004; OECD 2004). Nämä tulokset ovat jo käänty-
neet varsin suotuisiksi koulutuksen kehittämistoimiksi Tanskassa ja Saksassa,
vaikka pelko kohderyhmän leimautumisesta ja suvaitsevuu-
den vähenemisestä oli aluksi ilmeinen.

Koulutuksen kansainvälisissä arvioinneissa kulttuurisia, kielellisiä ja talou-
dellisia epäoikeudenmukaisuuksia pohditaan jatkuvasti. Esimerkiksi koulusaa-
vutuksia arvioivan IEA-järjestön tutkimustulosten julkistamisen yhteydessä af-
rikkalaiset ja aasialaiset ovat epäilleet angloamerikkalaisen valtakulttuurin syr-
jivän heitä. Ranskalaiset, italialaiset ja espanjalaiset ovat syyttäneet englantia
puhuvia maita kielellisestä ylivallassa. Pohjoismaalaisia puolestaan huolestuttaa
usein se, ettei hyvinvointiyhteiskunnan tasa-arvoajattelua ja inklusioperiaatetta
arvosteta tarpeeksi. Erilaisten kulttuuristen ja niihin liittyvien eettisten ristiriito-
jen pohtiminen ja tasapainottaminen käytännön arviointitilanteissa, menetelmä-
valinnoissa ja tulosten tulkinnoissa on kansainvälisen arvioinnin keskeisimpiä
haasteita (Bempechat ym. 2002; Hambleton 2002).

Millaisia ristiriitoja arvioinnin ohjaavan ja arvostelevan tehtävän välillä voi ilmetä?

Vaikka arviointi parhaimmillaan on motivoivaa, ohjaavaa ja kehittävää, sillä on
myös tulosten osoittamisen ja arvostelun tehtävä. Näitä hyvinkin erilaisia teh-
täviä on usein vaikea erottaa toisistaan, oli kyse oppimisen, opetuksen, kou-
lutusohjelman tai koko koulutusjärjestelmän arvioinnista. Vaikka arvioinnissa
haluasi painottaa oppimista ja opetusta tukevaa arviointia, arviointiperusteet
muodostuvat helposti hyviin – ja usein myös toisten – tuloksiin vertaamisen
kautta. Esimerkiksi opettaja palkitsee oppilasta motivoidakseen tätä jo pienestä-
kin yrityksestä, mutta rikkoo samalla tasapuolisen ja oikeudenmukaisen arvos-
telun periaatetta. Kun opettaja auttaa oppimisvaikeuksissa, mutta toimii kokeita
arvostellessaan tuomarina, sama henkilö näyttäytyy oppilaalle hämmäntävästi

erilaisessa roolissa. Huolenpito ja auttaminen joutuvat näin ristiriitaan arvostelun tasapuolisuuden kanssa.

Sama ristiriita näkyy yleisemminkin, kun vertaillaan formatiivista ja summatiivista arviointia: formatiivisen arvioinnin pitää antaa ohjaavaa ja motivoivaa palautetta oppimisen edistymisestä, kun taas summatiivisen arvioinnin tarkoitus on tehdä päätelmiä jonkin suorituksen laadusta. Toisaalta summatiivinen arviointi voi ohjata ratkaisevasti kurssivalintoja ja jatko-opintoja.

Ristiriitoja saattaa syntyä myös opetuksen, koulun tai koulutusohjelman laadun arvioinnissa. Opetusta tai kouluja arvostellaan usein samanlaisin laatuksiteerein, vaikka toiset opettajat tai koulut ovat mahdollisesti saaneet huomattavasti muita enemmän resursseja, täydennyskoulutusta, opetusvälineitä tai vanhempien tukea. Lisäksi kouluihin on voinut valikoitua hyvin eritasoisia oppilaita. Näin käy usein esimerkiksi lukioden kohdalla, vaikka ylioppilaskokeen tuloksista pyritään päättelemään lukioden tasoa (Linnakylä & Välijärvi 2005). Tällaiset eriarvoisuuskysymykset ovat erityisen haasteellisia kansainvälisissä arvioinneissa.

Miten arviointi on sekä vallankäyttöä että vastuun ottamista?

Koulutuksen arviointia voi lähestyä myös vallankäytön näkökulmasta, sillä arvostustensa perusteella toimiva arvioija käyttää valtaa arvioitaviin nähden (House 1995; House & Howe 1999; Newman & Brown 1996). Esimerkiksi oppilasarvioinnissa oppilaiden tai heidän vanhempiensa kriteerit eivät välttämättä ole samoja kuin opettajien tai ulkopuolisten arvioijien. Joissakin kouluissa pyritään huipputuloksiin, toisissa arvostetaan heikkojen tukemista, kolmannessa viihtymistä ja hyvinvointia. Samat henkilöt voivat toimia välillä arvioitavina, välillä arvioijina tai arvioinnin järjestäjinä tai tilaajina. Intressit voivat näin vaihdella, ja niiden mukaisesti vaihtelevat näkökulmat arviointiin. Opettajakin näkee arvioinnin opetustilanteessa eri tavoin kuin silloin, kun hän käyttää valtaa rehtorina, opettajajärjestössä, koululautakunnassa tai kunnanvaltuustossa.

Koulutuksen järjestäjä voi olla ristiriitatilanteessa, kun vähäisiä paikallisia resursseja on jaettava eri sektoreiden ja erilaisten koulutuspalveluiden ja koulujen kesken. Päätäjät ratkaisevat, tuetaanko menestyviä vai niitä, jotka ovat eniten tuen ja resurssien tarpeessa. Myös koulutuspolitiikan tai opetussuunnitelman toteutumisen arvioinnissa voi olla ristipaineita esimerkiksi silloin, kun yhtäältä vaaditaan työelämän odotusten huomioonottoa ja toisaalta sivistystehtävän vaalimista, tasa-arvoisten oppimismahdollisuuksien turvaamista ja sosiaalisesti vastuuntuntoisten kansalaisten kasvattamista.

Arviointitiedon varassa tehdään joka tapauksessa sekä pedagogisia että hallinnollisia ja poliittisia päätöksiä, jolloin käytetään eksplisiittisesti valtaa. Arvioijan tuottamaan tietoon liittyy aina arvostuksia, ja tietoa käytetään päätök-

senteossa (House & Howe 1999). Vastuullinen arvioija tiedostaa itsekin työnsä eettisen haasteellisuuden (Coombes 1992; Newman & Brown 1996; Pring 2003). Hän edistää oikeudenmukaisuutta ja kansalaisvastuullisuutta, osallistaa eri arviointitahoja yhteiseen keskusteluun, jolla avataan valta-asetelmia ja etsitään yhteisten arvojen mukaisia periaatteita hyvien arviointikäytäntöjen ja -kulttuurin kehittämiseksi (House & Howe 1999).

Tässä suhteessa uudet, entistä autenttisemmat arviointimuodot, kuten erilaiset näytteet, näyttötutkinnot ja portfoliot arviointikeskusteluineen, pyrkivät jo menetelmällisesti osallistamaan arvioitavia, saamaan heidät neuvottelemaan yhteisistä arviointikriteereistä ja hankkimaan palautetta. Näin ne voimaannuttavat arvioitavia oman toimintansa reflektiiviseen pohdintaan ja uusien opiskeluhaasteiden suunnitteluun (Atjonen 2007; Kankaanranta ym. 2007; Linnakylä & Välijärvi 2005). Arviointiin osallistuminen vähentää siihen usein liittyvää jännitystä tai epävarmuutta omasta osaamisesta. Osallistuminen edistää ajattelutapaa, jonka mukaan arviointi voi olla sekä oppimisen että näytön paikka, joka vahvistaa itseluottamusta ja tervettä ylpeyttä hyvin tehdystä työstä.

Kun arviointivastuu annetaan ulkopuoliselle, arvioinnin järjestäjän ja arviointitavien on voitava luottaa siihen, että arvioijalla on tehtävänsä vaatima pätevyys, jonka varassa hän voi toimia tehokkaasti ja oikeudenmukaisesti asiantuntijavastuunsa kantaen (House 1993; 1995; Newman & Brown 1996). Tämän taustalla on puolestaan ajatus, että arviointi on eettistä työtä ja koskee arviointiin eri tavoin osallistuvia. Tilivelvollisuus on kaksitahoinen. Arvioitavan pitää tehdä oma työnsä hyvin pyrkiessään oppimiseen tai kehittymiseen. Samoin arvioijan tulee tehdä oma työnsä hyvin vastuunsa tuntien ja kehittämistä edistäen. Vaikka arvioijalla on enemmän valtaa kuin arvioitavalla, hänellä on myös enemmän vastuuta.

Asiantunteva ja vastuullinen arvioija rakentaa teoreettisesti kestävä arviointikehyksen, hyödyntää monipuolisesti ja tarkoituksenmukaisesti erilaisia metodologioita. Hän tuottaa tuloksia, jotka ratkaisevat koulutuksen, opetuksen ja oppimisen todellisia ongelmia ja lisäävät ymmärrystä arviointikohteesta – olipa kyse tiedollisista oppimistuloksista, maahanmuuttajanuorten sopeutumisesta suomalaisen kouluun, aikuiskoulutuksen esteistä tai koko suomalaisen koulutusjärjestelmän laadusta. Arvioija ymmärtää tulosten historiallisen, kulttuurisen, taloudellisen ja sosiaalisen kontekstin merkityksen ja tekee johtopäätöksiä nämä huomioonottaen.

Vastuullinen arvioija ei yleistä tuloksia, jos otanta ei riitä yleistykseen ja ilmaisee selvästi tulosten tulkinnan rajoitukset. Hän ei korosta eroja, jotka loppujen lopuksi ovat vähäisiä – vaikka ne olisivat suurissa aineistoissa tilastollisesti merkitseviä – eikä toisaalta vähättele ongelmia, joita ilmenee esimerkiksi etnisten, kielellisten tai sosio-ekonomisten ryhmien välillä. Arviointitutkijan tulisi pyrkiä oikeutettuihin johtopäätöksiin ja kehittämisehdotuksiin, jotka edistävät erilais-

ten oppijoiden opiskelua ja opetusta tasa-arvoperiaatteen hengessä ja oikeudenmukaisuutta kunnioittaen. (Linnakylä & Välijärvi 2005.)

Koskeeko vastuu myös tuloksia?

Tulosvastuuta tarkasteltaessa on pakko ottaa vastaan hyvinkin vaativia eettisiä haasteita, kun näkökulmaa laajennetaan oppilasarvioinnista oppilaitosten, kuntien koulutoimen, korkeakoululaitoksen tai kansallisen koulutuspolitiikan arviointiin. Ulkoisen arvioinnin tehtävänä on osoittaa rehellisesti oppimisen, opetuksen ja koulujärjestelmän heikkoudet ja vahvuudet. Tulokseton ja tehoton opetus, aikansa elänyt opetussuunnitelma tai puutteellinen opiskeluympäristö voi olla kohtalokas sekä koululaisille että koko kansakunnalle. Osaaminen ja sen kehittäminen ovat tulevaisuuden tärkein voimavaramme. Tässä yhteydessä olemme tekemisissä erityisesti seurauseettisten kysymysten kanssa.

Koulutuksen arvioinnin on paljastettava hyviä käytäntöjä, vahvuuksia ja kehittämishaasteita myös tuloksellisuuden ja tehokkuuden näkökulmasta. Silti se ei saa tehokkuuden nimissä unohtaa sosiaalista vastuuta, tasa-arvoa ja heikoimpien tukemista, mikä on suomalaisen koulutuspolitiikan arvoperusta (vrt. Lounonen 2000; Oravakangas 2005). Kansakunnan sosiaalisen oikeudenmukaisuuden tajusta kertoo paljon se, miten se kohtelee vähäosaisimpia. Vaikka joidenkin yksilöiden huippuosaaminen voi toimia muidenkin pyrkimysten innoittajana, osaamisen laajempi hyödyntäminen voi jäädä vähäiseksi, ellei sosiaalisia ja kulttuurisia arvoja painoteta tarpeeksi. Rahalla ei saa mitata kaikkea inhimillistä kasvatuksellista toimintaa.

Lisääntyvätkö arvioinnin eettiset ongelmat?

Arvioinnin eettiset ongelmat näyttävät lisääntyneen arviointien yleistyessä ja monimuotoistuuksessa. Arviointitiedon tarve on kasvanut nopeasti yhteiskunnan, työelämän ja kulttuurin muutoksen myötä. Tarvetta ovat lisänneet muun muassa hallinnon ja päätösvallan hajauttaminen, tähän liittyvä informaatio- ja tulosohjaus, entistä avoimempi ja kansainvälisempi koulutuspolitiikka, inhimillisen pääoman merkityksen vahvistuminen, julkisten menojen leikkaukset, koulutusjärjestelmän muutokset, oppilaitosten ja opiskelijoiden itseohjautuvuus sekä pedagogisten ratkaisujen valinnaisuus (Linnakylä & Välijärvi 2005).

Arviointi-intoa voisi kuvata jo melkein buumiksi, johon voi liittyä toisarvoisia imagonkohottamispyrkimyksiä. Välillä arvioinnista tunnutaan etsittävän apua silloinkin, kun luotettavaa ja pätevää tietoa – niin vaikea kuin sitä epistemologisessa mielessä onkin määritellä ja tuottaa – ei välttämättä edes haluta, vaan muutosten legitimointi tai jarruttaminen tai erilaisten eturyhmien intressit vaa-

tivat tukea (Vuorenmaa 2001). Sisäisiä arviointeja käynnistetään joskus ulkoisen arvioinnin pelossa, ulkoista arviointia taas resurssien lisäämisen toivossa. Aika usein arvioinnin vaikutukset ja seuraamukset ovat olemattomia, vaikka arviointi käynnistettiin juuri kehittämisen vauhdittamiseksi.

Resurssien lisäämisen toivossa heikkouksia voidaan joskus liioitella. Jos taas huippuosaamista palkitaan, mahdolliset haittavaikutukset ja häviäjistä huolehtiminen unohdetaan helposti. Arviointiprosessin aikana arvioitavia ei aina muisteta informoida arvioinnin todellisesta tarkoituksesta ja menettelytavoista. Tiedon keruu voi olla epäsystemaattista, ja tulokset tulkitaan pikemminkin subjektiivisten ennakko-odotusten kuin todellisen evidenssin varassa. Joskus taas kriteerit määritellään tiettyjen tulosten toivossa.

Median vahvistunut merkitys yhteiskunnassamme lisää myös arvioinnin eettisiä ongelmia. Mediassa arviointitulokset saavat julkisuutta eniten silloin, kun ne ovat huonoja ja paljastavat uhkia ja herättävät pelkoa (Gardner ym. 2001). Eettisesti erityisen ongelmalliseksi on muodostunut median vaatimus koulujen ranking-listojen julkaisemisesta. Media vetoaa avoimeen ja vapaaseen tiedonkulkuun demokratian perusoikeutena ja perustelee aktiivista tiedottamista kansalaisten oikeudella tietää, mitä julkisilla varoilla kustannettu koulutus tuottaa. Myös vanhempien tarpeeseen saada tietoa kouluvalintoja varten halutaan vastata. (Brooks 2002.) Taustalla vaikuttanee lisäksi uusliberalistinen näkemys, jonka mukaan kilpailu parantaa aina laatua (Maw 1999; Linnakylä & Välijärvi 2005; Paechter 2003). Arviointitutkijat puolestaan tietävät, että koulujen paremmuusjärjestykset ovat usein epäoikeudenmukaisia, ja ne antavat harvoin kokonaiskuvan koulun tasosta. Julkisuus rikkoo helposti luottamuksellisuuden ja oikeudenmukaisuuden periaatteita. Tutkijat tietävät myös, ettei kilpailu aina nosta yleistä laatutasoa vaan vahvistaa koulujen välisiä eroja. Oppilaiden osaamiseen vaikuttavat myös monet tausta- ja resurssitekijät, joita on vaikea tuoda esille ranking-listojen yhteydessä. (Hautamäki & Kuusela 2005; Malin 2005.)

Koulutusjärjestelmien ranking-listat eivät ole suotavia, vaikka ne saavat näkyvyyttä etenkin silloin, kun kansallinen koulutus loistaa listojen kärjessä. Maa on harva lähtee kuitenkin vaihtamaan paremmuusjärjestyksen takia. Jos taas kansalliset tulokset ovat heikkoja, voi seurauksena olla opettajien ammattitaidon vähättelyä. Hyviä oppimistuloksia tuottavan koulutusjärjestelmän syvällisemmät jatko- ja tausta-analyysit voivat sentään tukea hyvien pedagogisten käytäntöjen tunnistamista ja jakamista.

Tarvitaanko arviointiin eettisiä standardeja?

Arvioinnin käytäntöjen kirjavuus sekä metodologiset ja eettiset ongelmat ovat johtaneet monessa maassa eettisten standardien määrittelyyn (esim. Guiding principles for evaluators 2007; Standards for educational testing 1999; The Joint Committee on Standards for Educational Evaluation 2006; The student evaluation standards 2003). Keskeisintä näissä standardeissa on huomioida arvioitavien edut ja oikeudet, joita ei tule loukata arviointiprosessin missään vaiheessa. Toisaalta arvioinnin etiikka painottaa myös luottamusta arvioijan vastuullisuutta, oikeudenmukaisuutta ja pätevyyttä kohtaan sekä sellaisen metodologian valintaa, joka tyydyttää sekä tutkimuksen että eettisen toiminnan vaatimukset ja vahvistaa samalla arvioijien omaa eettistä pohdintaa ja kehittymistä.

Kirjoitetuista eettisistä koodeista ja suosituksista huolimatta arvioijan on otettava rohkeasti itse kantaa eettisiin kysymyksiin, joihin ei ole yhtä oikeaa vastausta tai käsikirjaa, vaan jotka vaativat monien osapuolten kuuntelua ja yhteisen näkemyksen rakentamista. Silti kirjoitetuilla eettisillä ohjeilla on esimerkiksi arvioijan ammattieettisten suuntaviivojen luojana oma merkityksensä, koska ne lisäävät tietoisuutta hyvien eettisten käytäntöjen rajoista.

Kansainvälisissä arvioinneissa valta-asetelmat, arvostukseen liittyvät intressit ja eettiset ristiriidat näkyvät kansallisia selvemmin. Monikansallisten arviointien suunnittelu herättää aina kiivaita koulutuspoliittisia ja eettisiä keskusteluja. Esimerkiksi parhaillaan suunnitteilla olevan OECD-maiden aikuisten avaintaitojen arvioinnin yhteydessä eri maiden asiantuntijoiden näkemykset arvioinnin toteutuksesta poikkesivat merkittävästi sekä metodologisesti että erityisesti eettisesti. Kenen ehdoilla arvioidaan: yksilön vai yhteiskunnan; koulutuksen vai työelämän? Arvioidaanko sitä, mitä globalisoituva talous ja tietoyhteiskunnan työtehtävät vaativat vai pyritäänkö korostamaan ihmisen monipuolista henkistä kasvua, lähimmäisestä huolehtimista, yhteisöllisyyttä ja aktiivista kansalaisuutta?

Pohjoismainen näkökulma poikkeaa usein muista eurooppalaisista ja amerikkalaisesta, etenkin keskusteltaessa sosiaalisen koheesion merkityksestä sekä koulutuksen tavoitteena että tietoyhteiskunnan voimavarana. Näissä keskusteluissa arvioinnin asiantuntijoilta vaaditaan paitsi vahvaa metodologista osaamista, myös vahvaa eettistä selkärankaa ja taitavaa argumentointia. Tämä puolestaan vaatii sekä omien että kansallisten arvojen tiedostamista ja pohtimista tänä päivänä ja tulevaisuudessa (Gardner ym. 2001; Morris 2005).

Millaisia arvioinnin eettisiä periaatteita ja arvioijan hyveitä voidaan vaatia?

Edellä on jo puhuttu hyvän arvioinnin ominaispiirteistä ja esitelty arvioijalta vaadittavia ominaisuuksia. Ne ovat osin samoja kuin tutkijalta odotetaan etenkin silloin, kun toteutetaan vaikkapa toiminta- tai kehittämistutkimusta kvalitatiivisen paradigman mukaan. Silloin pyritään jatkuvasti peilaamaan tutkimusta suhteessa käytännön toimivuuteen ja tekemään yhteiseen hyvään perustuvia pragmaattisia päätöksiä seuraavasta tutkimusaskeleesta. Newman ja Brown (1996) ovat käsitelleet viittä arviointieettistä periaatetta, jotka nekään eivät anna tyhjentäviä vastauksia arjen kysymyksiin eivätkä ole toisistaan riippumattomia. Eniten he painottavat haitan tai vahingon välttämisen periaatetta, mikä tarkoittaa arvioinnissa tavallisimmin psyykkisiä haittoja, esimerkiksi stressin lisäämistä tai henkilösuhteiden kriisiytymistä. Toisinaan taitamattomasti tai epäeettisin piilopyrkimyksin toteutetusta arvioinnista voi olla myös tarkoittamattomia taloudellisia seurauksia tai haitallisia sivuvaikutuksia.

Haitan välttäminen voi jättää varjoonsa hyvän tekemisen, jonka pitäisi motivoida arviointia. Arviointitiedon keruun pitäisi aina palvella kehittymistä ja suuntautua tulevaisuuteen. Hyvän tekemiseen liittyy eettinen seurausoletus: hyvän tekemisestä seuraa lisää hyvää – ellei heti, niin ehkä myöhemmin (Einhorn 2005). Hyvän tekemistä ei pidä ajatella vain velvollisuutena tai seurauksia laskelmoiden vaan myös hyveenä, tärkeänä ja tavoiteltavana. Arvioinnissa menestymisen tulisi edellyttää hyvyyttä: toisten huomioon ottamista, sosiaalista taitavuutta ja luottamuksen arvoiseksi osoittautumista.

Oikeudenmukaisuuden periaatetta voidaan pitää Platonin tavoin ylimpänä hyveenä. Oikeudenmukaisuus tarkoittaa reiluutta, kohtuullisuutta ja tasapuolisuutta. Sen vastakohtana epäoikeudenmukainen arviointi loukkaa tai laiminlyö joidenkin oikeuksien toteutumista. Joskus on kyse jopa ihmisoikeuksien kyseenalaistamisesta, esimerkiksi jos syrjitään erityistä tukea tarvitsevia oppilaita, kansallisia vähemmistöjä tai maahanmuuttajia. Monet arviointietikot korostavat, että totuudellinen ja oikeudenmukainen arviointi voi jäädä etäiseksi ja kylmäksi. Vastapainoksi tarvitaan huolenpitoa, läheisyyttä ja vastavuoroisuutta, joita esimerkiksi osallistava ja autenttinen arviointi tarjoavat (Juujärvi 2003; Noddings 2003.)

Arvioinnissa pitäisi noudattaa myös uskollisuuden periaatetta, eli on tärkeää pitää alkuperäisistä lupauksistaan kiinni ja olla lojaali arviointiin osallistujille. Viimeksi mainittu ei ole aina helppoa, koska esimerkiksi ohjelma-arvioija saattaa joutua pohtimaan kiperästi, kenelle hän olisi lojaali: asiakkaalle, projektin ohjausryhmälle, ohjelmaan osallistuneille vai veronmaksajille. Uskollisuus korostuu pitkäaikaisissa arviointisuhteissa eli esimerkiksi alakoulun opettajan työssä.

Autonomiaa Newman ja Brown tarkastelevat sekä arviointieettisenä periaatteenä että arvioijan tavoiteltavana hyveenä. Itsemääräämisoikeus eli autonomisuus on tärkeä inhimillinen tarve ja kokemus, joka on tärkeä sekä arvioijalle että arvioitavalle. Vaikka arvioinnista pitäisi olla aina kehittämissuureita, ei ketään lopulta voida tai kannata pakottaa muuttumaan. Esimerkiksi jonkin koulutusohjelman toteuttajien pitää voida arviointinsa jälkeen pohtia itsenäisesti tuloksia sekä niiden seuraamuksia.

Kansalliset tai kansainväliset ulkoiset arvioinnit rikkovat autonomian periaatetta silloin, kun ne ilmestyvät kouluun lupaa kysymättä, jonkin valtaa pitävän arviointitahon kehoituksesta. Autonomiansa hukannut arvioija on etenkin monitahoarvioinneissa vaikeuksissa, ellei hän pysty määrittelemään omaa toimintatilaansa ja riippumattomuuttaan monien odotusten ristipaineissa. Arvioija toimii yleensä monessa roolissa etenkin ohjelma-arvioinneissa: tiedon kerääjänä, raportoijana, hallinnoijana, yhteisön ja arvioija-ammattin edustajana. Tästä herää helposti eettisiä jännitteitä.

Arvioinnissa, kuten kasvatuksessa, oppimisessa ja opettamisessakin, ollaan vääjäämättä tekemisissä ihmisten ja heidän henkilökohtaisten ominaisuuksiensa kanssa. Arvioijakin käyttää persoonallisuuttaan ja omia arvostuksiaan työvälineenä erityisesti osallistavassa arvioinnissa (Patton 2002; Fetterman 1998), oppilasarvioinnista puhumattakaan. Newman ja Brown kuvaavat autonomian lisäksi kahta muuta arvioijan eettistä hyvettä, jotka ovat henkilösuhteiden ja itsensä ylittämisen kyky.

Henkilösuhteiden kyvyllä he tarkoittavat monipuolisia ja rakentavia ihmishuuhdetaitoja. Arvioijan pitää olla taitava kuuntelija, kysymysten esittäjä ja sosiaalisen ilmapiirin aistija. Taito sovitella ristiriitoja ja tarkastella kipeitä kysymyksiä rakentavasti on arvokas. Arvioijan pitää huomata kuunnella kaikkia etu- ja sidosryhmiä tasapuolisen ja oikeudenmukaisen kokonaiskuvan rakentamiseksi. Itsensä ylittäminen liittyy hyvän tekemiseen ja haluun sitoutua palvelemaan arvioitavien kehitystä. Persoonan kannalta voitaisiin puhua Aristoteleen sanoin suurisieluisesta ihmisestä, joka ei ole liian herkkänahkainen ja joka osaa asettaa asioita oikeisiin mittasuhteisiin.

Miten kulkea kohti eettisesti kestävää arviointia?

Arviointietiikkaa on käsitelty kirjallisuudessa varsin vähän, vaikka sille olisi mitä ilmeisimmin tarve. Tutkimusetiikkaa koskeva keskustelu ja julkaisuutoiminta on sen sijaan ollut viime vuosina voimistumassa. Tästä keskustelusta voidaan ammentaa hyödyllisiä näkökohtia myös arviointiin arvottavana, kantaa ottavana ja tulkitsevana toimintana. Ulkoisten arviointien ja niiden toteuttajien kasvava määrä voi lisätä riskejä ylilyönteihin, joissa arvioitavan oppilaan, opettajan, kou-

lun, ohjelman tai koulutusjärjestelmän etuja poljetaan ja aiheutetaan vahinkoa tai vähintäänkin paha mieltä. Lopputuloksen pitäisi aina olla kehittämistä kannustava ja tukeva. Tällaisen arvioinnin toteuttamisessa ei riitä pelkästään metodinen ammattitaito, vaan myös eettinen kompetenssi ja sensitiivisyys nousevat arvoonsa.

Kehittämisen ja parantamisen näkökulma edellyttää, että arviointitietoa kerätään silloin, kun siitä on oikeasti hyötyä. Summatiivinen arviointi tehdään usein opetusjakson tai projektin päätteeksi, jolloin arvioinnilla on vain taaksepäin katsova feedback-merkitys. Englanninkielinen feed forward -sanonta kuvastaa hyvin, miten arvioinnin tulee antaa palautetta, josta on hyötyä tulevaisuutta ajatellen (Weeden ym. 2002). Prosessien aikainen välivaiheiden formatiivinen arviointi mahdollistaa sen, että yksilöiden tai organisaatioiden oppimisen siihenastisia vahvuuksia ja heikkouksia voidaan tunnistaa ja päätellä, millaisia tavoitteita jatkolle asetetaan ja miten niihin voidaan parhaiten päästä.

Arvon antaminen oikeudenmukaisesti, totuudellisesti ja huolenpitoa kunnioittaen on mahdollista vain silloin, kun arvioinnin tavoitteet on asetettu selkeästi ja yhteisesti. Arviointituloksia voidaan peilata verraten niitä tavoitteisiin, ja yhdessä voidaan pohtia seuraavia toimenpiteitä. Jos tavoitteita ei osata tai haluta määritellä, monet piilossa olevat ja samalla epäeettiset vaikutukset saavat sijaa. Tuloksia voidaan silloin käyttää toisiin tarkoituksiin kuin tietoja on alun perin kerätty.

Miten etiikan tarve voidaan arvioinnissa kiteyttää? Hyvän ja pahan tai oikean ja väärän pohdiskeluun ei yleensä jouduta silloin, kun kaikki ovat asiasta yhtä mieltä tai tilanne on ongelmaton. Eettisen pohdinnan tehtävä on tiedostaa ja analysoida eettisiä ongelmia, esittää ratkaisuja ja perustella ne sekä hahmottaa eettisiä suuntaviivoja ja kehittää moraalista tietoisuutta. Yhteiskunta tai koulutuspolitiikka ei tarjoa enää valmiita arvojärjestelmiä, joihin arvioija voi nojata työssään. Arvioijalla on oltava rohkeutta eettisiin pohdintoihin ja kannanottoihin. Eettinen pohdinta edistää samalla arvioinnin kriteereinä olevien arvojen tiedostamista ja niiden määrätietoista noudattamista sekä haastaa tarkastelemaan kriittisesti, kysymään oikeutusta ja kuuntelemaan omaatuntoa.

Lähteet

- Atjonen, P. 2004. Pedagoginen etiikka koulukasvatuksen karttana ja kompassina. Suomen kasvatustieteellinen seura. Kasvatusalan tutkimuksia 20.
- Atjonen, P. 2007. Hyvä, paha arviointi. Eettisiä kysymyksiä arvon antamisesta. Helsinki: Tammi.
- Bempechat, J., Jimenez, N. V. & Boulay, B. 2002. Cultural-cognitive issues in academic achievement. Teoksessa A. Porter & A. Gamoran (toim.) Methodological advances in cross-national surveys of educational achievement. National Research Council Washington, DC: National Academy Press, 117-149.

- Brooks, M. 2002. Assessment in secondary schools. A new teacher's guide to monitoring, assessment, recordings, reporting and accountability. Philadelphia: Open university.
- Castells, M. 1999. End of millenium. The information age: economy, society and culture. Vol. III. (revised ed.) Oxford: Blackwell.
- Chelimsky, E. 1997. The coming transformations in evaluation. Teoksessa E. Chelimsky & W. R. Shadish (toim.) Evaluation for the 21st century. A handbook. Thousand Oaks: Sage, 1–26.
- Coombes, M. 1992. The ethics of educational evaluation. *Social Alternatives* 11 (3), 39–42.
- Einhorn, S. 2005. *Konsten att vara snäll*. Stockholm: Forum.
- Farrell, A. (toim.) 2005. Ethical research with children. Maidenhead: Open university, 1–14.
- Fetterman, D. 1998. Empowerment evaluation: An introduction to theory and practice. Teoksessa D. Fetterman, S. Kaftarian & A. Wandersman (toim.) Empowerment evaluation. Knowledge and tools for self-assessment & accountability. London: Sage, 3–46.
- Fitzpatrick, J.L. 2005. Honesty. Teoksessa S. Mathison (toim.) *Encyclopedia of evaluation*. Thousands Oaks, Ca.: Sage, 187.
- Gardner, H., Csikszentmihalyi, M. & Damon, W. 2001 *Good work. When excellence and ethics meet*. New York: Basic Books.
- Grisay, A. 1984. Les mirages de l'évaluation scolaire. *Revue de la Direction Générale de l'Organisation des Études*, XIX, 8.
- Guiding principles for evaluators. American evaluation assessment. Saatavilla [www.muodossa: <http://www.eval.org/Publications/GuidingPrinciples.asp>](http://www.eval.org/Publications/GuidingPrinciples.asp). (Luettu 13.4.2007.)
- Hallamaa, J., Launis, V., Lötjönen, S. & Sorvali, I. (toim.) 2006. *Etiikkaa ihmistieteille*. Tietolipas 211. Helsinki: SKS.
- Hambleton, R. K. 2002. Adapting achievement tests into multiple languages for international assessments. Teoksessa A. Porter & A. Gamoran (toim.) *Methodological advances in cross-national surveys of educational achievement*. National Research Council Washington, DC: National Academy Press, 58–79.
- Hautamäki, J. & Kuusela, J. 2005. Mitataan oppilaita, mutta päätellään koulusta. Teoksessa H.K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnista*. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä, 229–241.
- Heikkinen, H. L. T., Huttunen, R. & Kakkori, L. 1999. ”Ja tämä tarina on tosi...” Narratiivisen totuuden ongelmasta. *Tiedepolitiikka* 24 (4), 39–52.
- House, E. R. 1980. *Evaluating with validity*. Beverly Hills, CA: Sage Publications.
- House, E. R. 1993. *Professional evaluation. Social impact and political consequences*. Newbury Park, CA: Sage.
- House, E. R. 1995. *Evaluoinnin futuurin perfekti*. Teoksessa S. Takala (toim.) *Arviointi ja koulutuksen laadun kehittäminen*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 23–32.
- House, E. R. 1997. Ethics of evaluation studies. Teoksessa J. P. Keeves (toim.) *Educational research methodology, and measurement: An international handbook*. 2. p. Cambridge: Cambridge University Press. Pergamon, 257–261.
- House, E. R. 2005. Deliberative democratic evaluation. Teoksessa S. Mathison (toim.) *Encyclopedia of evaluation*. Thousand Oaks, Ca.: Sage, 104–108.
- House, E. R. & Howe, K. R. 1999. *Values in evaluation and social research*. Thousand Oaks: Sage.
- Husén, T. 1986. *The learning society revisited*. Oxford: Pergamon.
- Hvistendahl, R. & Roe, A. 2004. The literacy achievement of Norwegian minority students. *Scandinavian Journal of Educational Research* 48 (3), 307–324.
- Häyry, M. 1999. Huomioita tutkimusetiikasta ”keskitason periaatteiden valossa”. Teoksessa S. Lötjönen (toim.) *Tutkijan ammattietiikka*. Opetusministeriö. Koulutus- ja tiedepolitiikan osaston julkaisusarja 69, 10–14.

- Juujärvi, S. 2003. The ethic of care and its development. A longitudinal study among practical nursing, bachelor-degree social work and law enforcement students. University of Helsinki. Department of Social Psychology. Social psychological studies 8.
- Kankaanranta, M., Grant, A. & Linnakylä, P. (toim.) 2007. e-Portfolio in lifelong learning. University of Jyväskylä: Institute for Educational Research.
- Kuula, A. 2006. Tutkimusetiikka. Aineistojen hankinta, käyttö ja säilytys. Tampere: Vastapaino.
- Launonen, L. 2000. Eettinen kasvatusajattelu suomalaisen koulun pedagogisissa teksteissä 1860-luvulta 1990-luvulle. Jyväskylä Studies in Education, Psychology and Social research 168.
- Linnakylä, P. 1998. Koulutuksen kansallisen ja kansainvälisen arvioinnin haasteista. Kasvatus 2 (29), 175–189.
- Linnakylä, P. & Välijärvi, J. 2005. Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulun kehittämiseksi. Jyväskylä: PS-kustannus.
- Mabry, L. 1997. Implicit and explicit advocacy in postmodern evaluation. Teoksessa R. Stake & L. Mabry (toim.) *Advances in program evaluation. Evaluation and the post-modern dilemma*. London: Jai press, 191–200.
- Malin, A. 2005. School differences and inequities in educational outcomes. PISA 2000 results of reading literacy in Finland. University of Jyväskylä. Institute for Educational Research. Research reports 15.
- Maw, J. 1999. League tables and the press – value added? *Curriculum Journal* 10 (1), 3–10.
- Morris, M. 2005. Ethics. Teoksessa S. Mathison (toim.) *Encyclopedia of evaluation*. Thousand Oaks, Ca.: Sage, 131–134.
- Newman, D. & Brown, R. 1996. *Applied ethics for program evaluation*. London: Sage.
- Noddings, N. 2003. *Happiness and education*. Cambridge: University press.
- Nuutinen, P. 2000. Opettajat vallassa ja vallan alla. Teoksessa J. Enkenberg, P. Väisänen & E. Savolainen. *Opettajatiedon kipinöitä. Kirjoituksia pedagogiikasta*, 177–189. Saatavilla www.muodossa.com: <<http://sokl.joensuu.fi/Verkkojulkaisut/kipinat/PDFt/PirjoN2.pdf>>. (Luettu 13.4.2007.)
- OECD. 2004. *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- Oravakangas, A. 2005. Koulun tuloksellisuus? Filosofisia valtuksia koulun tuloksellisuuden problematiikkaan suomalaisessa yhteiskunnassa. Jyväskylän yliopisto. Chydenius-instituutin tutkimuksia 2.
- Paechter, C. 2003. On goodness and utility in educational research. Teoksessa P. Sikes, J. Nixon & W. Carr (toim.) *The moral foundations of educational research. Knowledge, inquiry and values*. Maidenhead: Open University, 105–117.
- Patton, M. 1997. *Utilization-focused evaluation*. Thousand Oaks: Sage.
- Patton, M. 2002. *Qualitative research & evaluation methods*. 3rd ed. London: Sage.
- Peschar, J. 2004. Cross-curricular competencies: Developments in a new area of education outcome indicators. Teoksessa J. H. Moskowitz & M. Stephens (toim.) *Comparing learning outcomes. International assessments and education policy*. New York: Routledge Falmer, 59–20.
- Pring, R. 2003. The virtues and vices of an educational researcher. Teoksessa P. Sikes, J. Nixon & W. Carr (toim.) *The moral foundations of educational research. Knowledge, inquiry and values*. Maidenhead: Open university, 52–67.
- Rawls. 1971. *Theory of justice*. Oxford: University Press.
- Räsänen, R. 2002. Arvot, opettajuus ja opettajankoulutus valtaviiran ja moninaisuuden ristiaallokossa. Teoksessa R. Räsänen, J. Jokikokko, M-L. Järvelä & T. Lamminmäki-Kärkkäinen (toim.) *Interkulttuurinen opettajankoulutus. Utopiasta todellisuudeksi toimintatutkimuksen avulla*. *Acta Universitatis Ouluensis E* 55, 15–30.
- Simons, P. 2000. Damned if you do, damned if you don't. Ethical and political dilemmas in evaluation. Teoksessa H. Simons & R. Usher (toim.) *Situated ethics in educational research*. London: Routledge & Falmer, 39–55.

- Standards for educational and psychological testing. 1999. Washington, D.C.: American Educational Research Association.
- Stufflebeam, D. 1997. A standard-based perspective on evaluation. Teoksessa R. Stake & L. Mabry (toim.) *Advances in program evaluation. Evaluation and the postmodern dilemma*. London: Jai press, 61–88.
- The joint committee on standards for educational evaluation. Saatavilla [www-muodossa : <http://www.wmich.edu/evalctr/jc/>](http://www.wmich.edu/evalctr/jc/). (Luettu 13.4.2007.)
- The student evaluation standards 2003. Joint Committee on Standards for Educational Evaluation. Thousand Oaks, CA: Corwin.
- Vuorenmaa, M. 2001. Ikkunoita arvioinnin tuolle puolen. *Jyväskylä Studies in Education, Psychology and Social Research* 176.
- Weeden, P., Winter, J. & Broadfoot, P. 2002. *Assessment. What's in it for schools?* London: Routledge Falmer.



Riippumattomuus arviointitoiminnan periaatteena

Riippumattomuus on ulkoisen arviointitoiminnan keskeinen periaate. Periaatteen toteutumisen arviointi on osa arviointitoimijoiden itsearviointia ja ulkoisen arviointitoiminnan arviointikohteita. Koska arviointien toteuttaminen on aina jossain määrin sidoksissa erilaisiin konteksteihin, on riippumattomuutta arvioitava siitä näkökulmasta, onko riippumattomuus toteutunut arviointitoiminnassa ja sen piirissä toteutetuissa hankkeissa riittävässä määrin. Arvioijat eivät voi toimia riippumattomina esimerkiksi suhteessa niihin arvoihin, jotka ovat sisäänrakennettuja kohdetoimintaan ja sitä säätelevään ohjaukulttuuriin. Artikkelissa tarkastellaan muun muassa niitä tilanteita, joissa riippumattomuuden toteutumista pidetään tärkeänä. Riippumattomuus tukee arviointitoiminnan keskeisiä arvoja, joita ovat esimerkiksi oikeudenmukaisuus, yhdenvertaisuus ja totuudellisuus. Riippumattomuus edistää arviointitoiminnan uskottavuutta ja eettisesti kestävää toimintaa. Artikkelissa tarkastellaan riippumattomuuden sisältöä, säädösperustaa, toteutumisen ehtoja ja toteutumattomuuden haittoja. Riippumattomuus nähdään erityisesti arviointitoiminnalle ominaisena metodologisena periaatteena ja arviointien laatua parantavana tekijänä.

Asiasanat: riippumattomuus, ulkoinen arviointi, toimintaperiaate, arvot

Arvioinnilla ja tutkimuksella hankitaan tietoa koulutuksesta. Molemmissa on olennaista, mitä, miten, milloin, kenelle ja missä tarkoituksessa tietoa tuotetaan. Tiedon tuottamisessa menetelmällinen pätevyys ja luotettavuus ovat tärkeitä. Kun arviointitietoa tuotetaan julkishallinnolle ja edistämään muun muassa kansalaisten perusoikeuksien toteutumista, kuten sivistyksellisiä perusoikeuksia, on vaatimus tiedon tuottamisen riippumattomuudesta otettava huomioon.

Arviointeihin liittyvät yleensä arvioijien käsitykset kehittämistarpeista sekä näkemykset kehittämistä edellyttävistä toimenpiteistä. Arvottaminen kehittämissuosituksineen tekee arvioinnista erityisen intressiherkän. Eri intressit voivat aiheuttaa sen, että arvioijiin pyritään vaikuttamaan. Ne voivat myös aktivoida kohteena olevia ja kohteesta vastuussa olevia päätöksentekijöitä. Arvioinnin eettisiin periaatteisiin sitoutuneet arvioijat eivät voi esimerkiksi käydä kauppaa siitä, mitä he sanovat tai jättävät sanomatta.

Arviointien perustana olevan tiedon tulee olla luotettavin menetelmin tuotettua ja totuudellista. Ulkopuolinen arviointi ei voi tapahtua hallinnon, toiminnassa mukana olevien tai sen kehittämistä vastuussa olevien ehdoilla. Avoimuus, läpinäkyvyys ja osallistavuus ovat välttämättömiä arvioinnin kehittämistehtävän onnistumiseksi. Jakku-Sihvosen ja Hämäläisen (2000, 14) mukaan arvioinnin kohteiden ja perusteiden tulee olla julkisia ja arvioitavien tiedossa.

Arvioinnin riippumattomuuden takaaminen voi olla ongelmallista. Kaakinen (2004) pelkistää arvioinnin uskottavuuden perustuvan viime kädessä arvioinnin eettisiin periaatteisiin ja niiden johdonmukaiseen noudattamiseen sekä arviointiyhteisön kehittyneeseen itsekontrolliin. Temmes (2004) puolestaan korostaa teoreettisen ja kehittyneen arviointikäytännön kannalta arvioinnin ulkopuolisuutta ja itsenäisyyttä suhteessa kohteeseen ja tilaajaan.

Täysin riippumatonta arviointia ei voi olla olemassa. Esimerkiksi niin sanottu riippumaton lehti ei voi toimia ikään kuin julkaisijaa, mainostajia, toimittajia, lukijoita ja päätoimittajaa tai heidän tarpeitaan ei olisi olemassa (Angell 2000). Onkin tarkasteltava erityisesti sitä, milloin riippumattomuus on asianomaisen toiminnan kannalta riittävää. Milloin riippuvuus voi olla jopa haitallista arvioinnin onnistumiselle? Mitä erityisiä vaatimuksia riippumattomuudelle on asetettava arviointitoiminnassa?

Riippumattomuutta voidaan tarkastella muun muassa arviointiorganisaatioiden itsensä määrittelemänä. Toimivatko ne määrittelemiensä toimintaperiaatteiden mukaan? Miten arviointitoiminnan riippumattomuus voi toteutua julkishallinnossa niin, että arviointi palvelee samalla hallintoa tai on osa sitä?

Riippumattomuuden toteutumista voidaan tarkastella esimerkiksi arvioinnin valintatilanteissa. Perusvalinnat tehdään arviointijärjestelmää luotaessa, kun määritellään arvioinnin arvoja, tavoitteita ja toimintaperiaatteita. Ne ovat osa arvioinnin metodologista perustaa. Ohjaavien arvojen ja periaatteiden tulisi olla myös jatkuvan kriittisen tarkastelun kohteina – johdetaanhan menetelmävalin-

nat osin arvoista ja periaatteista käsin. Esimerkiksi valmiiden mittareiden käyttäminen merkitsee myös mittareiden arvoihin ja tausta-ajatteluun sitoutumista.

Arvioinnin kytkeminen kehittämisprosesseihin merkitsee aina jossain määrin tinkimistä riippumattomuudesta. Kehittävässä arvioinnissa tämä seikka tulee esiin selvästi. Lisäksi Suomen kaltaisessa pienessä maassa henkilökohtaisiin suhteisiin perustuvat riippuvuudet voivat olla joskus merkityksellisempiä kuin muodolliset riippuvuudet (ks. Kaakinen 2004). Arvioijien ja arviointiorganisaation arvoriippuvuuksien läpinäkyvyyden parantaminen on osa arviointieettisen tiedostamistason kohottamista.

Arviointitoiminnan valintatilanteita liittyy muun muassa päätöksiin arvioinnin arvoista ja toimintaperiaatteista, arviointitoiminnan organisoinnista ja resursoinnista, arviointiohjelmista ja -kohteista, arviointihankkeiden tavoitteista, niiden laajuudesta ja ajoituksesta, arviointimenetelmistä ja -mittareista, arvottamisen kriteereistä ja tasovaatimuksista, arvottamisen mekanismeista ja arvioivista päätelmistä sekä kehittämistarpeiden ja -suositusten valinnoista. Kysymys riippumattomuudesta liittyy arviointitoiminnan kaikkiin prosesseihin ja valintoihin.

Arviointijärjestelmän tarkastelu toisiinsa sidoksissa olevina päätöksentekovaiheina antaa riippumattomuuden kuvaamiselle yleiskehikon, koska se lisää valintoja ohjaavien perusteiden läpinäkyvyyttä. Riippumattomuuden toteutumista arvioinnissa voidaan tarkastella myös osana julkisen hallinnon ja ohjauksen vastuullista toimintaa. Tavallisimmin puhe riippumattomuudesta rajoittuu keskusteluun, jossa tarkastellaan arvioinnin riippumattomuutta vain sen hallintosektorin piirissä, johon arvioitavat kohteet kuuluvat.

Riippumattomuus ilmiönä ja riippumattomuuden käsite

Riippumattomuutta tarkastellaan tässä lähinnä kansallisen, niin sanotun koulutuksen ulkopuolisen arvioinnin (ks. PL 21 §) kannalta. Yhteiskunnallisessa keskustelussa riippumattomuus rinnastetaan muun muassa puolueettomuuteen, autonomisuuteen, liittoutumattomuuteen ja lahjomattomuuteen.

Oikeuslaitoksen riippumattomat tuomioistuimet edustavat riippumattomuutta lainsäädäntövallasta ja toimeenpanovallasta. Jos koulutuspolitiikka on arvioinnin kohteena, on riippumattomuus toimeenpanovallasta tärkeää. Se ei ole kuitenkaan täysin mahdollista, joten arviointitoiminnan asema koulutuksen hallinnossa tulisi määritellä riittävän tarkasti.

Hallintolain 6 §:ssä (434/2003) mainittu puolueettomuus liittyy hallinnon oikeusperiaatteisiin. Arviointitoiminnassa ei kuitenkaan puhuta puolueettomuudesta, koska arviointia ei rinnasteta viranomaistoimintaan, jossa tehtäisiin päätöksiä lakeihin ja säädöksiin perustuen tai valmisteltaisiin niitä. Puolueettomuus

liittyy arviointiin erityisesti silloin, kun tasa-arvo ja oikeudenmukaisuus nähdään keskeisiksi arvoiksi.

Lääkärinvalassa (Saarni & Henriksson 2005, 9) riippumattomuus korostuu silloin, kun vedotaan muun muassa tasapuolisuuteen, totuudellisuuteen ja professionaalisuuteen. Valassa todetaan muun muassa: ”Täytän lääkärin velvollisuuteni jokaista kohtaan ketään syrjimättä enkä uhkausestakaan käytä lääkärintaivani ammattietiikkani vastaisesti. Käytän vain lääketieteellisen tutkimustiedon ja kokemuksen hyödyllisiksi osoittamia menetelmiä.” Arvioijalla on metaforisesti pitkälti samat vaatimukset omassa työssään.

Liiketoiminnassa riippumattomuus liittyy erityisesti päätöksenteon oikeudenmukaisuuteen. Lahjonta voi vaikuttaa valintoihin, päätöksiin ja toimintaan eri tavoin. Korruptio on päätöksentekokulttuurissa moraalisesti kyseenalaista. Kyse on oman aseman väärinkäytöstä eri tavoin etujen saamiseksi itselle ja tai omalle lähipiirille.

Lobbauksen, lahjonnan tai sensuurin arvellaan tuottavan etuja sen harjoittajille suoraan tai välillisesti. Hyöty saattaa olla esimerkiksi etua kilpailijoihin nähden tai onnistumista jo tehdyn virheen salaamisessa.

Tiedonvälityksessä riippumattomuus on demokraattisessa yhteiskunnassa avainasia. Sensuurilla vallanpitäjä pyrkii suodattamaan ja valikoimaan julkistettavia tietoja ja kannanottoja. Kansallisten arviointien keskeiset tulokset (PL 2003, 21§) tulee julkistaa. Tätä seikkaa korostaa myös vaatimus arvioinnin avoimuudesta ja oikeudenmukaisuudesta (asetus 150/2003). Velvollisuus arvioida itse omaa toimintaa ja saattaa oma toiminta myös ulkopuolisten tahojen arvioitavaksi tukee arviointitoiminnan avoimuutta.

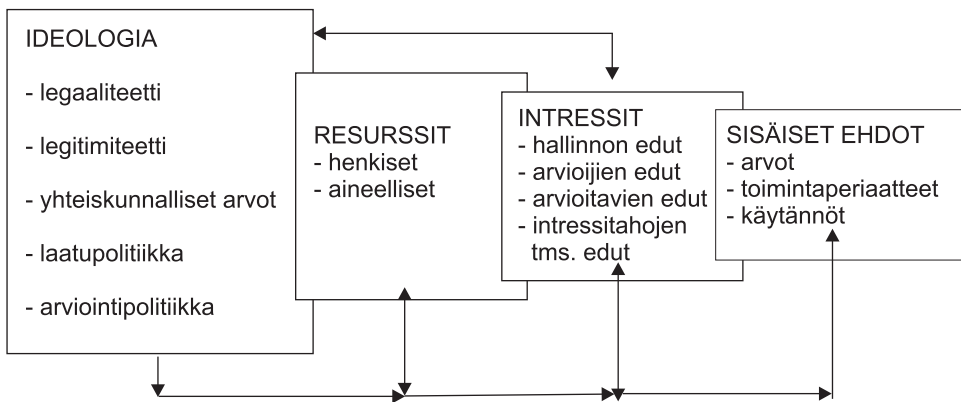
Yliopistojen tutkimustoiminnan vapaus, toiminnan luotettavuuden vahva imago ja hallinnollinen autonomia rinnastetaan usein arkipuheessa riippumattomuuteen. Tutkimustoiminnalta odotetaan myös objektiivisuutta. Tämä ajattelu tapa riippumattomasta kontekstista oli mukana, kun tehtiin päätös Koulutuksen arviointisihteeristön sijoittamisesta Suomessa hallinnollisesti Jyväskylän yliopiston yhteyteen vuonna 2003.

Arvioinnin yhteydessä institutionaalisista normeista ei kuitenkaan voida vielä puhua samalla tavoin kuin tieteen yhteydessä, vaikka arviointi jakeakin universaalisuuden, tiedon yhteisomistuksen ja pyyteettömyyden normit tieteen kanssa. Arviointi kuitenkin etsii vielä rooliaan tiedon järjestelmällisenä epäilijänä (vrt. esim. Tuunainen 2004). Kehittämistyön tukemisen näkökulmasta saattaa epäilyn painottaminen arvioinnissa olla myös heikkous, mutta se voi tukea myös riippumattomuutta profession vahvistumisen kautta.

Arviointitoiminnassa riippumattomuutta voidaan kuvata esimerkiksi seuraavien tekijöiden kokonaisuutena. 1) Tuodaanko asiat esiin niin kuin ne ovat (avoimuus, totuudellisuus)? 2) Onko toiminta eri tahojen näkökulmasta oikeudenmu-

kaista ja tasa-arvoista? 3) Luottavatko asiakkaat ja yhteistyökumppanit siihen, että toiminta on lainsäädännön sekä toimijan ilmoittamien arvojen ja tavoitteiden mukaista? 4) Onko toiminta korkealaatuista ja järjestelmälliseen kehittämiseen perustuvaa? 5) Onko toiminta avointa ja ulkopuolisten tahojen arvioitavissa?

Edellä sanotun perusteella riippumattomuuden vaatimus on moniulotteinen. Kokonaisuuden tarkastelussa on hyvä ottaa huomioon kuvion 1 mukainen jatkuva ja valintojen kokonaisuus, jonka sisällä riippumattomuus voi toteutua kontekstissaan eri tavoin.



Kuvio 1. Arvioinnin valintoja määrittävät tekijät julkishallinnossa

Riippumattomuus on yksi Koulutuksen arviointineuvoston toimintaperiaatteista (Koulutuksen arviointineuvosto 2004). Arviointitoiminnan edellytyksiin riippumattomuus liittyy siitä näkökulmasta, luovatko toiminnan resurssit, rakenteet ja arviointipoliittiset ohjauspäätökset aidosti edellytyksiä tuoda näkyviin se, mikä pitääkin. Toivottavat tulokset ovat usein yhteydessä siihen, kenen näkökulmasta asiaa tarkastellaan. Tämä tarkoittaa esimerkiksi sitä, ettei ole olemassa esteitä tarkastella koulutusjärjestelmän vahvuuksia ja heikkouksia niin kuin ne ovat. Tämä edellyttää, että koulutusjärjestelmän panoksia, prosesseja ja tuloksia voidaan järjestelmällisesti valottaa muun muassa trenditiedon ja keskeisten kohdealueiden (resurssit, prosessit, tulokset) keskinäisten yhteystarkastelujen avulla. Edellä todetun perusteella riippumattomuuden periaatteen tulisi ulottua strategisesta päätöksenteosta aina yksittäisten hankkeitten tulosten julkistamiseen ja levittämiseen saakka.

Optimaalinen riippumattomuus arviointitoiminnan ulkopuolisista toimijoista ei välttämättä takaa riippumattomuuden toteutumista arvioinnin sisäisessä toiminnassa. Avoimuuden ja riippumattoman ulkopuolisen arvioinnin puuttuessa arviointitoimijakin voi joutua oman sisäisen kulttuurinsa loukkuun. Toiminta-

kulttuurista voi tulla ”ahdas takki”. Omat normit voivat kapeuttaa dynaamista liikkumatilaa. Riippumattomuudelle tulisi luoda edellytyksiä selkeillä säädöksillä ja oman toiminnan jatkuvalla kehittämisellä.

Arviointitoiminnan ja arvioinnin riippumattomuudella tarkoitetaan tässä sitä, missä määrin arviointia virallisesti harjoittava taho, sitä edustava ryhmä, virkamies tai asiantuntija voi toimia arvioinnin arvojen, toimintaperiaatteiden ja tavoitteiden mukaisesti siten, että toiminnasta ja sen tuloksista seurauksina pääteltävissä olevat haitat tai niistä saatavat edut eivät vaikuta valintoihin. Riippuvuuksien arviointi ja niiden eliminointi on tärkeää järjestelmätasolta hanketoiminnan yksityiskohtiin saakka.

Ulkopuolisuuden vaatimus arvioinnissa saattaa olla puute silloin, kun kohdetta tulisi ymmärtää syvällisesti myös sisältä. Tällöin mukaan saataisiin monipuolista asiantuntemusta ja sitoutettaisiin arvioitavat tai muut sidosryhmät arvioinnin keinoin mukaan kehittämistoimintaan. Ulkopuolisuuden vaatimus voidaan ottaa tällöin huomioon osittain – ainakin metodisesti. Ulkopuolisuus on riippumattomuuden kannalta etu, mutta kohteen sisäisen ymmärtämisen ja asiakkaiden sitouttamisen näkökulmasta myös puute. Menetelmällisesti tämä tarkoittaa tarvetta kehittää itsearviointi ja ulkopuolisen arvioinnin metodista integrointia.

Arvioinnin sisällä voi korostua pyrkimys lisätä omaa valtaa, tulkita vallan rajat yksilön, ryhmän tai organisaation kannalta yksipuolisesti ja indoktrinoitua näin osaksi muodostuneita valtajärjestelmän reagoitaitapumuksia. Tämä riippuvuus on organisaation sisäistä riippuvuutta, jonka havaitseminen ja tunnistaminen voi olla äärimmäisen vaikeaa ilman avointa vuorovaikutusta esimerkiksi arviointiasiantuntijoiden verkostoissa.

Riippumattomuuden säädöspohja

Arviointitoiminnan riippumattomuus nousi tärkeänä asiana esiin eduskunnalle annetussa koulutuksen arviointia koskevassa selvityksessä (2003). Myös opetusministeriön asettama Koulutuksen arviointijärjestelmän kehittämistyöryhmä (2007, 48) kiinnitti huomiota arvioijien riippumattomuuteen ja Koulutuksen arviointineuvoston aseman riippumattomuuteen.

Kun kansalliselta ulkoiselta arvioinnilta edellytetään avoimuutta, luotettavuutta ja oikeudenmukaisuutta (A 150/2003, 3§), on kysymys edellytyksistä, jotka eivät voi toteutua täysin ilman riittävää riippumattomuutta. Hallintolaissa (434/2003) esteellisyyssäädökset tukevat riippumattomuuden toteutumista erityisesti toiminnan ja prosessien sisäisen riippumattomuuden kannalta.

Koulutuksen arviointineuvosto (2004, 16) katsoo, että sen järjestämän arviointitoiminnan riippumattomuuden edellytyksenä on riippumaton asema ope-

tushallinnosta ja koulutuksen järjestäjistä, riittävät taloudelliset toimintaedellytykset ja tunnustettu asema arvioijana. Koulutuksen arviointineuvoston (emt. 16) mukaan arvioinnin vaikuttavuus edellyttää, että koulutuksen arvioitsija on riippumaton opetushallinnosta (koulutuksen päätöksenteosta) ja koulutuksen kehittämiseen osallistuvista tahoista sekä riippumaton niiden sosiaalisista järjestelmistä.

Koululainsäädäntöämme uudistettaessa sivistysvaliokunta katsoi (SiVM 3/1998 vp, 20–21), että valtakunnallisen koulutuksen tuloksellisuuden arvioinnin järjestämisessä on keskeistä, että arviointia suorittaa luottamusta herättävä, puolueeton, toiminnassaan autonominen ja asiakaslähtöinen, monipuolinen ja kansainvälistä yhteistyötä harjoittava asiantuntijayksikkö. Valiokunnan mielestä arviointien organisoinnilla tulee huolehtia arvioinnin kaikinpuolisesta luotettavuudesta ja toimivuudesta.

Koululainsäädännössä (esim. PL 21 §) koulutuksen arvioinnin tarkoituksiksi määriteltiin koululainsäädännön tarkoituksen toteuttamisen turvaaminen, koulutuksen kehittäminen ja oppimisen edellytysten parantaminen (esim. PL 21 §). Asetus koulutuksen arvioinnista (150/2003, 2§) toteaa sanottua täydentäen, että koulutuksen arvioinnin tavoitteena on hankkia ja analysoida tietoa valtakunnallisen ja paikallisen kehittämistyön ja päätöksenteon pohjaksi sekä tukea opiskelijoiden oppimista, opetushenkilöstön työtä ja oppilaitosten kehittämistä. Mainittujen säädösten arvioinnille määrittelemä tarkoitus ja tavoitteet edellyttävät arvioinnilta ulkopuolisuutta arvioinnin kohteena olevasta toiminnasta, toimijoista ja kehittäjistä. Velvoite kansallisiin arviointeihin osallistumisesta on ainoa koulutuksen järjestäjiä koskeva velvoite ulkoisissa arvioinneissa.

Kun koululainsäädännössä puhutaan toisaalta ulkopuolisesta arvioinnista (PL 2003, 21 §) ja toisaalta ulkopuolisesta arviointitoiminnasta (asetus 150/2003, 4§), on ulkopuolisuudella haluttu korostaa riippumattomuutta siitä toiminnasta tai organisaatiosta, johon arviointi kohdistuu (vrt. Opetushallitus 1998, 14). Koululaeissa todetaan myös, että koulutuksen järjestäjän tulee arvioida antamaansa koulutusta (esim. PL 2003, 21§), jolloin on taas kyse itsearviointista.

Opetushallitus on koulutuksen järjestäjien toimintaa arvioidessaan osin samassa asemassa riippumattomuuden kannalta kuin koulutuksen järjestäjä on toimeenpannessaan oppilaitoksien arviointeja. Koulutuksen järjestäjä ei ole riippumaton, kun se arvioi hallinnoimiaan oppilaitoksia. Siitä huolimatta paikallinen arviointitoiminta täydentää kansallista ulkopuolista arviointia ja on tärkeä osa organisaatioiden omaa laadunhallinnan kokonaisuutta.

Koulutuksen arviointineuvoston yhteys ministeriöön todetaan asetuksessa koulutuksen arvioinnista (A 150/2003, 4§) seuraavasti: ”Ulkopuolisesta arviointitoiminnasta ja sen kehittämisestä vastaavat opetusministeriö ja sen yhteydessä asiantuntijaelimenä toimiva koulutuksen arviointineuvosto.” Koulutuksen arviointineuvostolla on näin ollen selvä tehtävä osana opetusministeriön hallinnon

kokonaisuutta. Hallinnon sisäisenä toimintana tilanne ei ole perusteltavissa riippumattomuuden kannalta.

Koulutuksen arviointineuvosto (2004, 13) toteaa, että riippumattomuuden taakamisessa on tärkeää, ettei hallinnolla ja päätöksentekojärjestelmällä ole käsyyvaltaa suhteessa arviointiorganisaatioon ja sen voimavaroihin, arviointiohjelmaan eikä arviointiprosesseihin. Näin ollen on keskeistä, mitä riippuvuus ministeriöstä merkitsee rajoituksina, mahdollisuuksina ja toteutuvina käytänteinä.

Riippumattomuuden arvopohjaa

Arviointitoiminnassa keskeisiä asioita ovat arvottaminen ja pyrkimys kehittämiseen. Arvottamisessa on kyse siitä, millainen arvo, hyöty tai merkitys arviotavaan kohteeseen liitetään. Arviointi on arvotonta, ellei se arvota tuloksiaan (esim. Scriven 1997). Arvon antamisen edellytykset muodostuvat prosessuaalisesti arviointitoiminnan valintojen ketjussa. Keskeistä näissä prosesseissa on se, millä perusteilla arvioija valitsee arviointinsa kriteerit, millainen arvo tai merkitys kohteelle annetaan kriteereittäin ja mitkä arvot vaikuttavat yleensä arvioijien valintoihin. Riippumattomuutta voidaan tarkastella esimerkiksi edellä mainittujen valintojen vapauden ulottuvuudella, tilaajan määrittämien riippumattomuusehtojen sekä arvioijien omien arvioinnin eettisten periaatteiden näkökulmista.

Arvon määrittämisen logiikka ja mekanismit ovat avainkysymyksiä arvioinnin luotettavuuden ja vertailtavuuden kannalta. Esimerkiksi tiedot ja näytöt, jotka ohjaavat arvottamista ja kehittämisuositusten laadintaa, vaihtelevat usein painoarvoltaan ja merkitykseltään arvioijittain. Monitahoisessa arvioinnissa arvottamisprosessiin tulee mukaan arvojen kirjo ja niiden välisen tasapainon etsiminen. Kysymys on konsensukseen pyrkimisestä. Monitahoisuus harmonisoi painotuseroja ja tarjoaa foorumin arvoja koskevaan aitoon dialogiin. Tällöin voivat kuitenkin tulla ongelmalliseksi erilaiset jääviystilanteet ja arvioivien päätelmien löysä yhteys tuloksiin.

Arvovalinnat perusteluineen ovat kulttuuri- ja kontekstisidonnaisia. Arvioinnit tapahtuvat monia arvoja edustavassa kokoonpanossa. Arvioijien tausta voi olla hyvinkin kirjava hallinnosta eri tieteenalojen edustajiin saakka. Arviointien avoimuuteen kuuluu kuitenkin, että on voitava kuvata, analysoida ja arvioida niitä tekijöitä, jotka vaikuttavat ja ovat mukana valinnoissa. Erilaisten riippuvuuksien osoittaminen on osa tämän tietoisuuden rakentamista.

Riippumattomuus liittyy Koulutuksen arviointineuvoston arvoihin (2004, 15–16) analysoimalla vastauksia seuraaviin kysymyksiin: tukeeko arviointijärjestelmä yhdenvertaisuuden toteutumista (tasa-arvo), onko arviointitieto luotettavaa ja tieteellisesti perustelluin metodein tuotettua (totuudellisuus) sekä kohdellaanko arvioitavia tahoina ja yksilöinä samalla tavalla (oikeudenmukaisuus)?

Koska arviointi on yksi kansallinen tukipalvelu koulutuksen kehittämiseksi, toiminnalla tuetaan tasa-arvon ja koululainsäädännön tarkoituksen toteutumista. Arviointitehtävä voi sisältää myös sen toteamisen, onko toimittu lainsäädännön mukaan, mutta toisaalta arvioinnin tehtävä on myös luoda edellytyksiä lainsäädännön kehittämiseksi.

Riippuvuuden haitat korostuvat silloin, jos se taho, josta arviointi on riippuvainen, saa siitä etua ja vastaavasti jokin toinen vastaava taho ei etua saa. Arvioinnin haitoista tai jopa niiden kumuloitumisesta voidaan puhua silloin, jos arviointijärjestelmä mekanismina ei tuekaan tasa-arvoa, oikeudenmukaisuutta ja totuudellisuutta. Toiminnan arvoperusta pettää.

Arvoja koskeva asiantuntijuus on arvioijien ydinosaamista (ks. Lyytisen ja Nikkasen artikkeli tässä julkaisussa). Erilaisten hallinnollisten, kulttuuristen, valtapoliittisten ja kontekstuaalisten arvojen näkyväksi tekeminen ja niistä käytävä jatkuva keskustelu on arvioijien työtä. Arvioijien ja arvioinnin arvoja tulee myös arvioida. Ilman arvioinnin avoimuutta voivat erilaiset riippuvuutta ylläpitävät tekijät muodostua osaksi arviointijärjestelmän kyseenalaistamatonta ja legitiimiä kulttuuria.

Osa arvioinnin arvoista saattaa jäädä piiloarvoiksi. Kyky tiedostaa niitä kuuluu jokaisen arvioijan asiantuntemukseen. Joskus saatetaan omaa arviointia jopa raivokkaasti puolustaa, vaikka ei kyetä osoittamaan edes arvioidun kohteen tärkeyttä (Vuori 2004, 2). Arvioijilla voi olla myös arviointi-ideologisia riippuvuuksia, jotka vaikuttavat sitoutumisiin, esimerkiksi yksittäisten mittareiden arvopohjaan, tai nämä sitoutumiset tulevat ilmi intresseinä erilaisissa valinnoissa.

Arviointien seurauksena yhteisöjen sekä instituutioiden merkitys ja kilpailuasetelmat saattavat muuttua. Arvioinneissa käytetyistä kriteereistä, niiden painotuksista ja tuloksina kerrotuista hyödyistä ja arvomäärittelyistä voidaan olla myös eri mieltä. Arvot ja arvioinnit ovat subjektiivisia. Toista arvoa ei voida objektiivisin perustein osoittaa toista arvoa arvokkaammaksi (arvosubjektivismi). Arviointi on myös keino pitää esillä koulutuksellisten arvojen läpinäkyvyyttä monipuolisesti. Tällöin kehittämiseen vaikuttaminen tarkoittaa esimerkiksi arvokeskustelun rikastamista ja vahvistamista sekä arvoketjujen puutteiden osoittamista.

Arvojen analysointi, niistä käytävä keskustelu sekä arvojen näkyväksi tekeminen toimintakulttuurissa on arvioijien ammatillista perustyötä. Arvot vaikuttavat valintoihin ja ovat mukana erilaisina opittuina valintataipumuksina. Laajempi yhteiskunnallisen arvomaailman tuntemus on arvioijille näin ollen välttämätöntä. Arvioija ei ole riippumaton vallitsevan yhteiskunnan arvoista. Tätä tiedostamista kohottamalla riippuvuutta voidaan myös paremmin kontrolloida.

Riippumattomuuden toteutumisen edellytyksistä

Jakku-Sihvonen ja Hämäläinen (2000, 12) katsovat, että ulkoinen arviointi vaatii kehittämistä juuri hallinnollisesti riippumattoman tahon suorittamana. Koska täydellistä riippumattomuutta ei ole, kyse on siitä, milloin riippumattomuuden katsotaan täyttävän sille asetetut hallinnolliset vaatimukset, olevan legitiimiä ja eettisesti kestävää sekä riittävää toiminnan tarkoituksenmukaisuuden kannalta. Arvioinnille ominaisten menettelytapojen noudattaminen arvioinnin eri vaiheissa on arvioinnin itsenäisyyden kannalta olennaista (Temmes 2004).

Arvioijan autonomialla on rajansa. Halu rajoittaa autonomiaa arvioinnin ulkopuolelta ja toisaalta halu poiketa autonomiasta arvioinnin sisäisessä toiminnassa on osoitus myös yhteiskunnassa vallitsevien erilaisten arvojen synnyttämästä dynamiikasta. Siksi riippumattomuuden toimintaperiaatteen käytännöllinen sisältö on arvioijalle merkityksellinen toiminnan eettisestä näkökulmasta ja tärkeä myös arvioinnin teoreettisen ja metodologisen kehittämisen kannalta.

Kansallisessa arviointitoiminnassa korostetaan riippumattomuutta hallinnon sisäisistä prosesseista ja sen omasta ohjausvaikutuksesta. Lyytisen (2004) mukaan hallintokulttuurista riippuvuutta kuvaavat esimerkiksi hallinnon valta- ja valvontaorientaatio, hierarkkinen ohjaus, säilyttävyyden ja rajoja määrittävä luonne sekä hallinnon omien intentioiden ja omien toimintatapojen tietoinen ja tiedostamaton käyttö arviointikriteerinä. Erityistä huomiota hallinnon ohjauksessa tulisi kiinnittää siihen, että arviointitieto on mahdollisimman pitkälle riippumaton niiden virkamiesten toiminnasta, joilla on vastuu toiminnan kehittämisestä (Jakku-Sihvonen & Hämäläinen 2000, 13). Sulkunen (2003, 120) osoittaa kyseenalaiseksi muun muassa sen intressisyyden, jossa sama taho rahoittaa sekä kehittämishankkeet että niiden arvioinnin.

Organisoitaessa arviointitoimintaa asiantuntijoiden verkostona (esim. PL 2003, 21 §) lisääntyvät edellytykset hyödyntää monitahoisuutta, monitieteisyyttä ja erilaisten näkökulmien yhteensovittamista, mikä tukee myös riippumattomuuden toteutumista. Lyytisen (2004) mukaan asiantuntijaverkoston riippumattomuudessa korostuu muun muassa 1) professionaalinen kehittämisorientaatio, 2) näkökulmien monipuolisuus ja ulkoisuus, 3) verkoston sisäisiä intressejä tasa-painottava vaikutus ja 4) usko asiantuntija-auktoriiteettiin. Riippumattomuutta tuetaan näin arviointitoiminnan organisoimisen monipuolisuudella, monitahoisuudella ja sen mukanaan tuomalla metodisella laaja-alaisuudella. Useimmissa EU-maissa onkin perustettu riippumattomia tai osittain riippumattomia yksiköitä koulutuksen arviointiin. Nämä yksiköt sijoittuvat julkisen vallan ja oppilaitosten välimaastoon (Jakku-Sihvonen & Hämäläinen 2000, 13).

Jos riippumattomuus määritellään säädöksissä yksiselitteisesti, ei hallinnon valtajärjestelmälle jää tilaa omalakiselle arvioinnin säätelyyn. Arviointikäsitteiden osittainen epämääräisyys, johon viitattiin jo hallituksen esityksessä eduskun-

nalle koulutusta koskevaksi lainsäädännöksi (HE 1997/86), on myös esimerkki mahdollisuudesta käyttää arviointikäsitteistön vakiintumattomuutta hyväksi omia näkemyksiä ajattaessa tai toimivaltaa edelleen allokoitaessa.

Koska riippumattomuus voi toteutua parhaimmillaankin vain jossain määrin, on kysyttävä, missä määrin on riittävässä määrin. Arvoiltaan selkeä arviointijärjestelmä, avoin toimintakulttuuri ja arvioinnin riippumattomuuden rajat selkeästi osoittavat säädökset ja arviointitoiminnan laadunhallinnan tarkoituksenmukaisuus tukevat riippumattomuuden toteutumista.

Riippuvuutta aiheuttavat tekijät liittyvät myös tiedon tuottajaan itseensä, sen ohjausvaltaan ja yhteyksiin, arvioinnin kohteena oleviin tahoihin sekä arvioinnin sidos- ja intressiryhmiin. Tyypillinen riippuvuus syntyy erityyppisestä asianosaisuudesta, intressisyydestä suhteessa kohteisiin tai niiden parissa toimiviin tahoihin tai ohjausvallan tarpeista ohjata sektorinsa kaikkea toimintaa. Arviointitoiminnan sisäisiä riippuvuustekijöitä voidaan luokitella seuraavasti:

1) Ohjaus- tai kumppanuusriippuvuus

Arvioinnin ulkopuolinen taho määrittelee, mitä, miten, missä laajuudessa ja milloin arvioidaan tai arvioija osallistuu arviointitoimintaa ohjaavaan päätöksentekoon; riippuvuus saattaa toteutua erityisesti pysyvämässä ohjaus- tai kumppanuussuhteessa.

2) Intressiriippuvuus

Arvioivan yksikön asiantuntijoille tai arvioinnin päätöksentekoon osallistuville asiantuntijoille (tai heidän taustayhteisölleen) voi olla etua tai haittaa siitä, mitä arvioidaan, miten arvioidaan tai milloin arvioidaan.

3) Asianosaisriippuvuus

Arviointi liittyy asianomaiseen, hänen etuunsa, oikeuteensa tai velvollisuuteensa esimerkiksi toiminnan normittajana, hallinnoijana, kehittäjänä tai resursoijana.

4) Asiakas- tai edustusriippuvuus

Arvioinnin kohteena oleva laitos, kohteena olevat ihmiset tai kohteena olevat asiat ovat arvioijaan nähden edullisessa asemassa poliittisesti, maailmankatsomuksellisesti, syntyperän, kielen tai muun syyn vuoksi.

5) Taloudellinen riippuvuus

Arviointitaho, -yksikkö tai arviointiasiantuntija on toiminnassaan ja sen perusteella ilmoitetuissa kannanotoissaan riippuvainen tahosta, joka on tilannut tai rahoittaa toimintaa.

Luottamuksen ilmapiirissä arvioinnin kohteena olevat tahot, hallinto mukaan lukien, luottavat arvioijien tietoon, tahtoon ja kykyyn noudattaa säädöksiä ja omalle toiminnalle hyväksytyjä periaatteita. Tällöin arviointitoiminnasta saadut viestit ja kannanotot eivät leviä kentälle ristiriitaisina, yllättävinä tai perustele-

mattomina. Arviointitoiminnan laadulla ansaittu ja vakiinnutettu asema antaa mahdollisuuden hyödyntää virallisia oikeuksia täysimääräisesti.

Riippumattomuuden toteutumisen uhkaskenaariona voidaan pitää toimintaa, jossa samat asiantuntijat laativat arvioinnin arvot, arvottamisen kriteerit, kouluttavat arvioijat, valitsevat ideologiansa mukaiset arvioijat ja osallistuvat itse arviointien suunnitteluun, tietojen keruuseen, prosessien ohjailuun, tulosten arvottamiseen sekä tulosten julkistamiseen ja levittämiseen. Arviointitaho voi tulla näin toimiessaan ennen pitkää sokeaksi omalle kulttuurilleen. Arvioijan suhde taastaansa, koulukuntiin, näkökulmiin ja professioihin on tarpeen tiedostaa (Temmes 2004).

Toisentyypistä uhkaa riippumattomuudelle edustaa se, jos arvioijien keskuudessa kriittisyys, kyseenalaistaminen ja pohdiskeleva keskustelu jäävät näennäiseksi, luokitellaan normittomuudeksi, vastustamiseksi tai keskustelua ei nähdä arvioinnin kehittämistä tukevaksi dialogiksi.

Riippumattomuus arvioinnin laatuvaatimuksena

Riippumattomuuden psykologisia taustavoimia ovat arvioijien oma itsekriittisyys, halu kehittää osaamistaan ja ammattieettisiä arvoja koskeva tinkimättömyys. Riippumattomuutta edistää myös jatkuva oman toiminnan kehittäminen, itsearviointi ja ulkopuolisen arvioinnin kohteeksi asettuminen sekä oman toiminnan vaikuttavuuden systemaattinen seuraaminen.

Riippumattomuuden uhkiksi voidaan nähdä uudistumisen puute, avoimen dialektisen keskustelun niukkuus sekä itsekritiikin vähäisyys ja sisäisen dynamiikan heikkous. Kyse on tällöin myös sisäisen organisoitumisen ja valtasuhteiden käytännöistä – usein indoktrinaatiosta, johon saatetaan ajautua esimerkiksi vahvan johtajuuden puuttuessa tai liian dominoivan vallan olosuhteissa. Verkostomaisuus, monitahoisuus ja kansainvälisyys ovat tällaisen kehityksen eräitä vastavoimia.

Arviointi ei ole kehittämislle normeja tai tahtotiloja suoraan luova institutio. Se on lainsäädäntö-, toimeenpano- ja tuomiovallan ulkopuolinen taho, joskus tarkkailija, totuudenpuhuja, ennakoija ja kriittinen analyttikko. Päätehtävä on kuitenkin kehittämisen tukeminen. Arvioinnin tulee vaikuttaa tietoperustaisesti päätöksentekijöihin, kehittäjiin ja yhteiskunnan päätöksentekokulttuuriin. Se on arvioinnin vahva intentio (esim. PL 2003, 21 §; asetus 150/2003, 2 §). Siitä syntyy arvioinnin vaikuttavuus.

Arviointitoiminnan riippumattomuus on koetteella silloin, kun mahdollisimman suuren vaikuttavuuden aikaansaaminen edellyttää myös interventioita kehittämisessä arvioinnin keinoin. Siitä ei ole enää pitkä matka arvioijien ke-

hittämistarpeiden ja kehittämissuosituksen toteutumisen käytännön tukemiseen. Silloin on jo selvästi astuttu rajan yli, riippuvuuden puolelle.

Arviointiin suuntautuvat vaikuttamisyrietykset, tulivatpa ne virallisen valtajärjestelmän kautta tai intressitahojen, sidosryhmien tai yhteistyökumppanien taholta, ovat luonnollisia. Kysymys onkin siitä, miten arviointijärjestelmä kykenee eliminoimaan näiden pyrkimysten intressisyyden ja toisaalta siitä, miten arvioijat kykenevät itse toimimaan yhteiskunnallisesti ajan hermolla. Oman toiminnan laadunhallinta, monitahoarviointi, monitieteisyys, arviointitoiminnan vaikuttavuuden jatkuva seuraaminen ja verkostomainen toimintakulttuuri luovat edellytyksiä, joiden kautta riippumattomuuden vaatimus toteutuu.

Lähteet

- Angell, M. & Farewell, A. 2000. *The New England Journal of Medicine* 2000; 342 (26). Massachusetts Medical Society. URL: <http://www.nejm.org>
- Asetus 20.2.2003/150. Valtioneuvoston asetus koulutuksen arvioinnista.
- Bingham, R. D. & Felbinger, C. L. 1989. *Evaluation in practice. A methodological approach*. New York & London: Longman.
- Hallintolaki 6.6.2003/434.
- He 1997/86. Hallituksen esitys Eduskunnalle koulutusta koskeväksi lainsäädännöksi.
- Hirsjärvi, S. (toim.) 1992. *Kasvatustieteen käsitteistö*. Helsinki: Otava.
- House, E. 1990. Evaluation with validity, values in evaluation and social research. *Professional evaluation. Ethics of evaluation studies. New Directions in Educational Evaluation* (toim.) Decentralised evaluation for Norway 1990.
- Jakku-Sihvonen, R. & Hämäläinen, K. 2000. *Laatua koulutuksen laatu politiikkaan. Opetushallitus*. Moniste 11/2000. Helsinki.
- Kaakinen, J. 2004. Arvioinnin paikka. Arvioitsija intressien ristipaineissa. *Hallinnon tutkimus* 23 (2), 94–99.
- Koulutuksen arviointijärjestelmän kehittämistyöryhmän muistio 2007. Opetusministeriön työryhmämuistioita ja selvityksiä 2007: 27.
- Koulutuksen arviointineuvosto. 2004. *Koulutuksen arvioinnin uusi suunta. Koulutuksen arviointineuvoston julkaisuja 1*. Jyväskylä.
- Lyytinen, H. K. 2004. Hallintokulttuurinen riippuvuus. *Opetusmoniste*. Jyväskylän yliopisto.
- Opetushallitus 1998. *Koulutuksen tuloksellisuuden arviointimalli*. Helsinki.
- Patton, M. Q. 1987. *How to use qualitative methods in evaluation. Qualitative research & evaluation methods. Utilization focused evaluation. Practical evaluation*. Newbury Park, CA: Sage.
- Perusopetuslaki 21.8.1998/628.
- Saarni, S. & Henriksson, M. (toim.) 2005. *Lääkäriin etiikka*. 6. painos. Suomen Lääkäriliitto.
- Scriven, M. S. 1997. Truth and objectivity in evaluation. Teoksessa E. Chelimsky & W. R. Shadish (toim.). *Evaluation for the 21st century*. Thousand Oaks, Ca: Sage, 477–500.
- SiVM. *Sivistysvaliokunnan mietintö 3/1998 vp*. 20–21. Liittyy hallituksen esitykseen koulutusta koskeväksi lainsäädännöksi 86/1997.
- Sulkunen, P. 2003. Onko arviointitutkimus asiantuntijavaltaa vai sen puutetta? *Sosiologia* 2, 115–120.

- Temmes, M. 2004. Arviointityön autuus ja salahaudat arvioijan näkökulmasta. Hallinnon tutkimus 23 (2), 86–93.
- Tuunainen, J. 2004. Akateemisen tutkimuksen vapaus muutosten puristuksessa. Saatavilla [www-muodossa: <http://www.edu.helsinki.fi/activity/people/tuunaine/>](http://www.muodossa: <http://www.edu.helsinki.fi/activity/people/tuunaine/>).
- Vuori, J. 2004. Arviointi(tutkimus) on arvotonta ellei se arvota arviointiaan. Hallinnon-tutkimus 23 (2), 1–3.



Arviointi, tutkimus ja arviointitutkimus koulutuksen tietotuotannossa

Artikkelissa tarkastellaan arvioinnin, tutkimuksen ja arviointitutkimuksen ominaispiirteitä ja niitä toisistaan erottavia näkökohtia muun muassa tiedon käyttötarkoituksen, teorian roolin, ongelmanasettelun, analyysien ja tulosten käytettävyyden kannalta. Arvioinnin funktioita esitellään tulosvastuun, tiedontuotannon ja kehittämisen kehyksissä sekä pohditaan vaikuttavuutta myös suhteessa tutkimukseen. Tutkimusta ja arviointia vertaillaan arvottamisten, teemallisen vapauden ja ajallisen perspektiivin näkökulmista. Artikkelissa osoitetaan, että tuotettavan tiedon, käytetyn metodologian ja tulosten hyödynnettävyyden osalta tutkimus, arviointi ja arviointitutkimus ovat lähentyneet toisiaan muun muassa kasvatusalan metodisen paradigman murroksen seurauksena. Myös näitä kaikkia koskevat laatuvaatimukset ovat kasvaneet toimintavolyymin lisääntyessä.

Asiasanat: arviointi, tutkimus, arviointitutkimus, metodologia, vaikuttavuus, arviointi- ja tutkimustehtävä

Koulutuksen arvioinnissa kohtaavat oppilaat ja opettajat, koulutuksen järjestäjät, opetushallinto, tutkijat ja poliittiset päättäjät. Samat henkilöt voivat toimia välillä arvioitavina, välillä arvioijina tai arvioinnin järjestäjinä tai tilaajina. Intressit ja tiedon tarve vaihtelevat, samoin arviointimenetelmät. Oppilas, opettaja,

koulutuksen järjestäjä tai koulutuspoliitikko tarvitsee tai otaksuu tarvitsevansa erilaista ja eri tavoin koottua tietoa. Samat henkilöt eri rooleissa voivat muuttaa odotuksiaan: rehtoriksi siirtyvä opettaja tai opettajaksi ryhtyvä koulutuspoliitikko joutuvat hankkimaan erilaista arviointitietoa kantaakseen vastuunsa uudessa tehtävässään.

Arviointitietoa tarvitaan moneen tarkoitukseen. On sekä inhimillistä että tärkeää tietää, miten oma toiminta, oppiminen, opetus, koulun johtaminen, hallinnointi tai koulutuspolitiikka onnistuu. Yhtä lailla halutaan tietoa siitä, mihin suuntaan ja millä tavalla tulisi oppimista, opetusta tai hallinnointia kehittää, jotta se olisi yhteisesti sovittujen arvojen ja sekä yksilöiden että kansakunnan menestyksen kannalta onnistunutta. Hyvän koulutusjärjestelmän ylläpitäminen vaatii niin ulkoista objektiivista kuin sisäistä monimenetelmäistä arviointia, jotta kehitystä voidaan vauhdittaa tuloksellisesti ja innostavasti. Arvioinnilla on siksi sekä tietoa kokoava ja palautetta antava (feedback) että eteenpäin suuntaava ja tulevaisuutta rakentava (feed forward) funktio.

Miten arvioinnin monitasoisuus ja -tahoisuus ilmenevät?

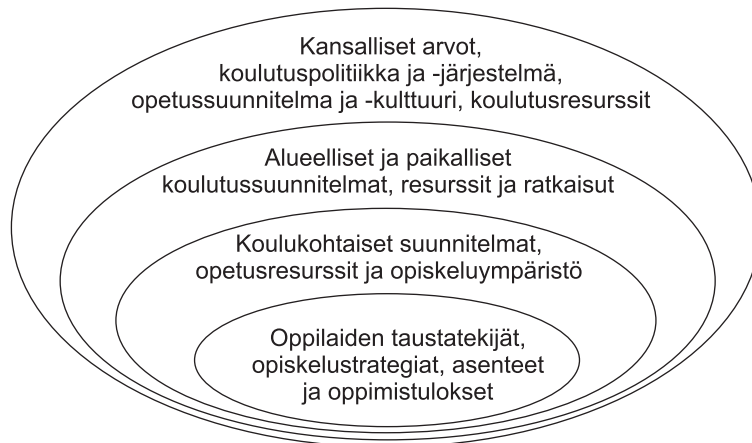
Koulutuksen arvioinnin ytimessä on oppilaiden tiedollisten, asenteellisten ja sosiaalisten oppimistulosten arviointi (engl. assessment). Oppimistuloksia on kuitenkin vaikea ymmärtää ja selittää, ellei tiedetä, miten oppilaat oppivat ja miten heitä opetetaan. Pedagogiset ratkaisut taas määräytyvät koulun oppimisympäristöstä, resursseista ja opetussuunnitelmasta, joihin paikallinen opetuksen järjestäjä luo mahdollisuuksia tai asettaa rajoituksia. (Atjonen 2005.)

Koulutuksen arviointi voidaan ymmärtää laajasti myös koko koulutusjärjestelmän, koulutuspolitiikan ja opetussuunnitelmien toteutumisen palaute- ja kehittämisympäristöksi (engl. evaluation). Tällä tasolla arviointi kohdistuu koulutusjärjestelmän rakenteisiin sekä opetuskulttuurin laatuun ja tuloksiin. Laatu-kriteereinä voidaan pitää opetussuunnitelman tavoitteita tai koulutuspoliittisia päämääriä, joiden toteutumisesta tarkastellaan suhteessa yhteisesti sovittuihin arvoihin, resursseihin ja rakenteisiin.

Koulutuksen laatua on vaikea ilmaista absoluuttisesti, joten se ilmaistaan yleensä suhteellisesti, joko muihin arvioitaviin tai ennalta sovittuihin kriteereihin verraten. Näin laatu voidaan ymmärtää esimerkiksi erinomaisuutena tai huippuosaamisena, virheettömyytenä tai tasaisuutena, heikkojen vähäisenä määränä, tulosten paranemisena, ohjelmien tai koko järjestelmän tehokkuutena tai taloudellisuutena (Harvey & Green 1993). Erinomaisuus näkyy muihin verrattuna parhaana tuloksena tai vaativimpien laatu-kriteerien täyttämisenä. Oppimistulosten tasaisuus taas ilmaistaan yleensä tulosten ja etenkin ääripäiden vaihteluna. Koulutuksen ja opetuksen tasaisuus osoitetaan yleensä koulutukseen pääsyn ja

oppimismahdollisuuksien tasa-arvon toteutumisena. Laadun tasaisuuteen liittyy läheisesti heikkojen tulosten vähäinen määrä tai ”riittävän hyvän” tason ylittäminen. Tulosten parantamisen osoittamiseen tarvitaan useita arviointikertoja tai jatkuvaa seurantaa. Tällöin lähtötaso ei ole olennaisin vaan muutoksen suuruus. Tehokkuuden ja taloudellisuuden kriteerit liittyvät läheisesti toisiinsa. Laatu on taloudellisesti tuloksellisin ja vaikuttavin koulutus ja opetus (Harvey 2002; Weeden ym. 2002).

Vaikka arviointi kohdistuisi ensi sijassa koulutusjärjestelmään ja opetussuunnitelmien ja ohjelmien toteutumiseen, todellista vaikuttavuutta seurattaessa on arvioitava myös kansallisen koulutuksen traditioita, koulujen ja opettajien päätösvaltaa, toimintakulttuuria sekä opettajien koulutusta, palkkausta ja statusta yhteiskunnassa (kuvio 1).



Kuvio 1. *Koulutuksen arvioinnin laaja kenttä*

Kuvio 1 muistuttaa, että vain yhden tason arviointitiedon ymmärtäminen ja selittäminen on usein vaikeaa, ellei sitä tarkastella arvioinnin laajemmassa kentässä sosiokulttuurisena ilmiönä (ks. Bempechat ym. 2002; LeTendre 2002). Esimerkiksi suomalaisten lasten hyvää lukutaitoa on vaikea ymmärtää, ellei samalla saada arviointitietoa myös lasten lukuharrastuksesta ja sen kehittämisestä koulussa ja kotona, kirjojen ja lehtien saatavuudesta, kirjaston käytöstä, koti- ja koulukulttuurista, erityisopetuksesta ja sen resursoinnista, opetussuunnitelman painotuksista tai tuntijaosta.

Mitä arvioinnilta odotetaan?

Arviointeja voidaan tarkastella myös niihin kohdistuvien odotusten ja intressien näkökulmista, jotka liittyvät läheisesti arvioinnin tarkoituksiin ja tahoihin, arvovalintoihin ja metodologioihin sekä ennen muuta tiedon hyödyntämiseen koulutuksen kehittämisessä (vrt. Leimu 2004).

Koulutuspoliittiset intressit liittyvät yhteiskunnan koulutukselle asettamiin yleisiin tavoitteisiin ja pyrkimyksiin. Suomessa koulutuksen arvioinnin keskeisin koulutuspoliittinen pyrkimys on seurata koulutuksellisen tasa-arvon toteutumista. Silloin kiinnostus kohdistuu erityisesti oppilaiden, eri sukupuolten, koulujen, alueiden sekä sosioekonomisten ryhmien tasa-arvoisiin koulutus- ja oppimismahdollisuuksiin. Lisäksi arviointia halutaan opetuksen tasosta sekä koulutuksen tehokkuudesta ja taloudellisuudesta. Kansainvälisissäkin arvioinneissa koulutuspoliittiset intressit ovat paljolti samoja, vaikka painotukset vaihtelevat (vrt. Baker 2004).

Joskus kansalliset etupyrkimykset kytkeytyvät koulutuspoliittisiin uudistuksiin, jolloin arvioinnilta toivotaan tietoa esimerkiksi uudistusten taloudellisista ja pedagogisista vaikutuksista. Hallinnolliset intressit painottavat erilaisten hallinnollisten ratkaisujen, esimerkiksi päätösvallan hajauttamisen, arviointia. Kulttuuriselta kannalta voidaan selvittää esimerkiksi oppisopimuskoulutuksen, erityisopetuksen tai maahanmuuttajaoppilaiden koulutuksen ja opetuksen järjestämistä (Linnakylä ym. 2006; Roe & Hvistendahl 2006). Historiallinen päähuomio saatetaan kohdentaa erilaisten koulunuudistusten etenemiseen ja seuraamusten arviointiin. Enenevästi tarvitaan tulevaisuusnäkökulmaa niin opetussuunnitelmien, oppimistuloksien kuin opiskelukäytäntöjen punninnassa esimerkiksi työelämän ja aktiivisen kansalaisuuden sekä kestäväen kehityksen asettamien odotusten kannalta (Schleicher 2004).

Arvioinnin pedagogiset intressit liittyvät opetuksen kehittämiseen. Ne kohdistuvat ainakin opetussuunnitelmien toteutumisen sekä opiskelumenetelmien ja oppimateriaalien käytön arviointiin. Opetussuunnitelmien arvioinnissa huomio voi kohdistua vaikkapa oppisisältöjen jäsentymiseen oppiaineiksi tai aihekokonaisuuksiksi. Oppimistavoitteiden ja standardien tarkkuusastetta on vertailtu ja arvioitu eri aikoina sekä kansallisesti että kansainvälisesti (Lam 2004; Linnakylä 2004; Scheerens 2004).

Viime aikoina on erityisesti toivottu arviointitietoa erilaisia oppijoita, sekä heikkoja että lahjakkaita, innostavien opetusmenetelmien soveltuvuudesta ja vaikuttavuudesta. Kouluhyvinvointi ja siihen liittyen kouluilmasto ja -kiusaaminen sekä opettajien ja oppilaiden väliset suhteet ja oppilaiden sitoutuminen opiskeluun ovat olleet kansallisen arviointihuomion kohteina. Kansainvälisen arvioinnin kiinnostus on suuntautunut erityisesti erilaisten opiskelustrategioiden ja -asenteiden arviointiin sekä tieto- ja viestintätekniikan hyödyntämiseen opetuk-

nessa. Opettajien koulutus, asema, itsenäisyys, solidaarisuus ja päätösvalta sekä näitä tukevat ja vaarantavat tekijät ovat olleet vertailevan arvioinnin kohteina (ks. Leimu 2004; OECD 2001 ja 2003). Kaikkien edellä mainittujen tekijöiden arviointiin on parhaimmillaan liittynyt sekä oppimistulosten että käytössä olevien voimavarojen arviointia sekä trianguloivaa palautetiedon keruuta.

Arvioinnin metodologiset intressit ovat myös monitahoisia (ks. Hambleton ym. 2005; Porter & Gamoran 2002). Arviointien pätevyyden kannalta on tärkeä rakentaa arvioinnille teoreettisesti vahva arviointikehys, joka kertoo sekä arvioitaville että ulkopuolisille, missä käsitemaailmassa liikutaan. Arviointikehys ilmaisee selkeästi esimerkiksi sen, mitä matematiikan osaamisella tai kouluhyvinvoinnilla juuri kyseisessä arvioinnissa tarkoitetaan ja miten arviointikohde jäsennetään niin, että sitä voidaan systemaattisesti, pätevästi ja luotettavasti arvioida. Jos matematiikan osaaminen ymmärretään opetussuunnitelman mukaisesti, myös tuloksia on tulkittava tästä näkökulmasta. Jos taas matematiikan osaaminen määritellään koulun ulkopuoliseksi matematiikan soveltamiseksi arkitilanteissa, tuloksia tulisi tarkastella tältä kannalta eikä puhtaan matemaattisen ajattelun tai kouluopetuksen näkökulmasta.

Joskus metodologinen mielenkiinto kohdistuu erilaisten tausta- ja kehystekijöiden mallintamiseen. Kansainvälisissä arvioinneissa tarvitaan erityisesti tietoa sekä arviointikehysten että mittavälineiden kulttuurisesta sopivuudesta, sillä vertailtavuuden mahdollistamiseksi käsitteellistäminen tai mittaustapa ei saisi suosia osallistujamaita eri tavalla (Bonnet 2002). Kulttuuristen vääristymien arvioimiseksi ja välttämiseksi sekä käännösten ekvivalenssin takaamiseksi onkin viime vuosina tehty monenlaista arviointitutkimusta (esim. Arffman 2004; Sul-kunen 2004). Esimerkiksi Arffmanin (2007) tuoreen väitöskirjan mukaan PISA-kokeiden käännökset ovat pääosin luotettavia ja vaikeustasoltaan yhdenvertaisia.

Metodologisesti erityisen tarpeellista olisi verrata ja arvioida erilaisten menetelmien käyttöä saman aiheen arvioinnissa. Mitä uutta ja erilaista esimerkiksi laadullinen arviointi tai tapaustutkimus tuo määrällisen arvioinnin tulosten ymmärtämiseksi? Miten sosiologinen ja psykologinen näkökulma eroavat kouluhyvinvoinnin tai koulutuksen tuottaman inhimillisen pääoman arvioinnissa? Miten oppimistulokset eroavat, jos otannan perusjoukkona on ikäluokka tai luokka-aste? Miten otannan edustavuus voidaan eri tavoin varmistaa? Suosivatko monivalintatehtävät poikia ja avoimet tuotostehtävät tyttöjä? Miten avoimien koetehtävien luotettavuus ja vertailtavuus taataan?

Metodologiset intressit voivat näin vaihdella hyvin laajoista tieteidenvälisistä ongelmista spesifeihin metodisiin kysymyksiin. Laajat arviointiaineistot tarjoavat varteenotettavan mahdollisuuden myös metodologiselle tutkimukselle (esim. Lie ym. 2004; Mejding & Roe 2006). Esimerkiksi PISA:n kaltaisissa laajoissa arvioinneissa maavertailuja tehdään melko yksinkertaisten tunnuslukujen valossa,

joten niiden syventäviin analyysihin kannattaisi keskittyä erillisissä jatkotutkimuksissa.

Myös *arvioinnin käytännölliset intressit* tietysti vaihtelevat. Yksi ratkaisua vaativa kysymys on aina arviointien ajoitus esimerkiksi eri koulutusvaiheisiin, kouluvuoteen, opetusjaksoon tai työpäivään. Huomiota pitää kiinnittää erilaisiin rajoituksiin, esimerkiksi voimavaroihin, arvioinnin toteuttajiin, ajankäyttöön ja erityisoppilaiden osallistumiseen. Monipuolista selvittämistä vaativat myös testimateriaalien sisältö ja muoto sekä tietokoneiden, laskimien, sanakirjojen ja muiden työvälineiden käyttö. Arviointitilanteen käytännön organisointi pitää mieltä tarkasti tulosten laadun varmistamiseksi: Miten koulut tai oppilaitokset tavoitetaan? Miten yhteistyö koulujen kanssa organisoidaan? Millaiset koeohjeet suullisesti tai kirjallisesti annetaan oppilaille?

Erityisesti viime vuosina on pyritty yhä enemmän painottamaan tiedon välitystä ennen arviointia, arviointiprosessin aikana ja sen jälkeen. Sähköinen viestintä on helpottanut arvioijien yhteydenpitoa arvioitaviin. Arviointitiedon monipuolisessa jakamisessa on myös vahvistunut käyttäjien intressien huomioiminen ja hyötyvaatimus niin tiedontuotannon, jatkokehittämisen kuin tutkimuksen välineenä. Palautetieto on erityisen tärkeää arviointiin osallistuneille, joille tietoa tulisi jakaa monipuolisesti ja vuorovaikutteisesti ja erilaisia esitysmuotoja kokeillen niin, että arviointitieto ymmärrettäisiin hyödylliseksi oman työn kehittämisen näkökulmasta. Laajat arviointiraportit eivät välttämättä avaudu kiireiselle kouluväelle. Keskustelevat työpajat ja seminaarit ovat usein tehokkaampia arviointitiedon omaksumisessa ja hyödyntämisessä.

Mihin tarkoitukseen arviointitietoa tuotetaan?

Arvioinnilla tulisi aina olla selkeä tarkoitus, josta sekä arvioijat että arvioitavat ovat tietoisia ja jonka he myös hyväksyvät (House 2005). Arvioinnin tarkoitus määrittää paljolti sen, mihin arviointi kohdistuu, mitä ja ketä arvioidaan, millaisia menetelmällisiä ratkaisuja valitaan, miten arviointi käytännössä toteutetaan ja miten arviointitietoa käytetään jatkossa.

Arvioinneilla on erilaisia tarkoituksia. Osa niistä suuntaa katseen menneeseen ja painottaa tulosten ja tulosvastuun näkökulmaa. Osa taas preferoi tulevaisuutta, ja päähuomio on ensisijaisesti kehittämispyrkimyksessä (Linnakylä 1974; 1998). Tarkoitusten erilaisuus havainnollistuu hyvin Chelimskyn (1997) erittelemistä arvioinnin kolmesta päätyypistä: tulosvastuuarvioinnista, tiedontuotannosta ja kehittämisarvioinneista (taulukko 1).

Tulosvastuuarvioinnissa huomio kohdistuu ensisijaisesti koulutuksen tuloksiin, vaikuttavuuteen, tehokkuuteen ja taloudellisuuteen, joista etenkin päättäjät, kouluhallinto ja koulutuksen järjestäjät ovat kiinnostuneita. Tulostietoa käyte-

Taulukko 1. *Kolme näkökulmaa arviointiin Chelimskyä (1997, 21) mukaillen*

	Tulosvastuu	Tiedontuotanto	Kehittäminen
Tarkoitus	tulosten, taloudellisuuden ja tehokkuuden mittaaminen	ymmärryksen syventäminen ja yhteyksien selvittäminen	järjestelmän parantaminen, opetuksen ja oppimisen kehittäminen
Metodologia	selvitykset, laadulliset ja määrälliset arvioinnit, indikaattorien rakentaminen	arviointitutkimukset, tutkimukset, meta-analyysit	laadulliset arvioinnit, itsearviointit, näytöt, toimintatutkimukset, ohjelma-arvioinnit
Tyypillinen käyttö	poliittisessa ja hallinnollisessa päätöksenteossa	tutkimuksessa, koulutuksessa ja tietolähteenä, indikaattorien pohjana	uudistusten tukemisessa, julkisessa poliittisessa keskustelussa
Arvioijan ja arvioitavan suhde	etäinen	etäinen tai läheinen riippuen metodista	läheinen, arvioija on ”kriittinen ystävä” ja osavastuussa kehittämisestä
Riippumattomuus	välttämätön	ehdoton	suhteellinen
Edunvalvonta	ei hyväksyttävä	ei hyväksyttävä	usein väistämätön
Arvioitavien hyväksyntä	neuvoteltava, usein vaikea saavuttaa	neuvoteltava, tulokset voidaan tarkistaa ja kiistää	helppo saavuttaa, sillä ei uhkaa
Objektiivisuus	melko vahva	vahva	epävarma
Asema julkisessa keskustelussa	vahva, saa julkisuutta	melko vahva, jos saa julkisuutta	epävarma

tään poliittisessa ja hallinnollisessa päätöksenteossa, johon sen toivotaan päätyvän mahdollisimman helposti ja nopeasti (Paechter 2003). Arvioija on tavallisesti etäinen ja riippumaton, luotettavan ja objektiivisen tiedon etsijä, joka ei itse ota kantaa poliittiseen keskusteluun, vaan kokoaa ja arvioi usein olemassa olevaa tietoa. Tyypillisiä ovat eri hallinnonaloilla tehtävät raportit ja selvitykset, joissa käytetään usein laadullisesti eritasoisia tietolähteitä, joita analysoimalla arvioidaan tilannetta ja tehdään johtopäätöksiä toiminnan kehittämiseksi. Selvitykset voivat arvioida esimerkiksi lasten ja nuorten aamu- ja iltapäivätoimintaa (Svedlin ym. 2002), oppilaskuntatoimintaa (Kauppinen ym. 2006), laboratorioalan koulutusta (Mäkelä 2006) tai koulutuksen paikallisen arvioinnin tilaa (Löfström ym. 2005).

Tiedontuotantoarvioinnin päätarkoitus on koulutusjärjestelmien, opetusohjelmien ja opetuskäytäntöjen ymmärtäminen sekä selitysmallien rakentaminen. Tämä liittyy tiedontuotannon läheisesti muuhun koulutustutkimukseen. Ongelmanasettelu lähtee tutkijoista ja kiinnittyy aiempaan tutkimukseen tai teoreettiseen keskusteluun. Arviointitietoa käytetään jonkin koulutuksellisen ilmiön,

ratkaisun tai yhteyksien selvittämiseen sekä erilaisten indikaattorien rakentamiseen. Arvioija pyrkii luotettavaan, pätevään ja yleistettävään tietoon, jota voi käyttää laajemminkin tutkimustarkoituksiin. Siksi arviointikehykseen ja tutkimusasetelmaan, otannan edustavuuteen ja mittavälineiden laatuun kiinnitetään erityistä huomiota. Arvioijalta odotetaan riippumattomuutta ja objektiivisuutta, koska se vahvistaa tulostiedon arvoa päätöksenteossa ja suunnittelussa.

Kehittämisarviointi palvelee koulutuksen uudistajien ja järjestäjien tarpeita ja pyrkii parantamaan koulutusjärjestelmää, opetussuunnitelmia ja -menetelmiä. Arviointi voi toimia joustavana työkaluna, jonka avulla selvitetään joko etu- tai jälkikäteen organisaatioiden, suunnitelmien tai menettelytapojen toimivuutta ja tuloksellisuutta. Arvioija on pikemminkin konsultin tai kriittisen ystävän roolissa, jolloin hän toimii läheisessä vuorovaikutuksessa tai yhteisessä kehittämissyhmässä, jossa arvioitavia osallistetaan kriittiseen itsetarkasteluun. Näin arviointi ei välttämättä ole kovin riippumatonta, objektiivista tai yleistettävää, vaikka se voi olla hyvinkin uskottavaa ja sillä saattaa olla huomattavaa merkitystä koulutusta tai opetusta kehitettäessä. (Chelimsky 1997; Kushner 2005; Räisänen 2005.)

Vaikka arviointityyppi ja -metodologia vaikuttavat arvioinnin hyödynnettävyyteen, Chelimsky (1997) painottaa, etteivät eri näkökulmat ole täysin toisiaan poissulkevia. Tiedontuotantoa voidaan edistää myös tulos- ja kehittämisarvioinneilla, ja joskus kehittämisarvioinnin tulosten perusteella käydään pedagogista tai poliittista keskustelua.

Miten arviointi voi vaikuttaa?

Yksi arviointien yleismaailmallinen ongelma kasvatus- ja yhteiskuntasektorilla on, etteivät ne tahdo saada aikaan käytännön seurauksia – kerätty tieto ei ole vaikuttavaa. Arvioinnin valottamia muutostarpeita vähätellään, laiminlyöntejä ei korjata eikä tietoa käytetä kehittämiseen. Vaikkapa lääketieteessä uusia hoitomenetelmiä koskeviin arviointitietoihin suhtaudutaan vakavasti, koska ihmisten terveydellä ja hengellä ei voi leikkiä. Koulutusosalalla tarkkoja vaikuttavuusarvioita on vaikea tehdä, koska Rajavaara (1999, 84) siteeraten ”yhteiskunnallisia ilmiöitä ja käytäntöjä ei voida viljellä laboratorioissa ja siirtää niistä takaisin yhteiskuntaan”.

Arvioinnin vaikuttavuutta voidaan Rajavaaran (1999) mukaan käsitellä niin sanottujen järkeilytyyliensä avulla. *Tavoitelähtöisessä järkeilytyylissä* ihannoidaan arvioinnin selkeyttä, täsmällisyyttä, konkreettisuutta, yksinkertaisuutta ja ymmärrettävyyttä. Sen huonona puolena on, ettei inhimillisiä kysymyksiä ole helppo taivuttaa määrällisen mitattavuuden muottiin. *Tarvelähtöiseen järkeilytyyliin* nojaavan arvioinnin tehtävänä on selvittää, vastaako vaikkapa jokin koulutus-

ohjelma tai -palvelu asiakkaiden tarpeita. Mukaan otetaan tyypillisesti arvioitavaan kohteeseen tai teemaan liittyvät sidosryhmät, jolloin vaikuttavuus yleensä paranee. *Kokeellinen järkeilytyyli* tarkoittaa, että vaikuttavuutta pyritään mittaamaan esimerkiksi koe- ja kontrolliryhmää vertaamalla. Oikeudenmukaisuudesta ja täsmällisyydestä huolimatta kokeellisuus pätee heikosti yhteiskunnallisiin kysymyksiin, joista kipeimpiin ei voida eettisistä syistä järjestää kokeita.

Mekanismeihin perustuvaa tyylä suosiva etsii erilaisia väliintulevia muuttujia vaikkapa koulutuspalvelujen hyödyn arvioinneissa. Mekanismin jäljittäminen pyrkii täyttämään sen tietoaution, jonka kokeellinen arviointi jättää sivuuttaessaan lopputulokseen vaikuttavat prosessitekijät. Osallistava, kehittävä, dialoginen ja voimaannuttava arviointi nojaavat *vuorovaikutteiseen järkeilytyyliin*. Tähän turvaututaan esimerkiksi erilaisissa kehittämis- ja toimintatutkimushankkeissa sekä itsearvioinneissa. Vuorovaikutteinen perustelu korostaa osallistujien panosta ja itsemääräämisoikeutta, jolloin arviointi mahdollistaa oppimisen.

Tieteellisiin vaikuttavuusnäyttöihin perustuva arviointien oikeutus näkee näytöt tärkeiksi, jolloin arvioinnin osuvuus ja pysyvyys ainakin periaatteessa kohenevat: jonkun koulutusohjelman tai opetusmenetelmän vaikuttavuudesta pitää olla selvästi osoitettavia todisteita. Yksi lähestymistapa on niin sanotut ”hyvät käytännöt”, joiden kuvaamisella ja levittämällä halutaan edistää tutkitun tai käytännössä koetellun tiedon hyväksikäyttöä. Tässä sekoittuu sekä pehmeä että kova arviointimetodologia kiinnostavalla ja vastaanottajiin uskottavuutta luovalta tavalla. *Auditoiva järkeilytyyli* tähdentää työntekijöiden valvonnan sijasta tulosten kautta tapahtuvaa valvontaa.

Vaikuttavuuden perustelutyylit muistuttavat arvioinnin tavoitteiden ja tarkoituksen tärkeydestä. Joskus niin arviointi kuin tutkimuskin lähtevät tarpeeseen vastaamisesta, toisinaan niiden pitää kohde huomioiden turvautua erityisesti vuorovaikutteiseen lähestymiseen, mikäli ne mielivät saada aikaan käytännön vaikutuksia. Monimutkaisissa, erilaisiin eturyhmiin kiinnittyvissä arvioinneissa ja tutkimuksissa taas voidaan joutua hankkimaan mahdollisimman täsmällisesti hankittua näyttöä tarkastelukohteen tilasta (evidence-based assessment).

Onko arviointi tutkimusta tai tutkimus arviointia?

Arvioinnin historiallinen kehitys (esim. Cook 1997; Raudasoja 2005; Vuorenmaa 2001) osoittaa, miten eri aikoina on arvostettu erilaisia arviointiparadigmoja. Kovin kaukana menneisyydessä eivät ole ajat, jolloin ”oikea arviointi” tai ”hyvä tutkimus” oli vahvasti kvantitatiivisen ja psykometrisen arviointisäännösten alaista systemaattista asiantuntijatyötä. Ihanteena oli ulkopuolinen, riippumaton ja objektiivinen kuvaus (vrt. edellä tavoitelähtöinen ja tieteellinen järkeilytyyli). Sittemmin on kuljettu kvalitatiivisen, osallistavan ja hyötyä korostavan arvi-

oinnin suuntaan. Näin erilaiset toiminta- ja kehittämistutkimukselliset otteet ja pehmeämmät tiedonkeruutekniikat (esim. ryhmähaastattelut ja -itsearviointit) ovat kasvattaneet suosiotaan opetus- ja kasvatusalan monitahoisten ilmiöiden arvioinnissa (vrt. edellä mekanismeja jäljittävä tai vuorovaikutteinen järkeilytyyli).

Arviointityön laajuus ja moni-ilmeisyys

Arviointitoiminta on laajentunut huomattavasti niin, että jotkut puhuvat jo arviointiähkystä. Vaikka arviointi on periaatteessa tähdellistä toimintaa, ja kaikkien tahojen mahdollisuus tehdä arviointia on merkki demokratiasta ja yhteiskunnan avoimuudesta, arviointikin voi tehdä laajenemisellaan karhunpalveluksen itselleen. Esimerkiksi Vuorenmaa (2001) osoittaa maailmalla jo tapahtuneen arvioinnin inflatoitumista: liiallinen arviointi väsyttää ja vie arvon koko toiminnalta. Suomessakin nykyään koulut ja opettajat joutuvat toistuvasti informantin rooliin, kun opetusministeriö, Opetushallitus, Koulutuksen arviointineuvosto, Suomen kuntaliitto, lääninhallitukset, koulupalvelukeskukset, tutkijat ja kansalaisjärjestöt tekevät kartoituksia, selvityksiä, seurantoja, mielipidetiedusteluja, asennekyselyjä, barometreja, tilastotiedon keruita, arviointeja ja tutkimuksia.

Selvitykset, kartoitukset ja mielipidetiedustelut on saatettu usein toteuttaa kevein teoreettisin ja metodisin perustein, eivätkä ne täytä tieteellisen tutkimuksen systemaattisen tiedonkeruun, aineiston analyysin ja tulkinnan vaatimattomimpiakaan kriteereitä. Kartoituksia (survey), seurantaa (follow up) tai barometreja voidaan tehdä yhtä kevyin perustein, mutta toisinaan tutkimuksellinen tarkkuus ja huolellisuus leimaavat myös niitä. Esimerkiksi oppimaan oppimisen valmiuksien arviointi (Hautamäki ym. 2003 ja 2005) on hyvä esimerkki laadukkaasta seuranta-arvioinnista, mutta se onkin toteutettu tutkijakoulutuksen saaneiden toimesta.

Paradigmamurros kvantitatiivisesta kvalitatiiviseen – ja viime aikoina niitä rinnakkain hyödyntävään (mixed methods) – lähestymistapaan on vaikeuttanut tarkan rajan vetämistä arvioinnin ja tutkimuksen välille. Vaikka arviointi on arvosidonnaista, ei tutkimuskaan ole ”puhdasta tietoa”. Tutkimuksen tavoitteita ja ongelmanasettelua sekä tulosten tulkintaa ohjaa tieteellisen relevanssin lisäksi myös yhteiskunnallinen konteksti. Arviointi voi muistuttaa tavoitteiltaan ja metodologialtaan tutkimusta, vaikka johtopäätösten teko on arvottavaa. Tällöin puhutaan jo yleensä arviointitutkimuksesta (Jakku-Sihvonen & Heinonen 2001). Tutkimusten tuloksista tehdään arvo- ja kontekstisidonnaisia päätelmiä ja suosituksia, aivan kuten arvioinnissakin.

Tutkimus, arviointi ja arviointitutkimus

Vaikka arviointi- ja tutkimustoiminta muistuttavat monin tavoin toisiaan, joitakin eroja niiden kesken voidaan tehdä taulukosta 2 ilmenevällä tavalla.

Kun ”puhdas” (kvantitatiivinen, positivistinen) tutkimus asettaa ongelmia ja testaa teoriasta ja aiemmista tutkimustuloksista johdettuja hypoteeseja, arviointi on ohjelma- tai ongelmakeskeistä. Se pyrkii tukemaan pedagogisia ja koulutuspoliittisia ratkaisuja sekä yhteiskunnallisesti relevanttia kehittämistyötä, kun taas tieteellinen tutkimus etsii totuutta ja yleistäviä teorioita tai kehittää niitä edelleen. Kun tutkimustilanteessa väliintulevia tekijöitä pyritään kontrolloimaan mahdollisimman tarkasti, arvioinnissa tavoitellaan toimintaa mahdollisimman aidoissa oloissa, monia yhtäaikaaisesti vaikuttavia tekijöitä huomioiden.

Johtopäätökset tehdään arvioinnissa tavallisesti kertamittausten perusteella ennalta sovittuihin kriteereihin verraten eikä toistoihin tai ristiinvaldointiin luottaen kuten tutkimuksessa (Raivola 1995; Leimu 1996). Tosin kansainvälisissä arvioinneissa (esim. PISA) mittavälineiden pätevyyden ja luotettavuuden takaa-

Taulukko 2. *Koulutuksen arvioinnin, arviointitutkimuksen ja tutkimuksen eroja ja yhtäläisyyksiä (vrt. Jaku-Sihvonen & Heinonen 2001, 26)*

Vertailtava ominaisuus	Arviointi	Arviointitutkimus	Tutkimus
Tiedon käyttötarkoitus	päätöksenteko ja ohjaus	tiedonhankinta	yleistyksyet
Teorian rooli	toteutusta tukeva arviointikehys	käsitteellinen arviointikehys	aiempaan tutkimukseen perustuva teoriatausta
Ongelmanasettelu	koulutuksen ja opetuksen ongelmat	koulutuksen ja opetuksen ongelmat ja ratkaisumallit	teorioiden testaaminen, todentaminen ja kehittäminen
Hypoteesit	ei olettamuksia ennakkoon	hypoteesien rakentaminen	hypoteesien testaaminen
Analyysit	tulosten kuvaus	yhteyksien etsiminen ja mallintaminen	väliin tulevien tekijöiden kontrollointi mallinnuksessa
Validointi	harvinainen	toistomittaukset	ristiinvaldointi
Tulosten yleistettävyyys	suppea, kohteeseen rajoittuva	perusjoukkoon yleistäminen	yleistäminen
Toiminnan päämäärä	arvottaminen: hyvien ja huonojen puolien, hyötyjen, etujen ja haittojen osoittaminen	vaihtoehtoisten ratkaisujen hyvien ja huonojen puolien, etujen ja haittojen sekä rajoitusten pohtiminen ja kontekstualisointi	totuuden etsiminen: yleistävien johtopäätösten, mallien ja teorioiden kehittäminen

minen edellyttää nykyisin useita esitestauksia ja monivaiheisia validointiprosesseja (Peschar 2004; Klieme 2004).

Teoriakehys ja arvottaminen

Sekä kansallisissa että kansainvälisissä arvioinneissa on tutkimuksen tapaan ennistä enemmän kiinnitetty huomiota arviointien teoreettis-käsitteellisen viitekehysten vahvuuteen, mittauksen ekologiseen monitasoisuuteen, mittavälineiden autenttisuuteen, osioanalyysien vahvuuteen ja selitysmallien rakentamiseen tulosten tulkinnan ja johtopäätösten teon pohjaksi. Toisaalta tutkimukselta, aivan kuten arvioinneiltakin, odotetaan validiutta, joka kattaa sekä mittaukset että johtopäätösten relevanssin (Leimu 2004; Linnakylä 2001).

Arvioinnissa toiminnan päämääränä on *arvottaminen*, kun tutkimus pyrkii karttamaan arvoarvostelmia. Arvioinnin tieteellisyyttä epäillään juuri arvosidonnaisuuden vuoksi, josta on kuitenkin opetus- ja kasvatusalalla vaikea aidosti irtaantua (House & Howe 1999). Arvosidonnaisuuden perusteella voitaisiin nimittäin hylätä muutkin soveltavat tieteet, kuten yhteiskuntatiede, tilastotiede tai lääketiede, joiden tehtävänä on tuottaa hyödyllistä, yhteiskunnallisesti, sosiaalisesti ja kulttuurisesti arvokasta tietoa. Arvioinnin arvostus soveltavana tieteenä on maailmanlaajuisesti vahvistunut, koska arviointi palvelee muutosta ja kehittämistä. Arviointitutkimuksen vahvuutena on taas metodologinen pätevyys, monitieteisyys ja tieteidenvälisyys sekä tieteelliseen toimintaan kaivattu käytännöllinen relevanssi.

Vapausasteiden eroja

Arviointiin verrattuna tieteellisellä tutkimuksella näyttäisi olevan mahdollisuus suurempaan *päätelmälliseen ja menetelmälliseen vapauteen*, mutta toisaalta sen toteuttamisen ehdot ovat vaativammat kuin arvioinnilla. Tutkijan tulee perustella huolellisesti esimerkiksi kohdejoukon valinta, noudattaa analyysimenetelmää mahdollisimman oikeaoppisesti ja osoittaa tarkasti tulosten johtaminen aineistosta. Tutkimuksen tulokset ja johtopäätökset liitetään aiempaan tutkimukseen ja alan teoreettiseen kehittämiseen tarkoin lähdeviittein. Arvioinnin tuloksista ja johtopäätöksistä keskustellaan pikemminkin koulutuspoliittisin, pedagogisin ja varsin praktisin suosituksin.

Tutkija voi valita *tutkimusaiheensa* vapaammin kuin arvioija, joka tavallisesti saa toiselta henkilöltä tai taholta toimeksiannon tehdä jonkin kohteen arviointi. Tässä suhteessa tutkijan autonomia on erityisen tärkeä. Yleensä tutkimukselta edellytetään selvää rajausta ja keskittymistä joihinkin asioihin syvällisesti, eikä tutkittavaan asiaan liittyviä tekijöitä yritetäkään kattaa laajasti. Sen sijaan arviointi kohdentuu monesti kysymyksiin, joissa ratkaistaan käytännön kannal-

ta tärkeitä pulmia ja joissa tarvitaan mahdollisimman monesta tekijästä tietoja, vaikka ne jäisivätkin pintapuolisiksi.

Monitahoarvioinneissa on tärkeää, että *kaikki* arviointiin osallistuvat tahot ja henkilöt *perehtyvät koko arviointiaineistoon*. Kun esimerkiksi kansallisessa perusopetuspedagogiikan arvioinnissa (Koulutuksen arviointineuvosto 2007) koottiin kyselylomakkeilla tietoja ala- ja yläkoulujen pedagogisesta tilasta, ei analysointi- ja päätelmätyönjakoa tehty kysymyksittäin tai teemoittain, vaan jokainen arvioija tulkitse kaikki tulosityhteenvedot ensin itsekseen, ja sitten kokoonnuttiin muodostamaan arvokonsensus. Tutkijat olisivat tässä tehneet rationaalisen työn: kukin olisi syventynyt huolellisesti vain yhteen tai kahteen kysymykseen.

Myös *ajallinen perspektiivi* erottelee tarkasteltavia toimintoja. Arviointi pitää yleensä toteuttaa suhteellisen lyhyessä ajassa, ja se kohdentuu usein jollakin tavalla ajankohtaistuneeseen teemaan. Kerätyt tiedot eivät saa vanheta parannustoimia ajatellen, eikä etenkin poliittisilla päättäjillä ole aikaa odottaa tuloksia (Paechter 2003). Ajankäytön vuoksi arvioinnin perusteellisuudesta on useimmiten tingittävä sen kaikissa vaiheissa. Tutkimus ja arviointitutkimus taas etenevät selvästi hitaammassa tahdissa. Teoriakehyksen rakentaminen, aineiston huolellinen koonti ja analysointi sekä päätelmien tekeminen vaativat jopa useita vuosia, jotta tulokset ovat varmasti uskottavia, oli kyseessä laadullinen tai määrällinen lähestymistapa. Sovelluksiin yltäminen ei tutkimuksissa ole niinkään kiinni tutkimustehtävän pragmaattisuudesta tai ajankohtaisuudesta kuin sen teoreettisesta kantavuudesta.

Tavoitteena rinnakkaisuus ja lomittaisuus

On helppo yhtyä Nikkasen ja Lyytisen (1996, 99–103) esitykseen, että sekä arvioinnin että tutkimuksen on oltava prosesseiltaan ja tuloksiltaan *perusteltua*, kommunikoitavaa, *tietyllä tavalla yleistä*, totta tai ainakin todennäköistä, korjautuvaa, objektiivisuuteen pyrkivää (ei tahallisesti subjektiivista), *kriittistä*, *autonomista* ja edistävää.

Sekä arviointi että tutkimus lähtevät liikkeelle mahdollisimman huolellisesti mietityn kysymyksen esittämisestä (mitä tutkitaan tai arvioidaan), pohtivat ongelmanratkaisun tapoja (ketä tutkitaan tai arvioidaan ja miten materiaali kerätään sekä analysoidaan) ja useimmiten tuloksista kirjoitetaan raportti. Kumpikin tietysti toivoo, että omasta toiminnasta ja sen tuloksesta olisi jonkinlaista hyötyä tai iloa edes jollekin taholle, heti tai tulevaisuudessa.

Kun arvioinnin metodologiset vaatimukset ovat lisääntyneet sekä kansallisesti että kansainvälisesti, myös arvioinnin, tutkimuksen ja arviointitutkimuksen lomittaisuus on vahvistunut. On tärkeää, ettei tutkimuksen ja arvioinnin eroja korostettaisi tarpeettomasti. Oleellista on tehdä niin arviointi- kuin tutkimustehtävät mahdollisimman luotettavasti, pätevästi ja systemaattisesti sekä raportoida

tulokset koulutusta, opetusta ja oppimista edistäen. Samalla pitää pohtia kriittisesti niitä rajoituksia, joita työn kuluessa on jouduttu eri syistä tekemään ja jotka vaikuttavat johtopäätösten arvoon.

Tiiviiksi päätännöksi arviointia, tutkimusta ja arviointitutkimusta koskevaan teemaan voidaan sanoa Raudasojan (2003, 55) tavoin, että kaikki arviointitutkimus on arviointia, mutta kaikki arviointi ei ole tutkimusta. Päätelmää voisi jatkaa toteamalla, että opetus- ja kasvatusalalla on sellaista perustutkimusta, jossa ei ole arvon antamiseen tähtääviä elementtejä, mutta useimmiten alan soveltava tutkimus on luonteeltaan normatiivista, koska sen kohteet ovat pääsääntöisesti arvosidonnaisia.

Miten ja milloin tuotetaan kehitystä palvelevaa erilaista tietoa?

Opetus- ja kasvatustieteen kehittämiseksi tarvitaan sen eri tasoilla erilaista tietoa. Joskus riittää pikaisesti toteutettu selvitys, jonka kohde on valittu puhtaasti harkinnanvaraisena näytteenä. Toisinaan päättäjiä voi palvella barometri, johon on valittu muutama rajallinen indikaattori, joiden tuottamaa tietoa seurataan ajan funktiona. Yksittäinen erikseen tehty arviointi saattaa muodostaa vain pienen elementin päätöksenteossa, johon liittyy lukuisia muita vaikuttavia tekijöitä.

Aika-ajoin voidaan olla kiperässä ongelmatilanteessa, jonka yhteydessä tarvitaan huolellisesti, kattavasti ja monimenetelmällisesti koottua tietokokonaisuutta. Tämän tietokokonaisuuden varassa pitää voida tehdä isoja päätöksiä, suunnata koko organisaation tulevaisuutta ja ratkaista mittavien voimavarojen kohdentaminen. Silloin tarvitaan järeämpiä keinoja eli korkeatasoista tutkimusmenetelmällistä osaamista ja kokemusta, joka maksaa ja vaatii aikaa.

Tietotarpeiden tyydyttämiseen saattaa riittää poikkileikkaustarkastelu (esim. laaja tai suppea kyselylomakeaineisto), mutta esimerkiksi opetussuunnitelmat tai opetusmenetelmäuudistuksissa kaivataan kipeästi arvioivaa tai tutkimuksellista seuranta prosessien etenemisestä. Oikeaoppinen tutkimusseuranta tietysti vaatii, että seurannan kriteerit on tarkasti valittu siten, että ne kykenevät kuvaamaan kohteen validisti vaikkapa viiden vuoden aikana (esim. koulukokeilun alussa, keskellä ja lopussa). Myös kohderyhmien pitää olla samat, jotta voidaan havainnoida muutosta juuri tarkastellussa kohteessa.

Kehittävään ja osallistavaan arviointiin liittyy nykyisin paljon positiivisia odotuksia. Jo itsearviointin vaihe, jolle vaikkapa arviointiryhmän vierailu arviointikohteessa perustuu, saattaa olla arvioitaville merkityksellinen alku kohti muutoksia. Omien vallitsevien käytänteiden inventointi mahdollistaa joidenkin asioiden toteamisen hyviksi. Ne voidaan säilyttää ennallaan, mikä vähentää vastarintaa. Tällöin jää voimia mieltä parannuksia havaittuihin ongelmakohtiin.

Joskus objektiivisuutta tavoittelevan ulkoisen arvioinnin pitää paljastaa todella isoja ongelmia, ennen kuin se otetaan koetun uhkaavuutensa vuoksi vakavasti. Ongelmien korjaaminen taas edellyttäisi, että niistä tiedettäisiin vaiheessa, jolloin ne eivät ole vielä isoja. ”Ismiä” kehittävistä arviointistakaan ei tietysti pidä tehdä.

Arvioijakentän laajeneminen ja kirjavoituminen herättävät tarpeen koordinoida arviointeja. Tietoa kyllä kerätään ja raportoidaan, mutta siitä puuttuu kansallista suunnitelmallisuutta (vrt. Opetusministeriö 2007). Suunnitelmallinen koordinointi ja toimijoiden verkostoituminen on tehokas tapa lisätä tietoisuutta uusista tiedonkeruista ja estää niiden päällekkäisyyttä tai sellaista lomittaisuutta, joka jättää tärkeitä tarkastelukohteita toistuvasti vaille huomiota.

Koordinoinnissa tulisi myös parantaa niin arviointia kuin tutkimusta tekevien mahdollisuuksia käyttää tehokkaasti eri yhteyksissä aiemmin koottuja arviointiaineistoja. Esimerkiksi kansallisten ja kansainvälisten arviointien tuloksia voidaan ymmärtää entistä paremmin, jos arviointiaineistoja voidaan hyödyntää jatkotutkimuksissa ja meta-analyyseissa. Suuret arviointiaineistot tarjoavat mahdollisuuksia metodisten sovellusten kehittämiseen ja selitysmallien rakentamiseen. Nämä voivat puolestaan luoda pohjaa seuraavien arviointien teoreettisestikin vahvan kehityksen rakentamiselle. Arviointiaineistot ovat aarteita myös uusien arviointiasiantuntijoiden koulutuksessa.

Lähteet

- Arffman, I. 2004. Kansainvälisten lukukokeiden kääntäminen ja kulttuurinen soveltaminen. Teoksessa P. Linnakylä, S. Sulkunen & I. Arffman (toim.) *Tulevaisuuden lukijat. Suomalaisnuorten lukijaprofiileja*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 49–71.
- Arffman, I. 2007. The problem of equivalence in translating texts in international reading literacy studies. Jyväskylän yliopisto. Koulutuksen tutkimuslaitos. Tutkimuksia 21.
- Atjonen, P. 2005. Arviointi opetuksen ja oppimisen tueksi. Teoksessa H. K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnista*. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä, 143–152.
- Baker, E. 2004. Reforming education through policy metaphors. Teoksessa T. Fitzner (toim.) *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie, 150–163.
- Bempechat, J., Jimenez, N. V. & Boulay, B. 2002. Cultural-cognitive issues in academic achievement. Teoksessa A. Porter & A. Gamoran (toim.) *Methodological advances in cross-national surveys of educational achievement*. National Research Council Washington, DC: National Academy Press, 117–149.
- Bonnet, G. 2002. Reflections in a critical eye: On the pitfalls of international assessments. *Assessment in Education* 9, 387–400.
- Chelimsky, E. 1997. The coming transformations in evaluation. Teoksessa E. Chelimsky & W. R. Shadish (toim.) *Evaluation for the 21st century. A handbook*. Thousand Oaks: Sage, 1–26.

- Cook, T. 1997. Lessons learned in the evaluation over the past 25 years. Teoksessa E. Chelimsky & W. Shadish (toim.) *Evaluation for the 21st century*. London Sage, 30–52.
- Hambleton, R. K., Merenda, P. F. & Spielberger, C. D. (toim.) 2005. *Adapting educational and psychological tests for cross-cultural assessments*. Mahwah, NJ.: Lawrence Erlbaum.
- Harvey, L. 2002. The end of quality. *Quality in Higher Education* 8 (1), 5–22.
- Harvey, L. & Green, D. 1993. Defining quality. *Assessment & Evaluation in Higher Education* 18 (1), 9–34.
- Hautamäki, J., Hautamäki, A., Lindblom, B. & Scheinin, P. 2003. Oppimaan oppiminen yläasteella. Tilanne vuonna 2001 ja muutokset vuodesta 1997. Opetushallitus. Oppimistulosten arviointi 6. Helsinki.
- Hautamäki, J., Kupiainen, S., Arinen, P., Hautamäki, A., Niemivirta, M., Rantanen, P., Ruuth, M. & Scheinin, P. 2005. Oppimaan oppiminen ala-asteella. Tilanne vuonna 2003 ja muutokset vuodesta 1996. Opetushallitus. Oppimistulosten arviointi 1. Helsinki.
- House, E. R. 2005. Deliberative democratic evaluation. Teoksessa S. Mathison (toim.) *Encyclopedia of evaluation*. Thousand Oaks, Ca.: Sage, 104–108.
- House, E. R. & Howe, K. R. 1999. *Values in evaluation and social research*. Thousand Oaks: Sage.
- Jakku-Sihvonen, R. & Heinonen, S. 2001. Johdatus koulutuksen uudistuvaan arviointikulttuuriin. Arviointi 2. Opetushallitus. Helsinki.
- Kauppinen, J., Kaihari-Salminen, K. & Hyvärinen, R. 2006. Perusopetuksen oppilaskunta-toiminnan kartoitus 2005. Opetushallitus. Monisteita 9/2006. Helsinki: Edita Prima.
- Klieme, E. 2004. Assessment of cross-curricular problem-solving competencies. Teoksessa J. H. Moskowitz & M. Stephens (toim.) *Comparing learning outcomes. International assessments and education policy*. New York: Routledge Falmer, 81–107.
- Koulutuksen arviointineuvosto. 2007. Perusopetuspedagogiikan kansallinen arviointi. Julkaisematon arviointisuunnitelma. Jyväskylän yliopisto. Arviointisihteeristön arkisto.
- Kushner, S. 2005. Programme evaluation. Teoksessa S. Mathison (toim.) *Encyclopedia of evaluation*. Thousand Oaks, Ca.: Sage, 334–338.
- Lam, T. C. M. 2004. Issues and strategies in standards-based school reform: The Canadian experience. Teoksessa T. Fitzner (toim.) *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie, 103–149.
- Leimu, K. 1996. Arviointi koulutustutkimuksessa. Teoksessa R. Laukkanen & K. Stenvall (toim.) *Arviointi koulutus- ja tiedepolitiikassa*. Tampereen yliopisto. Hallintotiede A. Tutkimuksia 9, 59–78.
- Leimu, K. 2004. Kansainvälisyys koulutuksen arviointitutkimuksessa. Teoksessa K. Leimu (toim.) *Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 7–26.
- LeTendre, G. K. 2002. Advancements in conceptualizing and analyzing cultural effects in cross-national studies of educational achievement. Teoksessa A. Porter & A. Gamoran (toim.) *Methodological advances in cross-national surveys of educational achievement*. National Research Council Washington, DC: National Academy Press, 198–228.
- Lie, S., Linnakylä, P. & Roe, A. (toim.) 2003. Northern lights on PISA. Unity and diversity in the Nordic countries in PISA 2000. University of Oslo: Department of Teacher Education and School Development.
- Linnakylä, P. 1974. Oppilaiden ja opettajien asenteet yhteisiä kokeita kohtaan sekä kokemukset ja odotukset yhteisten kokeiden funktioista. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitoksen julkaisuja 222.
- Linnakylä, P. 1998. Koulutuksen kansallisen ja kansainvälisen arvioinnin haasteista. *Kasvatus* 29 (2), 175–189.

- Linnakylä, P. 2004. Educational 'standards' in Finland: Opportunities and threats. Teoksessa T. Fitzner (toim.) *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie, 43–53.
- Linnakylä, P., Malin, A. & Taube, K. 2006. What lies behind low reading literacy performance? A comparative analysis of the Finnish and Swedish students. Teoksessa J. Mejding & A. Roe (toim.) *Northern lights on PISA 2003 – a reflection from the Nordic countries*. Copenhagen: Nordic Council of Ministers, 143–158.
- Linnakylä, P. & Välijärvi, J. 2005. Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulun kehittämiseksi. Jyväskylä: PS-kustannus.
- Löfström, E., Metsämuuronen, J., Niemi, E. K., Salmio, K. & Stenvall, K. 2005. Koulutuksen paikallinen arviointi vuonna 2004. Opetushallitus. Arviointi 2. Helsinki.
- Mejding, J. & Roe, A. (toim.) 2006. *Northern lights on PISA 2003 – a reflection from the Nordic countries*. Copenhagen: Nordic Council of Ministers.
- Mäkelä, M.-L. 2006. Millä koulutuksella mihin tehtäviin laboratorioalalla? Selvitys laboratorioissa työskentelevien kemistien, laboratorioanalytikoiden ja laboranttien koulutuksista, sijoittumisesta ja työtehtävistä. Opetushallitus. Moniste 13. Helsinki.
- Nikkanen, P. & Lyytinen, H.K. 1996. Oppiva koulu ja itsearviointi. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- OECD. 2001. *Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.
- OECD. 2004. *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- Opetusministeriö. 2007. Koulutuksen arviointijärjestelmän kehittämistyöryhmän muistio. Työryhmämuistioita ja selvityksiä 27. Helsinki.
- Paechter, C. 2003. On goodness and utility in educational research. Teoksessa P. Sikes, J. Nixon & W. Carr (toim.) *The moral foundations of educational research. Knowledge, inquiry and values*. Maidenhead: Open University, 105–117.
- Peschlar, J. 2004. Cross-curricular competencies: Developments in a new area of education outcome indicators. Teoksessa J. H. Moskowitz & M. Stephens (toim.) *Comparing learning outcomes. International assessments and education policy*. New York: Routledge Falmer, 59–20.
- Porter, A. & Gamoran, A. (toim.) 2002. *Methodological advances in cross-national surveys of educational achievement*. National Research Council Washington, DC: National Academy Press.
- Raivola, R. 1995. Mitä evaluaatio on ja mihin sitä tarvitaan? Teoksessa A. Kajanto (toim.) *Aikuiskoulutuksen arviointi. Panoraamoja ja lähikuvia. Aikuiskasvatuksen 36. vuosikirja*. Kansanvalistusseura & Aikuiskasvatuksen tutkimusseura. Helsinki: Kirjasto-palvelu, 21–60.
- Rajavaara, M. 1999. Arviointitutkimuksen hyödynnettävyys. Teoksessa R. Eräsaari, T. Lindqvist, M. Mäntysaari & M. Rajavaara (toim.) *Arviointi ja asiantuntijuus*. Tampere: Gaudeamus, 31–54.
- Raudasoja, E.-M. 2005. Koulutuksen paikallisen arvioinnin kehittämisprosessi Oulun kaupungissa 1997–2001. Tampereen yliopisto. Acta Universitatis Tamperensis 1081.
- Roe, A. & Hvistendahl, R. 2006. Nordic minority students' literacy achievement and home background. Teoksessa J. Mejding & A. Roe (toim.) *Northern lights on PISA 2003 – a reflection from the Nordic countries*. Copenhagen: Nordic Council of Ministers, 113–128.
- Räisänen, A. 2005. Kehittävään arviointiin. Teoksessa H. K. Lyytinen & A. Räisänen (toim.) *Kehittämisuuntaa arvioinnista. Koulutuksen arviointineuvoston julkaisuja 6*. Jyväskylä, 109–127.
- Scheerens, J. 2004. The formative implication of standards. Teoksessa T. Fitzner (toim.) *Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform*. Bad Boll: Evangelische Akademie, 202–219.

- Schleicher, A. 2004. Deutschland im internationalen Bildungswettbewerb. Teoksessa T. Fitzner (toim.) Bildungsstandards. Internationale Erfahrungen – Schulentwicklung – Bildungsreform. Bad Boll: Evangelische Akademie, 277–286.
- Sulkunen, S. 2004. Millaiset tekstit ovat nuorten mieleen? Lukukoetekstien autenttisuus suomalaisnuorten kulttuurissa. Teoksessa P. Linnakylä, S. Sulkunen & I. Arffman (toim.) Tulevaisuuden lukijat. Suomalaisnuorten lukijaprofiileja. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 5–32.
- Svedlin, R., Hotari, A.-M. & Lyra-Katz, A. 2002. Selvitys lasten ja nuorten aamu- ja iltapäivätoiminnasta. Opetushallitus. Moniste 8. Helsinki.
- Vuorenmaa, M. 2001. Ikkunoita arvioinnin tuolle puolen. Jyväskylä Studies in Education, Psychology and Social Research 176.
- Weeden, P., Winter, J. & Broadfoot, P. 2002. Assessment. What's in it for schools? London: Routledge Falmer.

II

Arviointitoiminnan organisoinnin tapoja ja menetelmiä



Verkostomaisuus koulutuksen arvioinnin organisointitapana

Verkostona vai virastona?

Koulutuksen arviointijärjestelmiä voidaan tarkastella sen mukaan, miten tehokkaasti ja laaja-alaisesti ne pystyvät hyödyntämään kansallisia ja kansainvälisiä asiantuntijoita ja asiantuntijaorganisaatioita arviointien toteuttamisessa. Voidaan puhua verkostomaisuuden ulottuvuudesta, jonka ääripäinä ovat yhtäällä tiukasti yhteen (vieläpä hallinnon sisällä olevaan) arviointivirastoon keskitetty arviointitoiminta ja toisaalla mahdollisimman laajasti asiantuntijamarkkinoilla tarjolla olevaa osaamista arvioinneissa käyttävä, alati muuntuva ja täydentyvä verkosto.

Organisatorisesti ja resurssien kannalta konkreettinen kysymys onkin, kuinka suuri arviointiorganisaatio tai -yksikkö viisi miljoonaisen väestön tarpeisiin on sopiva. Tarvitaanko erillinen instituutti tai arviointivirasto kiinteine henkilökuntineen ja virkoineen, vai onko eri asiantuntijaorganisaatioiden ja tahojen osaamista verkostomaisesti hyödyntävä järjestelmä paremmin perusteltu vaihtoehto? Lopulta asia päätetään päätöksenteon korkeimmalla tasolla, eduskunnassa. Hyvin perusteltukaan ei ole aina poliittisesti mahdollinen. Koulutuksen arvioinnin järjestämiseksi tulevaisuudessa on vuonna 2007 esitetty useita vaihtoehtoisia organisointitapoja. Opetusministeriön asettama ja ylijohtaja Sakari

Karjalaisen johtama koulutuksen arvioinnin kehittämistyöryhmä esitti Korkeakoulujen arviointineuvoston ja Koulutuksen arviointineuvoston säilyttämistä erillisinä eliminä ja näiden sihteeristöjen sijoittamista yksiin tiloihin opetusministeriön yhteyteen ilman hallinnollista yhdistämistä (OPM 2007). Opetushallituksen asemaa, roolia ja tehtäviä arvioinut selvitysmies Timo Lankinen (2007) esitti raportissaan kaiken koulutuksen ulkoisen arviointitoiminnan keskittämistä uudelle perustettavalle Koulutuksen arviointikeskukselle, johon kuulusivat niin korkeakoulujen kuin muunkin koulutuksen arviointi. Oppimistulosten arviointi jäisi Opetushallituksen tehtäväksi.

Verkostomaisuus on ennen kaikkea toimintatapaa luonnehtiva ominaisuus. Verkosto on resurssi, jonka avulla arvioinnit organisoidaan ja toteutetaan. Verkostosuhde voi vaihdella tarkoituksen mukaan löyhästä kiinteään, tilapäisestä pysyvään, epävirallisesta viralliseen, paikallisesta kansainväliseen, osapuolien autonomiasta hierarkkisempaan määräysvaltasuhteeseen.

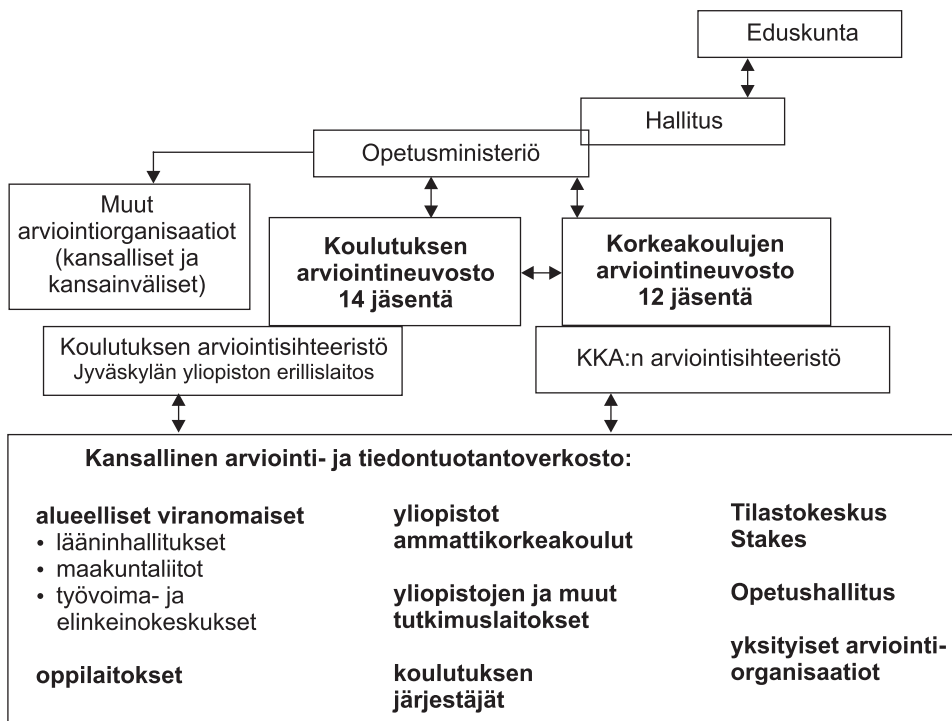
Vuonna 2003 Suomessa koulutuksen arvioinnissa käyttöön otettu verkostomalli mahdollistaa erikoistuneiden asiantuntijoiden käyttämisen kunkin arviointikohteen tarpeiden mukaisesti. Uuden organisointitavan ero aiempaan on hyvin selkeä, kun arvioinnin organisointitapojen ulottuvuuden toisessa ääripäässä on erillinen ”arviointivirasto” kymmenine vakinaisine virkamiehineen, jotka suorittavat arviointeja ja laativat niistä raportteja. Aiemmin Opetushallituksessa työskenteli yli 30 henkilöä arviointiasioissa ja arviointiraportit laadittiin virkatyönä, samalla kun organisaatio oli keskeinen toimija koulutuksen kehittämistyössä. Koulutuksen arviointineuvostossa arviointitoimintaa on organisoinut pääsihteerin ja viiden pääsuunnittelijan muodostama arviointisihteeristö. Sen enempää arviointineuvosto kuin sen sihteeristökin eivät tee arviointeja. Arvioinnit laaditaan ulkopuolisten riippumattomien asiantuntijoiden toimesta. Arviointineuvosto organisoii arviointitoiminnan ja luo mahdollisimman hyvät edellytykset arviointien tekemiseksi laadukkaasti, tehokkaasti ja nopeasti (ks. Lyytinen & Hämäläinen 2004).

Arvioinnin organisoinnissa painotetaan ensisijaisesti arviointiprosessin keveyttä ja joustavuutta. Arviointiprosessien tehokkuutta ja taloudellisuutta edistetään kiinnittämällä huomiota arviointihankkeiden sisällölliseen ja ajalliseen rytmitykseen sekä niiden määrään, tiedonhankinta- ja analysointimenetelmiin, tulosten raportointiin, verkoston käyttöön, vaihtoehtoihin organisointitapoihin ja olemassa olevan tiedon hyödyntämiseen.

Aktiivinen verkosto on periaatteessa tehokkain malli parhaiden asiantuntijoiden käyttöön. Alun perin markkinatalous- ja yritysmaailmasta tutun verkostoitumisen keskeiset vaikuttimet ovat tiedon tehokas käyttö, nopeus ja joustavuus (esim. Ollus & al. 1998). Saman mallin mukaan koulutuksen arvioinnissa pyritään verkostomaisuuteen, jotta saataisiin käyttöön:

1. kulloisenkin arviointikohteen ja tehtävälueen ajantasaisin ja syvällisin erityisasiantuntemus
2. eri alojen ja kohteiden erityisasiantuntijuuksien yhteen kokoamisen mukanaan tuomat synergiaedut
3. nopeasti laadukasta tietoa ja tuloksia
4. monipuolisesti kentän, sidosryhmien ja eri intressitahojen näkemykset tulkintoineen.

Verkostomainen arviointijärjestelmä rakentuu jo olemassa olevan infrastruktuurin varaan. Arviointihankkeiden toteuttamisen ohella verkostoyhteistyössä kehitetään arviointijärjestelmää. Keskeisiä yhteistyöalueita ovat tietojärjestelmien, indikaattoreiden ja arviointimenetelmien ja -tutkimuksen kehittäminen, koulutuksen järjestäjien ja oppilaitosten henkilöstön kouluttaminen ja koulutuksen järjestäjille suunnattujen arviointipalveluiden kehittäminen. Arviointijärjestelmän kehittämisessä muun muassa Opetushallitus on tärkeä yhteistyötaho Tilastokeskuksen ja Stakesin asiantuntijoiden ohella. Suomen kansallinen koulutuksen arviointiverkosto on koottu kuvioksi 1.



Kuvio 1. Kansallinen koulutuksen arviointijärjestelmä Suomessa

Hyvin toimivan verkoston ominaisuuksia

Koulutuksen arvioinnin kenttään kytkeytyvät toimijat ovat verkoston solmu-kohtia. Toimijat voivat olla henkilöitä ja muodollisia tai epämuodollisia ryhmiä niin yksityisistä kuin julkisistakin organisaatioista. Verkostojen luominen, ylläpitäminen ja säilyminen perustuvat toimijoiden erikoistumiseen ja monien erikoisosaamisalueiden välttämättömyyteen kokonaisuuden kannalta. Verkosto on luonnehdittavissa samanaikaisesti sekä heterogeeniseksi että homogeeniseksi, mihin sen vahvuus olennaisesti perustuu. Tietty määrä yhdenmukaisuutta (ainakin kielessä, ajattelussa ja tavoitteissa) tarvitaan, jotta erilaisuus saadaan hyödynnettyä sovitun päämäärän saavuttamiseksi.

Toimiakseen verkosto tarvitsee koordinoivan osapuolen, joka huolehtii siitä, että toimijat pyrkivät yhteiseen päämäärään. Koordinoija huolehtii siitä, että verkostoon on helppoa ja mielekästä osallistua. Ajantasaista tietoa tuottavan arviointijärjestelmän yksi avainkysymys on, miten koko verkosto (solmukohtineen) pidetään koko ajan sillä tavoin aktiivisena, että sieltä on rekrytoitavissa asiantuntijaresursseja mahdollisimman nopealla aikataululla.

Verkostosuhde on tulkittavissa arviointiyhteistyösuhteeksi, joka tuottaa synergiaetuja yksilöille, ryhmille ja organisaatioille. Sille on ominaista verkostopartnereiden vastavuoroisuus, keskinäinen riippuvuus ja luottamus sekä löyhä keskinäinen sidonnaisuus. Verkostoitumisen päämääränä on menestystä tuottavan yhteistyökumppanuuden solmiminen. Jotta verkostosuhde toimisi hyvin, verkostotoimijoilla täytyy olla riittävän samankaltainen näkemys yhteistoiminnan tavoitteista ja sen kokonaisuudesta. Yhteistyökumppaneiden on tunnettava toistensa perustehtävät, tavoitteet ja toimintaperiaatteet sekä sovittava yhteistoiminnan periaatteista, menettelytavoista ja tosiasiallisista mahdollisuuksista osallistua yhteistyöhön. Yhteistyön toimivuuden kannalta on eduksi, jos osapuolet paneutuvat toistensa toiminnan prioriteetteihin ja tärkeänä pitämiin asioihin sekä tapoihin kommunikoida. Kielellisten merkitysten oikea tulkinta on välttämätöntä, kun edetään kommunikatiiviseen yhteistyöhön. Ensisijaisen tärkeää on, että osapuolet sitoutuvat ottamaan vastuuta paitsi omasta osuudestaan myös yhteistoiminnan kokonaisuudesta. Verkostoyhteistyöhön kiinnittymistä tukevat myös yhteistä hyötyä tai yleistä etua palvelevat yhteistyökohteet ja verkostopartnereille lisäarvoa tuottavat konkreettiset projektit. Avainlähtökohdat ovat siinä määrin vaativia ja haasteellisia, etteivät ne käytännössä toteudu hetkessä, vaan vaativat aikaa, määrätietoista ja pitkäjänteistä työtä, yhteisiä projektikokemuksia sekä aitoa dialogia verkostopartnereiden kesken. Alkuvaiheessa nousevat tärkeiksi konkreettiset ja nopeasti palkitsevat hyödyt. (Lyytinen 2006.)

Arviointiverkoston pitäminen toimintakykyisenä riippuu monista vaikeasti hallittavista tekijöistä. Toimintakykyisen verkoston keskeisiä ominaisuuksia voidaan eritellä esimerkiksi seuraavasti (Nikkanen & Lyytinen 1996, 59–60):

1. **Täydentävyys:** verkoston jäsenet voivat toistensa kautta täydentää aineellisia ja henkisiä resurssejaan. Yksi toimija on vahva yhdellä, toinen toisella osaamisalueella.
2. **Yhteensopivuus:** Verkoston aluetoimijoiden, eri organisaatioyksiköiden jne. tulevaisuudenkuvat, päämäärät, toiminta-ajatukset, toimintakulttuuri ja toimintastrategiat edistävät yhdessä asetettujen tavoitteiden saavuttamista.
3. **Yhteistyöhalukkuus:** Hyvät olosuhteet ja monipuoliset valmiudet eivät riitä, tuloksellisuus edellyttää motivoitumista aitoon yhteistyöhön.
4. **Sitoutuminen:** Sitoutuminen ymmärretään ihmisen vapaaehtoiseksi kiinnostumiseksi ja innostumiseksi panemaan itsensä likoon jonkin asian puolesta ilman ulkopuolista houkutinta tai pakotetta. Taustalla voi olla henkilökohtainen intressi, inhimillinen uteliaisuus, vastuun- tai velvollisuudentunne, korkea suoritusmotivaatio, sosiaalinen päteminen, henkilökohtaisesti koetut tärkeät arvot jne. Verkostotoimintaan sitoutuminen tapahtuu vaikuttamismahdollisuuksien sekä taustojen ja toiminnan vaihtoehtojen tuntemuksen kautta syntyvän ymmärryksen avulla.
5. **Luottamus:** Verkostoituminen edellyttää organisaatioilta sitä, että ne ovat toimintakulttuuriltaan avoimia eivätkä suljettuja. Luottamus ja avoimuus syntyvät päämäärien ja strategioiden yhteisen hyväksynnän kautta. Tiedonkulun avoimuus ja runsaus sekä informaatiokanavien avoimuus kaikille arvoon ja asemaan katsomatta vahvistavat luottamusta. Luottamus on verkostotoimintaa tehostavaa sosiaalista pääomaa (esim. Ilmonen & Jokinen 2002; Ruuskanen 2003; Blomqvist 2004).

Parhaimmillaan verkoston johtamisen kautta onnistutaan aktivoimaan verkon valmiutta rakentaa itseään ja täydentää eri osiensa kapasiteettia asetettujen tavoitteiden saavuttamiseksi. Kun useinkaan ei ole mahdollista luoda kiinteää yhteisöllisyyttä, tarvitaan toiminnan tuloksellisuuden takaajaksi yhteistä tavoitteenasettelua ja yhteisiä strategisia linjauksia ja vahvaa sitoutumista niihin. Koska verkosto on sosiaalisen teknologian tuote, jonka ylläpitäminen, muokkaaminen ja muuttaminen on jatkuva prosessi, johtaminen on keskeisesti nimenomaan yhteistyön edistämistä. Yhteistyö on omien kykyjen ja voimavarojen täydennystä. Samalla kun toimijat saavat lisää voimavaroja, ne saavat yhteistyösuhteen keskinäisessä työnjaossa sellaisen roolin, joka vapauttaa suuntaamaan omat voimavarat parhaiten osattuihin asioihin, ydinosaamiseen. Yhteistyösuhteen onnistumiseen vaikuttavat keskeisesti strateginen yhteensopivuus, organisaatiokulttuurin piirteet ja henkilösuhteet sekä operatiivinen yhteensopivuus, mihin verkoston rakentamisessa ja ylläpitämisessä tulee erityisesti kiinnittää huomiota. (Nikkonen & Lyytinen 1996.)

Yksilöllinen huippuosaaminen ei ole kollektiivisen onnistumisen tae. Taitava verkostomainen toiminta on sitä, että yhteistoiminnan tulos on enemmän kuin yksilöiden toimintojen ja osaamisen summa. Monipuolisia arviointitehtäviä sisältävä arviointihanke edellyttää usein sitä, että monipuolista asiantuntemusta yhdistetään arviointiprosessin eri vaiheissa tiedonhankinnasta tulkintaan ja arvottamiseen. Asiantuntijuuden jakamiseen perustuva lähestymistapa tuottaa jaetun kokemuksen ansiosta aina lisäarvoa. Siinä onnistuminen edellyttää myös toimivaa ryhmädynamiikkaa. (Lyytinen 2006.)

Luottamuskulttuuri verkostomaisuuden edellytyksenä

Arviointi on koulutuksen ja opetuksen kehittämisen väline, jonka päämääränä on oppimisen edistäminen. Arvioinnin vaikuttavuus edellyttää arvioitsijan ja arviointien kohteiden välistä luottamusta, uskottavan arviointikulttuurin luomista. Luottamusta ja uskottavuutta on vaikea rakentaa ilman riippumattomuutta ja puolueettomuutta. Verkostomainen toimintatapa on keskeinen rakenteellinen edellytys sille, että koulutuksen arvioija on riippumaton opetushallinnosta (koulutuksen päätöksenteosta) ja koulutuksen kehittämiseen osallistuvista tahoista. Hallinnon ulkopuolisena verkosto on myös riippumaton hallinnon sosiaalisesta järjestelmästä.

Luottamus mainitaankin usein hyvin toimivan verkostoyhteistyön yhdeksi tärkeimmäksi edellytykseksi. Toimijoiden keskinäisen luottamuksen puuttuminen halvaannuttaa koko verkoston toiminnan. Luottamuksen kadotessa verkosto jää pelkäksi kuvitelluksi rakenteeksi ilman tuloksellista toimintaa synnyttävää yhteistyötä. Esimerkiksi Blomqvist (2006) näkee luottamuksen toimivan verkostoissa eräänlaisena voiteluöljynä: vähentäessään kitkaa eri tahoilta tulevien toimijoiden välillä se parantaa toiminnan sujuvuutta ja lisää yhteistyöstä kaikille koituvia hyötyjä. Luottamuksen puutteen aiheuttamat kustannukset ja kitka rapauttavat tavoiteltavia synergiahyötyjä niin organisaatioiden sisällä kuin verkostoissa.

Epävarmassa ja kompleksisessa maailmassa menestyvimmit pystyvät hyödyntämään muita nopeammin ja taitavammin tietoa ja osaamista organisaation sisäisissä ja ulkoisissa verkostoissa. Menestyäkseen ja voidakseen hyvin verkostotalouden organisaatiot tarvitsevat luottoa niin työntekijöiltä, esimiehiltä, asiakailta, kumppaneilta kuin muiltakin sidosryhmiltä. Organisaation kilpailukyky onkin tietopääoman lisäksi yhä enemmän kiinni sen yhteistyökyyvystä: menestyminen kilpailussa edellyttää kykyä rakentaa luottamuksellisia, avoimeen kommunikaatioon ja sitoutumiseen perustuvia verkostosuhteita. (Emt.)

Liiketoiminnan ns. transaktiokustannukset muodostuvat etsintä-, neuvottelu-, sopimus- ja valvontakuluista. Kumppanuusyhteistyötä arvioitaessa on punnitta-

va myös mahdollisia transaktiokustannuksia ja -hyötyjä. Luottamus vähentää transaktiokustannuksia. Sen ansiosta on myös mahdollista hyötyä yhteistyöstä.

Tutkimuksissa on todettu luottamuksen syntyvän hitaasti kokemuksen ja oppimisen kautta. Yhteistyö lisää osaltaan luottamusta. Riittävälle tasolle noustuaan luottamus vahvistaa itse itseään. Toisaalta luottamus särkyä hetkessä. Kerran menetettyä luottamusta on erittäin työlästä ja hidasta, ellei mahdotonta, palauttaa entiselleen. (Blomqvist 2006, 20.)

Taulukko 1. *Kumppanuuksiin liittyvät transaktiokustannukset (-) ja -hyödyt (+)*

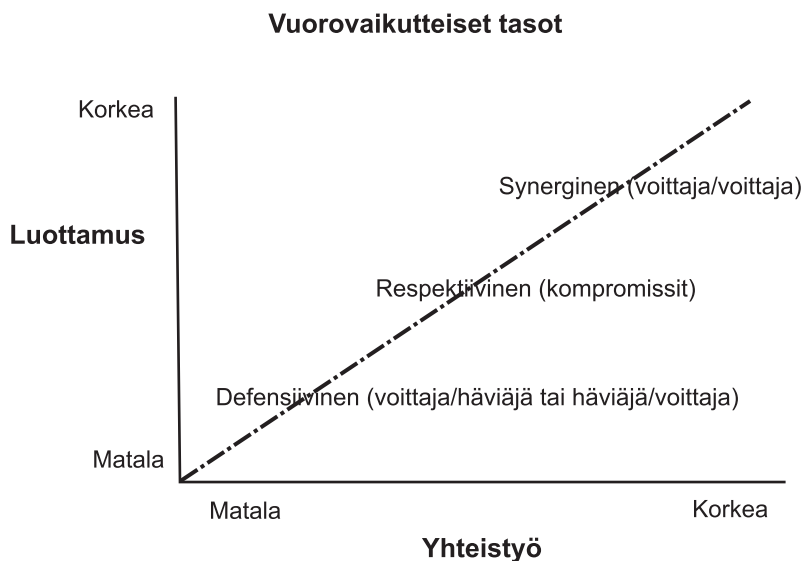
Kustannukset	Hyödyt
– kumppanin etsintäkustannukset	+ fokus ydinosaanamiseen
– neuvottelukustannukset	+ täydentävän osaamisen saaminen
– osaamisen arviointikustannukset	+ laajemman tuotevalikoiman rakentaminen
– sopimuskustannukset	+ strategisen yhteistyön rakentaminen
– suhteen rakentamiskustannukset	+ riskin jakaminen
– valvonta ja kontrollikustannukset	+ joustavuuden lisääntyminen
– suhteen johtamiskustannukset	+ lanseerausajan nopeutuminen
– oikeudenkäyntikustannukset?	+ oppiminen

Lähde: Blomqvist 2006.

Avoimuus on verkostossa kehittyvän luottamuksen keskeinen edellytys. Avoimuus edistää verkostomuotoisen toimintatavan luomaa keskustelemaan ja vuorovaikutteista arviointikulttuuria. Sen lisäksi, että arviointien tulokset tuodaan julkiseen keskusteluun, verkostomaisuus edistää osapuolten aktiivista osallistumista koulutusta ja sen arviointia koskevaan yhteiskunnalliseen keskusteluun yleisemminkin.

Verkoston rakentamis- ja vahvistamisprosessissa toimijoiden vuorovaikutuksella on keskeinen merkitys. Lyytinen pelkistää luottamuksen, yhteistyön ja vuorovaikutuksen suhteet kuvioksi 2.

Vähäiselle vuorovaikutukselle on tyypillistä puolusteleisuus, epäilevyys, suojautuminen, reviirin varjelu, legaalinen kielenkäyttö, pitäytyminen hallinnollisiin rooleihin sekä kilpailevuus. Vuorovaikutus on itsekeskeistä, muodollista ja laskelmoivaa. Vuorovaikutus voi myös leimata yhteistyökumppaneita kielteisesti, mikä synnyttää osapuolten välille voittaja–häviöjä-asetelman. Oma toiminta katsotaan oikeaksi ja oikeaoppiseksi, kumppaneiden toiminta puolestaan vääräksi ja vääräoppiseksi. Keskitasoinen vuorovaikutus on luonteeltaan reilua, aikuismaista, kohteliasta ja toista osapuolta kunnioittavaa. Osapuolet voivat ymmärtää



Lähde: Lyytinen 2006.

Kuvio 2. Vuorovaikutustasot luottamuksen ja yhteistyön rakentamisessa

toisiaan rationaalisesti paneutumatta syvällisesti uusiin yhteistyömahdollisuuksiin. Vuorovaikutus ja yhteistyö ovat kompromissikeskeisiä. Vastakkainasettelua pyritään tietoisesti välttämään. Vuorovaikutus on kuitenkin rehellistä, ja se voi tuottaa väliaikaisia voittaja–voittaja-tilanteita. Syvässä vuorovaikutuksessa myötäletään ja tullaan vahvasti toista vastaan. Vuorovaikutus on luovaa ja uusia mahdollisuuksia etsivää. Verkostosuhde perustuu sitoutumiseen, mikä taas vahvistuu partnerien oppiessa tuntemaan toisensa entistä paremmin. Vuorovaikutus vapauttaa energiaa ja tuottaa yhteistyön osapuolille lisäarvoa. Yhteistyötä leimaa jatkuva uudistumisen asenne. Yhteistoimintaedut ovat selkeitä ja näkyviä. Osapuolet viestivät toisistaan hyvántahtoisesti, oma-aloitteisesti ja luottamusta rakentavasti, jolloin luottamus vahvistaa itseään. Tällaisessa verkostossa ristiriidat ratkotaan rakentavasti. Ristiriitatilanteissa toimijoiden energia onnistutaan suuntaamaan pikemmin erimielisyyksien sisältöön kuin henkilöiden välisiin ristiriitoihin ja eturyhmäpolitiikointiin.

Suuri haaste on, miten voitetaan toimintakulttuurin ne esteet, jotka liittyvät valta-, rooli- ja reviiriorientoituneeseen yhteistyöhön. Vain nämä esteet ylittämällä voidaan saavuttaa kommunikatiivinen ja yhteistoiminnallinen yhteistyömalli, jossa osaaminen yhdistetään käytännössä yhdessä toimien. Viimeksi mainittu tavoite on lopulta kaikkien yhteinen etu. Erityisesti julkishallinnossa sen tulisi olla johtava periaate, kun voimavarat ovat rajalliset ja samanaikaisesti korostetaan suurempaa työn tuottavuutta kuin aiemmin. (Lyytinen 2006.)

Monitahoisuus verkoston luontaisetuna

Verkostomaisen toiminnan eräänlainen luontaisetua arviointitoiminnalle on monitahoisuus. Monitahoarvioinnin tärkein lisäarvo on, että mukaan saadaan laajalaisesti erilaiset näkemykset, tulkinnat, kannanotot ja arvioivat lausunnot arvioinnin kohteesta. Monitahoarvioinnissa tehdään tilaa avainryhmien väliselle vuorovaikutukselle ja neuvottelevuudelle. Lähtökohtaperiaatteita ovat oikeudenmukaisuus ja se, että kaikkien osapuolten, myös vähäisimmän hyödynsaajan, ääni tulee kuulluksi. Monitahoarvioinnissa avainryhmät, kuten eri toimijatahot ja sidosryhmät, osallistuvat arvioinnin suunnitteluun ja arviointiin. Näitä ryhmiä ovat tyypillisesti rahoittajat, päätöksentekijät, kehittäjät, kenttätoimijat, intressiryhmät ja vertaiset. Arvioinnin monitahoisuutta voivat toteuttaa myös monitahoisesti kootut arviointiraadit, joissa arvioinnin tuloksia arvioidaan. Usein monitahoarvioinnissa sovelletaan useita arviointimenetelmiä samanaikaisesti. Monitahoisuutta korostavassa näkökulmassa arviointi nähdään sitä paremmaksi, mitä paremmin siinä otetaan huomioon eri tahojen näkemykset. Arvioijat tekevät arviointinsa omien kokemustensa ja näkemystensä perusteella (henkilökohtaisuus), ja monitahoisuus syntyy näiden yksilöarviointien yhteen sovittamisesta (kollektiivinen konsensuaalisuus). Näkemykset arvioinnin perusteista ja johtopäätösten oikeutuksesta voivat vaihdella suuresti, mutta arviointiryhmän tehtävänä on löytää neuvottelevuuden avulla sellainen yhteinen ymmärrys (konsensus), johon se ryhmänä voi sitoutua. (Räisänen 2005, 119.)

Monet asiantuntijat pitävät keskeisten *stakeholdereiden* (sidosryhmien, intressitahojen) osallistumista arviointiin välttämättömänä. Asianosaistahojen osallistuminen arviointiin lisää erityisesti arvioijien mahdollisuuksia selvittää ja syventää omia tulkintojaan. Erilaisten ja toisistaan ehkä radikaalistikin poikkeavien näkökulmien esittäminen monipuolistaa arviointia. Asianosaiset voivat osallistumisellaan tuoda arviointiin myös lisää empiiristä evidenssiä. On syytä huomata, että sidosryhmien osallistumisen kautta hyötyjä saavutetaan arviointiin tulevien näkökulmien ohella myös vastuun jakamisesta. Välittävänä linkkinä toimivat sidosryhmät lisäävät myös mahdollisuuksia vaikuttaa suoraan päätöksentekoon ja mielipiteenmuodostukseen. (Räisänen 2005, 114.)

Toki sidosryhmien käyttöä arvioinnissa on myös kritisoitu muun muassa siksi, että tosiasiallisesti mahdollisuus selvittää arvioijien omia tulkintoja on varsin vähäinen ja sidosryhmistä saatava hyöty näennäinen. Lisäarvo riippuu siitä, onnistutaanko arviointiin valitsemaan ja osallistamaan sidosryhmien piiristä riittävän korkeatasoiset ja osaavat asiantuntijat. Läheskään aina se ei onnistu. Kriitikistä huolimatta esimerkiksi Patton (ks. esim. Patton 1982; 1997) pitää sidosryhmien osallistumista arviointiin välttämättömänä: asianosaiset selvittävät, kyseenalaistavat ja täydentävät arvioijien näkökulmia ja tuovat arviointiryhmään arvioijilta usein puuttuvan kokemuksen ja syvällisen ymmärryksen.

Sidosryhmät ovat tärkeitä myös oppimisen edistämässä sekä toimiessaan sanaanasaattajina, puolestapuhujina ja ”asianajajina” arvioinnin jälkeen. (Emt.)

Monitahoinen arviointi on osallistavaa ja auttaa kytkemään arvioinnin ja kehittämisen kiinteämmin yhteen. Osapuolten tasavertaisuuteen perustuva osallistava arviointi on kokemusten ja näkemysten vaihtoa ruokkiva toimintatapa. Osapuolten näkemysten ja tulkintojen esiin saaminen on välttämätön lähtökohta arviointiyhteistyön kehittämiseksi. Osallistuva verkosto myös pitää huolen siitä, että arviointi on osuvaa ja uudistuu kaiken aikaa. Päämääränä on, että koulutuksen järjestäjät ja oppilaitokset osallistuvat aktiivisesti arviointiin ja sitoutuvat avoimeen vuorovaikutukseen.

Vaikuttavuutta verkostojen välityksellä

Arvioinnin rooleja ja funktioita erittelevässä kirjoituksessaan opetusministeriön tuolloinen kansliapäällikkö Markku Linna (2006) kiinnittää huomiota siihen, että arviointien vaikuttavuuteen liittyy osaltaan, millaisen roolin arviointi koulutuksen kentällä ja kehittämisessä saa. Linnan mukaan arvioija voi olla myös ”uskotun miehen” roolissa. Silloin on tärkeää, kenen tai minkä tahon kanssa hän liittoutuu. Pahimmillaan arvioitsijoista tulee liian suuria muutoksia vierastavan organisaation johdon liittolaisia. Arviointi saa tällöin ulkoisen legitimoijan roolin. Toinen mahdollisuus on, että arvioinnissa otetaan positiivisella tavalla huomioon organisaation aidot kehittämispyrkimykset. Linna viittaa brittiläisen arviointitutkija Lee Harveyn analyysiin, jossa hän kävi läpi kaikki tärkeimmät korkeakoulujen arvioinnit Isossa-Britanniassa. Analyysi paljasti, ettei arvioinneilla ollut juuri mitään yhteyttä organisaation omiin opetuksen ja tutkimuksen kehittämishankkeisiin. Arviointi voi jäädä irralliseksi puuhasteluksi, jonka tulokset eivät ole ainoastaan laihoja vaan joskus jopa vahingollisia.

Kentän käytännön toimijoiden käyttäminen jo arvioinnin suunnittelussa on omiaan osallistamaan koko kenttää, mikä puolestaan edistää arviointitulosten hyödyntämistä ja vaikuttavuutta. Kun eri tahot tuovat yhteiseen pöytään oman tulkintansa arvioitavasta kohteesta ja osallistuvat arvioinnin muotoiluun, tulosten hyödynnettävyys ja sitä kautta käytön todennäköisyys lisääntyy (esim. Patton 1997; myös Lyytinen & al. 2002). Toisaalta riippumatonta tutkimusta ja arviointiasiantuntemusta edustava arvioijaryhmä on koulutuksen kenttään nähden sillä tavoin ulkopuolisessa asemassa, että sen arviointiin, johtopäätöksiin ja mahdollisiin suosituksiin luotetaan. Arviointiryhmä koostuu itsenäisistä asiantuntijoista, jotka omalla nimellään vastaavat arviointiraportissa arvioinnin tasosta.

Verkostomainen ja osallistava toimintatapa ovat yhdessä hyvä lähtökohta arvioinnin kehittävyydelle. Osapuolten osallistuminen arviointihankkeisiin ja asiantuntijaverkoston rakentamiseen suunnitteluvaiheesta alkaen on omiaan luo-

maan edellytyksiä sille, että arviointien tuloksia myös käytetään koulutuksen kehittämiseksi. Näin arviointi antaa oppilaitoksille, koulutuksen järjestäjille, koulutuspoliittisille päätöksentekijöille sekä muille koulutuksen sidosryhmille relevanttia tietoa prosesseista, jotka edistävät yksilöiden ja organisaatioiden oppimista. Ylipäättäänkin verkostomaisella arviointitoiminnalla edistetään yhteistä dialogia ja arvokeskustelua koulutusta ja kasvatusta järjestävissä työyhteisöissä ja koko yhteiskunnassa.

Verkostomaisuus kytkeytyykin elimellisesti kehittävän arvioinnin paradigmaan. Nykyisin jo vanhahtavaksikin luonnehdittua tarkastustoimintaa harjoittamaan saattaisi hyvin riittää esimerkiksi kentän toimijoista irrallinen virasto. Sen sijaan arvioinnin suunnittelu ja suorittaminen verkostona on kehittävän arvioinnin näkökulmasta mitä luontevin toimintamalli. Kehittävän arvioinnin paradigmassa verkostomaisuus on siis hyvin olennainen työkalu ja edellytys kaikkein tärkeimmän päämäärän eli koulutuksen kehittämisen saavuttamiseksi.

Verkostomaisuus Koulutuksen arviointineuvoston toiminnassa

Koulutuksen arviointia koskevat säännökset on sisällytetty kaikkeen vuonna 1999 voimaan tulleessa koulutusta koskevassa lainsäädännössä tarkoitettuun koulutukseen. Koulutuksen arvioinnista on säädetty samansisältöisesti perusopetuslain (628/1998) 21 §:ssä, lukiolain (629/1998) 16 §:ssä, ammatillisesta koulutuksesta annetun lain (630/1998) 24 §:ssä, ammatillisesta aikuiskoulutuksesta annetun lain (631/1998) 15 §:ssä, vapaasta sivistystyöstä annetun lain (632/1998) 7 §:ssä sekä taiteen perusopetuksesta annetun lain (633/1998) 7 §:ssä. Koulutusta koskevien lakien koulutuksen arviointisäännökset on uudistettu 1. päivänä maaliskuuta 2003 voimaan tulleilla laeilla, joilla perustettiin opetusministeriön yhteyteen erillinen koulutuksen arviointineuvosto.

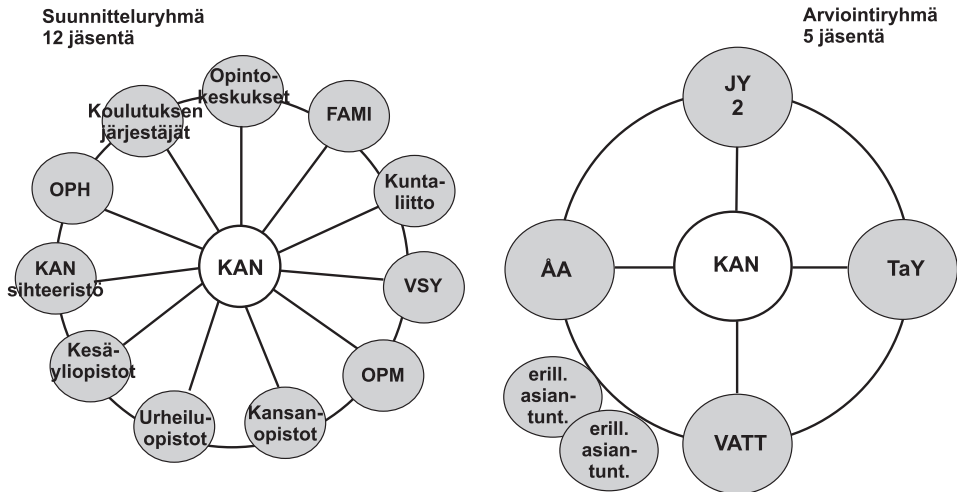
Edellä mainittujen säännösten mukaan koulutuksen järjestäjän tulee itse arvioida antamaansa koulutusta ja sen vaikuttavuutta sekä osallistua ulkopuoliseen toimintansa arviointiin. Paikallisesta arviointitoiminnasta ja sen kehittämisestä vastaavat ensisijaisesti koulutuksen järjestäjät. Koulutuksen arviointineuvoston perustamisen myötä koulutuksen arviointi siirrettiin hallinnosta riippumattomien asiantuntijoiden verkostolle. Uuteen arviointijärjestelmään siirtymisellä haluttiin varmistaa riippumattomuuden ohella arvioinnin osuvuus, tehokkuus ja taloudellisuus sekä vaikuttavuus. Lähtökohtana on se, että verkostomalli mahdollistaa syvällisesti erikoistuneiden asiantuntijoiden käyttämisen kunkin arviointikohteen tarpeiden mukaisesti. Verkostomaisen toimintatavan avulla edistetään tieteen, opetushallinnon, käytännön opetustyön ja sidosryhmien sekä muiden tarpeellisten tahojen asiantuntijoiden yhteistyötä.

Koulutuksen arviointineuvoston perustamiseen päädyttiin selvitysmiehenä toimineen professori Jarmo Visakorven laatimaan raportin pohjalta (Koulutuksen arviointineuvoston ja sen sihteeristön vaihtoehtoiset sijoitusratkaisut, 2002). Sitä oli edeltänyt opetusministeriön eduskunnalle 15.3.2001 antama selvitys, jossa esiteltiin kokemuksia koulutuksen arviointitoiminnasta, arviointitoiminnan organisoinnista sekä arviointijärjestelmän toimivuudesta. Arviointitoiminnan organisoimista tuolloin valmistellut opetusministeriön työryhmä päätyi ratkaisuun, jonka mukaan arviointineuvosto toimisi kevyenä perusorganisaationa ja varsinaisen arviointityön järjestettäisiin pääasiallisesti ulkopuolisten arviointiryhmien toimesta. Toimintamalli oli tuttu Korkeakoulujen arviointineuvoston toiminnasta (ks. Moitus & Saari 2004). Visakorven raportin mukaan ”käytännön arviointityön suorittamista varten on syytä perustaa arviointiverkosto, johon kuuluisivat edellä mainituissa yliopistoissa ja mahdollisesti muissakin yliopistoissa ja organisaatioissa toimivat arviointiyksiköt ja opetushallituksen vastaava järjestelmä. Verkoston kokoonpanosta ja toimivuudesta vastaisi arviointineuvosto, joka myös siis toimisi pääasiallisena työnantajana. Verkoston kullakin yksiköllä tulisi olla oma erikoisalansa, joka perustuisi paikalliseen osaamiseen. Verkostosta ei tarvitsisi säätää erikseen, mutta ajatuksen tulisi sisältyä arviointineuvoston toiminta-suunnitelmaan.” (Visakorpi 2002, 20–21.) Visakorven mukaan arviointineuvoston vastaisi arviointitoiminnan suunnittelusta, johtamisesta, kehittämisestä ja kansainvälisistä yhteyksistä, ja arvioinnit tehtäisiin pääasiallisesti neuvoston ulkopuolisin voimin.

Arviointineuvoston toimintaa koskeva keskeinen sanoma laissa kiteytyy ”ulkopuolisen arvioinnin” ja ”organisoimisen” ohella ilmaisussa ”arviointiasiantuntijoiden verkosto”. Esimerkiksi paikallisesta arviointitoiminnasta ja sen kehittämisestä vastaavat ensisijaisesti koulutuksen järjestäjät, kuten kuntalaki edellyttää. ”Organisoinnilla” laissa viitataan juuri siihen, että Koulutuksen arviointineuvosto itse ei arvioi, vaan organisoii arvioinnit toteutettavaksi verkoston avulla. Laissa verkoston toimijoista mainitaan erikseen yliopistot tieteellisen tutkimuksen tärkeimpänä instituutiona sekä Koulutuksen arviointineuvostoa edeltävän ajan keskeinen arvioinnin toteuttaja, Opetushallitus.

Kuviossa 3 havainnollistetaan verkostojen hyödyntämistä Koulutuksen arviointineuvoston organisoimassa vapaan sivistystyön arviointihankkeessa. Suunnitteluvaiheesta lähtien vapaan sivistystyön eri työmuodot (kansanopistot, kansalaisopistot, opintokeskukset, liikunnan koulutuskeskukset, kesäyliopistot) ovat tuoneet omat näkökulmansa, kuten myös esimerkiksi tiedonkeruujärjestelmänsä, arvioinnin käyttöön omien keskusjärjestöjensä kautta.

Verkoston rakentaminen alkaa arvioitavan kohteen kannalta relevanttien toimijoiden identifioimisesta. Toimijat ovat organisaatioita ja ihmisiä, jotka voidaan konkreettisimmillaan identifioida niminä. Verkostomaisessa arviointimallissa on tärkeää, että kentän toimijat osallistuvat toimintaan alusta lähtien. Arvioin-



Kuvio 3. Verkostomaisuus arviointihankkeiden suunnittelussa ja toteutuksessa: esimerkkinä vapaan sivistystyön arviointihanke 2004–2007

tiryhmän ohella myös arviointisuunnitelman tekevä suunnitteluryhmä rekrytoidaan verkostosta, joskin usein eri puolelta verkostoa kuin arviointiryhmä, joka koostuu pääosin arviointi- ja metodiosaamisen sekä tieteellisen tutkimuksen asiantuntijoista. Toki arviointiryhmässäkin voi olla kentän, esimerkiksi työelämän osapuolten piiristä tulevia, asiantuntijoita.

Verkostoilla ja niistä rekrytoiduilla asiantuntijoilla on erilaiset roolit hankkeen etenemisen eri vaiheissa. Tärkeimmässä asemassa ovat suunnittelu- ja arviointiryhmä:

Suunnitteluryhmä

- kerätään kentän, sidos- ja intressiryhmien ja kohteen asiantuntemuksen piiristä
- laatii suunnitelman ja hankkeen etenemisaikataulun eri vaiheineen
- tekee neuvostolle esityksen arviointiryhmään nimettävistä arvioijista.

Arviointiryhmä

- muodostuu riippumattomista asiantuntijoista arvioinnin ja tutkimuksen piiristä
- suorittaa arvioinnin ja tekee johtopäätökset suosituksineen
- laatii arviointiraportin ja vastaa siitä.

Arviointihankkeen suunnitteluryhmän kohdalla osallistuminen ja toimintaan sitoutuminen eivät ensialkuun näytä ongelmallisilta. Yleensä asianosaiset haluaivat olla vaikuttamassa siihen, millainen kuva heidän omasta toiminnastaan luo-

daan ja miten he voivat olla itse kehittämässä tuota kuvaa ja sen perustana olevaa todellisuutta. Ainakin he haluavat pitää huolen siitä, että arviointi on heidän oman toimintansa kannalta relevanttia. Asianosaisten sitoutumista toimintaan edesauttaa se, että osallistumalla aktiivisesti he samalla huolehtivat siitä, että arvioinnissa otetaan huomioon heidän omasta näkökulmastaan relevantit tekijät.

Verkostomaisen arviointitoiminnan arviointia

Koulutuksen arviointineuvoston ensimmäinen toimintakausi vuosina 2003–2007 on ollut eräänlaista verkostomaisen toimintatavan rakentamista ja sisäänajoa. Uuden toiminnan alkuvaiheessa keskeinen tehtävä on nimenomaan verkostojen luominen ja niiden toiminnan ja jatkuvuuden edellytysten rakentaminen. Neuvosto ja sihteeristö ovat luoneet perustaa verkostomaiselle toimintatavalle kirjoittamalla suomalaisen arviointiosaamisen kenttää tekemällä verkostovierailuja yliopistoissa ja tutkimuslaitoksissa sekä muissa koulutustutkimuksen ja arvioinnin asiantuntijaorganisaatioissa. Näissä verkostovierailuissa arviointisihteeristö tapasi ja kuuli noin kahtasataa asiantuntijaa vuonna 2004. Koulutuksen arviointineuvosto laaditutti ensimmäisestä toimintakaudesta arviointiraportin emeritusprofessori Tapio Vahervan johdolla 2007.

Jo pelkästään uutuutensa takia verkostomaisella toimintatavalla on monia esteitä ylitettävänä. Kysymys on tottumattomuudesta uuteen toimintatapaan. Aiemmin viranomaistoimintaa on harjoitettu sulkeutuneista virastoista käsin, ja verkostokulttuuri on ollut kehittymätöntä. Toiminta on tyypillisesti järjestynyt ylhäältä alas. Virastoperinteeseen kasvaneiden organisaatioiden ja ihmisten on vaikea nähdä uudella tavalla järjestetystä arviointitoiminnasta saatavaa yhteistä etua ja verkostoitumisen synergiaetuja. Synergiaetujen maksimoiminen edellyttää sitä, että verkoston jäsenet tuovat toimintaan ydinosaaamistaan. Ydinosaaaminen on organisaation tarkoin valitsema keskittymisalue, johon se suuntaa olennaisen osan voimavaroistaan ja jota kilpailijoiden on yleensä vaikea kopioida. Jotkut voivat pelätä verkostoon osallistumisen kautta menettävänsä yksinoikeuden kehittämäänsä ydinosaaamiseen, johon heidän kykynsä tuottaa lisäarvoa asiakkaille paljolti perustuu (Viitala 2003, 175–182; vrt. Karppi 1994).

Vaikka puolueettoman ja hallinnosta riippumattoman arvioinnin tarve on eduskunnan päätösten mukaan ilmeinen, menneinä vuosina arviointitoimintaan vakiintuneet intressit eivät ole ylitettävissä ilman vastustusta. Kenenkään ei ole helppo luopua saavutetuista eduistaan ja vakiintuneista intresseistään varsinkin, jos se merkitsee vallan ja voimavarojen siirtymistä itseltä pois.

Arviointihankkeiden suunnittelun monitahoisuudella tavoitellaan kentän sitoutumista ja arviointitulosten hyödynnettävyyttä, mutta tällaiseen suunnittelumalliin sisältyy myös rajoituksia ja riskejä. Mikäli monitahoisessa suunnittelussa

mukana olevat näkevät oman tehtävänsä nimenomaan taustaryhmänsä edunvalvontana, mielekkään kokonaisuuden saavuttaminen vaikeutuu. Voi käydä niin, että kaivautuessaan omiin poteroihinsa eri intressiryhmiä edustavat henkilöt antavat oman asiantuntemuksensa suunnittelun käyttöön vain niiltä osin, joka hyödyttää heidän erityisintressejään. Näin he eivät toimi yhteisen päämäärän hyväksi vaan ainoastaan valvovat, ettei heidän etunäkökohtiaan jätetä hankkeessa huomiotta. Verkostoon toki aina osallistutaan omista etunäkökohdista lähtien, mutta oman edun tavoittelun pitäisi lopulta johtaa johonkin yhteiseen etuun. Jotta verkostoa voisi luonnehtia osaamispääoman kasaumaksi, sen pitäisi palvella päämäärää, joka on yhteisesti sovittu tavoitteeksi ja joka ylittää yksittäiset etunäkökohdat (Hailikari & al. 2000, 15). Muutoin verkosto jää vain kilpaken-täksi, jossa osapuolet pyrkivät maksimaalisesti hyötymään toisistaan antamatta kuitenkaan omaa osaamistaan yhteiseen käyttöön.

Paitsi ryhmäkohtaisia intressejä arviointi sivuaa tietysti läheisesti myös val-lankäyttöä. Vaikka arvioinnin päämääränä nykyisin paljolti pidetään toiminnan kehittämistä, arvioinnin kohteena oleva ei välttämättä halua ulkopuolisia arvioijia esittämään arviota toimintansa laadukkuudesta ja tasosta, prosessien tarkoituksenmukaisuudesta ja tulokellisuudesta tai tuloksettomuudesta. Kun pelätään, mitä arviointi ”paljastaa”, ei mielellään haluta edistää sen enempää verkostomaista kuin muutakaan ulkopuolista arviointia.

Luottamuksen puutteen lisäksi verkostotoiminnan kehittymistä voi estää myös pelko verkostoon osallistumisen aiheuttamasta lisätyöstä ja resursseja vievistä kustannuksista. Uuden arviointitoiminnan vähäisiksi tiedetyt resurssit eivät ole omiaan houkuttamaan yliopistojen, tutkimuslaitosten ja muiden arviointiosaajien osallistumista verkostomaiseen arviointiin. Koulutuksen arviointineuvoston ensimmäisen toimintakauden yksi tärkeimmistä opetuksista koskien verkostomaista toimintatapaa on, että arviointiryhmien toimintaan on syytä resursoida enemmän voimavaroja. Arviointiryhmät on koottu alansa huippuasiantuntijoista, jotka ovat kiireisinä kiinni omassa päätyössään ja myös ylityöllistettyinä mukana monissa muissa hankkeissa. Arvioijien avuksi kaivataan lisää työvoimaa, jotta heidän tietämyksensä ja asiantuntemuksensa saadaan mahdollisimman täysimääräisesti hankkeen käyttöön.

Toisaalta pelkästään taloudellisista intresseistä lähtevä osallistuminen arviointihankkeisiin ei takaa verkostokulttuurin kehittymistä. Osallistumisen tulee myös kehittää eri osapuolia ja edistää jopa organisaation kilpailukykyä tutkimuksen, arvioinnin ja kehittämishankkeiden osalta. Tulosohjauksen voimistuminen ja tuottavuuspyrkimykset saattavat aiheuttaa sen, että organisaatio keskittyykin ainoastaan omiin pyrkimyksiinsä, mikä heikentää tosiasiallisia mahdollisuuksia verkostoitumiseen. Verkostoitumisen ideat eivät äärimmäisen kiireiseksi tiuken-netussa työrytmissä helposti toteudu. Verkostoituminen tuntuu houkuttelevalta, jos ajatellaan arviointia: arviointiosaaminen pystyy nimittäin kehittämään aidos-

sa arviointiympäristössä ja koulutuspoliittisesti merkittävillä ja ajankohtaisilla kohdealoilla. Houkuttelevuus luonnollisesti yhdistyy omaan aiempaan osaamiseen ja erityisalaan erityisesti tutkijoiden piirissä. Osallistuminen arviointiin tuo myös arvostusta, mainetta ja tunnettuutta siihen osallistuville. Niukasti resursoidut ja hartiapankkiperiaatteella toteutetut hankkeet eivät pitkällä tähtäimellä pysty luomaan sellaisia verkostosuhteita, joissa asiantuntijaorganisaatiot haluvat olla mukana. Pelkästään ylevillä tavoitelauseilla ryyditetty hengenpalo ei luo kestävästä verkostokulttuurista eikä sitouta sen kehittämiseen. Hankkeilla tulee olla myös vahva yhteiskunnallinen status. (Silvennoinen 2005; Lyytinen 2006.)

Arviointitoiminta edellyttää erikoishenkilöstöä ja tarkoituksenmukaisen infrastruktuurin uskottavan tiedon tuottamiseksi. Tähän tarvitaan sekä pysyvää että hankkeisiin rekrytoitavaa määräaikaista asiantuntija-henkilöstöä. Metodeihin ja tilastolliseen analyysiin erikoistunut henkilöstö olisi parasta olla pysyvää. Suurten aineistojen keruu oppilaitoksista ja koulutuksen järjestäjiltä edellyttää niin ikään toimivan infrastruktuurin. Yksi arviointiin liittyvä perustehtävä on arviointimenetelmien kehittäminen ja virikkeiden tuottaminen paikallistasolla tapahtuvaa arviointia varten.

Lähteet

- Blomqvist, K. 2004. Partnering in the dynamic environment: The role of trust in asymmetric technology partnership formation. Lappeenranta University of Technology. Acta Universitatis Lappeenrantaensis 122.
- Blomqvist, K. 2006. Luottamus verkostoyhteistyön edellytyksenä. Teoksessa H. Silvennoinen (toim.) Koulutuksen arviointi verkostoituu. Koulutuksen arviointineuvoston julkaisuja 18. Jyväskylä.
- Hailikari, M., Immonen, S., Kokko, N., Herrala, M., Salminen, A. & Ahola, M. 2000. Osamiskeskusmallin kehittäminen kuinka osaaminen saadaan leviämään. Helsinki: Työministeriö.
- Ilmonen, K. & Jokinen, K. 2002. Luottamus modernissa maailmassa. Jyväskylän yliopisto. SoPhi 60.
- Karppi, J. I. 1994. Onko yhteistoiminta mahdollista? Yleinen strategia julkisen intressin ilmentymänä. Teoksessa I. Raatikainen & J. Ahopelto. Verkostoajattelusta verkostotoimintaan. Kuopion yliopiston julkaisuja E. Yhteiskuntatieteet 21.
- Lankinen, T. 2007. Opetushallituksen asema, rooli ja tehtävät sekä koulutustoimialan ohjaus muuttuvassa toimintaympäristössä. Opetusministeriön työryhmämuistioita ja selvityksiä 46. Helsinki.
- Linna, M. 2006. Arviointin roolit ja funktiot. Teoksessa. H. Silvennoinen (toim.) Koulutuksen arviointi verkostoituu. Koulutuksen arviointineuvoston julkaisuja 18. Jyväskylä.
- Lyytinen, H. K. 2006. Koulutuksen arviointiosaaminen verkostossa – mahdollisuuksia ja kehittämishaasteita. Teoksessa. H. Silvennoinen (toim.) Koulutuksen arviointi verkostoituu. Koulutuksen arviointineuvoston julkaisuja 18. Jyväskylä.
- Lyytinen, H. K. & Hämäläinen, K. 2004. Koulutuksen kansallisen arviointin kehittäminen Suomessa. Saatavilla [www-muodossa: <http://www.koulutuksenarviointineuvosto.fi/linja.htm>](http://www.muodossa: <http://www.koulutuksenarviointineuvosto.fi/linja.htm>).

- Lyytinen, H. K., Vaso, J., Kangasoja, M. & Rönkä, A. 2002. Arvioi ja opi, opi ja arvioi: arvioimalla yhteisölliseen oppimiseen. Vantaa: Printel.
- Moitus, S. & Saari, S. 2004. Menetelmistä kehittämiseen. Korkeakoulujen arviointineuvoston arviointimenetelmät vuosina 1996–2003. Helsinki: Korkeakoulujen arviointineuvosto.
- Nikkanen, P. & Lyytinen, H. K. 1996. Oppiva koulu ja itsearviointi. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Ollus, M., Ranta, J. & Ylä-Anttila, P. (toim.) 1998. Yritysverkostot: kilpailua tiedolla, nopeudella ja joustavuudella. Helsinki: Sitra & Taloustieto.
- OPM. 2007. Koulutuksen arviointijärjestelmän kehittämistyöryhmän muistio. Opetusministeriön työryhmämuistioita ja selvityksiä 2007: 27 Helsinki.
- Patton, M. Q. 1982. Practical evaluation. Beverly Hills: Sage.
- Patton, M. Q. 1997. Utilization-focused evaluation. Thousand Oaks: Sage.
- Ruuskanen, P. 2003. Verkostotalous ja luottamus. Jyväskylän yliopisto. SoPhi 78.
- Räisänen, A. 2005. Kehittävään arviointiin. Teoksessa H. K. Lyytinen ja A. Räisänen (toim.) Kehittämissuuntaa arvioinnista. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä.
- Silvennoinen, H. 2005. Arviointitoiminta verkostoihin. Teoksessa H. K. Lyytinen ja A. Räisänen (toim.) Kehittämissuuntaa arvioinnista. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä.
- Viitala, R. 2003. Henkilöstöjohtaminen. Helsinki: Edita.
- Visakorpi, J. 2002. Koulutuksen arviointineuvoston ja sen sihteeristön vaihtoehtoiset sijoitusratkaisut. Selvitysmiehen raportti. Helsinki: Opetusministeriö.
- Vuorenmaa, M. 2001. Ikkunoita arvioinnin tuolle puolen. Uusia avauksia suomalaiseen koulutusta koskevaan evaluaatiokeskusteluun. Jyväskylän yliopisto. Jyväskylä Studies in Education, Psychology and Social Research 176.



Kehittävä arviointi

Kehittävän arvioinnin peruseriaatteena on arviointitiedon välitön hyödyntäminen arvioitavan hankkeen tai kohteen kehittämiseksi. Yleensä koulutuksen tai hankkeen vaikuttavuutta arvioidaan loppuvaiheessa tai hankkeen päättymisen jälkeen, mutta kehittävässä arvioinnissa tuotetaan hankkeen aloitusvaiheesta lähtien arviointitietoa, joka tukee mahdollisimman vaikuttavan ja onnistuneen hankkeen läpiviemistä. Kehittävä arviointi on jatkuvaa vuorovaikutusta käytännön toiminnan ja sen arvioinnin välillä.

Asiasanat: kehittävä arviointi

Jokainen on kouluaikoinaan tutustunut perinteiseen oppilasarviointiin, jonka tehtävä oli peilata oppilaan osaamista ja tuotosta suhteessa opetussuunnitelman tavoitteisiin. Käytännössä arvioinnin lopputulos näkyi oppilaalle arvosanojen 4 ja 10 välille osuvana numeroarviointina. Joskus kokeissa onnistui, joskus taas ei. Kuvitellaanpa, että tällaisen lopussa tapahtuvan kontrolloivan arvioinnin sijasta käytössä olisikin ollut ns. kehittävä arviointi. Oppimista ja sen edellytyksiä olisi arvioitu ja samalla kehitetty heti alusta lähtien, ja varmistettu tavalla tai toisella, että oppilas saavuttaa parhaat mahdolliset tulokset. Tällainen perinteisestä poikkeava kehittävä arviointimalli olisi varsin suuri mullistus kouluoppimiselle. Vastaavalla tavalla kaikki arviointitoiminta mullistuu, jos käyttöön otetaan perinteisten arvottavien, toteavien ja manageriaalisten arviointimallien sijasta arviointikohteen kehittämiseen pyrkivä kehittävä arviointi.

Kehittävän arvioinnin määritelmiä

Kehittävä arviointi ei ole yksittäinen arviointimalli tai menetelmä, vaan enemmänkin lähestymistapa, jonka perusperiaate on, että arviointitietoa hyödynnetään välittömästi arvioitavan hankkeen tai kohteen kehittämisessä. Tällöin puhutaan kehittävästä tai uudistavasta arvioinnista (Kauppi 1992). Arvioinnin avulla tuotetaan tietoa toiminnan kehittämiseksi ja kehittämällä toimintaa tuotetaan arviointitietoa. Kun kehittäminen on luonteeltaan iteratiivista, puhutaan kehittävästä arvioinnista. Uudistava arviointi puolestaan korostaa rakenteellisten muutosten ja laadullisesti erilaisten käytäntöjen kehittämistä.

Käytännössä ero perinteisempiin arviointimalleihin on siinä, että kun esimerkiksi koulutuksen tai hankkeen vaikuttavuutta arvioidaan yleensä loppuvaiheessa tai mielellään vasta viivästetysti hankkeen päättymisen jälkeen, tuotetaan kehittävässä arvioinnissa jo hankkeen aloitusvaiheesta lähtien sellaista arviointitietoa, joka tukee mahdollisimman vaikuttavan ja onnistuneen hankkeen läpiviemistä. Kehittävä arviointi on jatkuvaa vuorovaikutusta käytännön toiminnan ja sen arvioinnin välillä. Arviointi integroituu toiminnan suunnitteluun, toteutukseen ja kehittämiseen, jolloin se voidaan nähdä jatkuvana arviointitiedon keräämisenä sekä arviointitulosten ja niiden merkitysten pohdintana ja tulkintana (ks. Hämäläinen & Kauppi 2000). Tällainen arviointitieto voi olla esimerkiksi hankkeen tavoitteiden ja tarvemäärittelyn ennakoarviointia, käynnistysvaiheen arviointia tai hankkeen edetessä erilaisia ongelmakohtia paikantavaa arviointia. Jotta toiminnan hienosäätö on mahdollista, täytyy arviointitiedon olla välittömästi toimijoiden ulottuvilla ja vielä siinä muodossa, että siitä on hyötyä ja sovellusarvoa käytäntöjen kehittämiseksi.

Kehittävän arvioinnin tehtävä on siis varmistaa arvioitavan hankkeen mahdollisimman hyvä onnistuminen ja vaikuttavuus, kun sen sijaan perinteisemmät arviointimallit tarkastelevat arviointikohdetta neutraalisti ulkoapäin ja usein vasta jälkikäteen. Voidaan tietysti väittää, että myös perinteinen arviointi palvelee toiminnan kehittämistä. Onhan esimerkiksi erilaisten ohjelma-arviointien (esim. Nyyssölä 1995; ks. Heinonen 2001) ja koulutuksen evaluaatiotutkimusten (esim. Räisänen 1995) tavoitteena tarjota päätöksentekijöille eväitä sen suhteen, kannattaako kyseistä hanketta tai ohjelmaa rahoittaa myös jatkossa ja tarvitaanko siinä jotain hienosäätöä ennen sen seuraavaa toteutusta.

Kehittävä arviointi tulee monimutkaisemmaksi, jos arviointitiedon pohjalta tapahtuvaan kehittämiseen halutaan sitouttaa ja valtauttaa erilaisia toimijaryhmiä. Etenkin laajemmissa hankkeissa toimijoiden ottaminen mukaan arviointiin ja sitä kautta toiminnan kehittämiseen on usein onnistumisen perusehto. Eriyisesti dynaamisissa ja monimutkaisissa hankkeissa arvioitavista ilmiöistä on vaikea tuottaa luotettavaa arviointitietoa ajattelun ja tiedon luomiseksi sekä toiminnan kehittämiseksi. Mitattavia tavoitteita ei aina voida määritellä etukäteen,

jolloin tärkeäksi muodostuu toimijoiden osallistaminen jatkuvaan reflektointiin, tiedon luomiseen ja käytäntöjen kehittämiseen. Arvioinnista muodostuu väline toimijoiden valtauttamiseksi oman toimintansa tietoisempaan yhteisölliseen kehittämiseen. Kehittävä arviointi korostaakin usein toimijoiden osallistamista arviointiprosessiin (esim. Cousins & Whitmore 1998). Tämä osallistaminen on välttämätöntä, jotta arviointi voisi aidosti kytkeytyä toiminnan kehittämiseen. Toiminnan hajautuessa verkostoihin tärkeää on, että myös arviointi hajautuu sinne, missä toimitaan ja toimintaa kehitetään. Arviointi muistuttaa tällaisissa tapauksissa kehittävän tutkimuksen prosessia, jossa tiedon luominen, kokeileminen jne. yhdistyvät (ks. Valkama 2002).

Toimijoiden osallistaminen ja ”kiusaaminen” ei ole välttämätön kehittävän arvioinnin tunnuspiirre, vaikka se laajemmissa hankkeissa onkin perusteltua. Välittömän palautteen ja kehittävän arviointiotteen periaate voi toteutua myös muulla tavalla, esim. ns. ”Rolling Feedback” -mallin avulla (Brown & Evans 1990). Arviointia tekevän tutkijan tehtävä on tällöin tuottaa mahdollisimman pienellä viiveellä toiminnan kehittämistä tukevaa arviointitietoa arvioitavan hankkeen vetäjille ja toimijoille. He tekevät sitten arviointitiedon pohjalta johtopäätöksiä ja tarvittavia muutoksia hankkeen kehittämiseksi.

Kehittävän arvioinnin metodologinen tausta

Vaikka evaluaatio ja tutkimus on perinteisesti pyritty erottamaan toisistaan (Popham 1988), löytyy kehittävän arvioinnin ja perinteisen arvioinnin väliltä sama ero kuin perinteisen rationalistisen (positivistisen) tiedekäsityksen ja ns. naturalistisen paradigman (Guba & Lincoln 1989) väliltä. Luonnontieteistä periytyvän positivistisen tiedekäsityksen mukaan tutkijan (arvioijan) tulee suhtautua neutraalisti tutkimuskohteeseen, eikä sen ”elämään” saa missään tapauksessa mennä puuttumaan, koska tällainen interventio vaikuttaisi tuloksiin. Naturalistisessa yhteiskuntatieteisiin soveltuvassa paradigmassa todellisuus nähdään sen sijaan moniarvoisena, holistisena ja sosiaalisesti rakentuvana ilmiönä, joka ei ole ennustettavissa, mutta kuitenkin ymmärrettävissä. Lähtöoletus on, että tutkijan/arvioijan läsnäolo vaikuttaa aina joka tapauksessa. Kehittävässä arvioinnissa tämä vaikutusmahdollisuus otetaan vakavasti ja viedään vielä pidemmälle kehittävien interventioiden avulla.

Kehittävä arviointi on hyvin lähellä toimintatutkimusta (Action Research) ja kehittävää työntutkimusta. Tiedepiireissä toimintatutkimusta on perinteisesti pidetty ei-tieteellisenä, osin siitä syystä, että siinä rikotaan edellä mainittuja positivistisen tieteen sääntöjä ja ”sohaistaan tietoisesti muurahaispesää” (Kuula 1999). Toisaalta toimintatutkimusmallia on kritisoitu myös teoriattomuudesta, johon taas kehittävä työntutkimus pyrkii tarjoamaan ratkaisun (Engeström 1995).

Kehittävän arvioinnin ja toimintatutkimuksen välinen ero on hyvin häilyvä, ja joskus parhaat ja laajemmat kehittävän arvioinnin hankkeet ovat hyvin lähellä kehittävää työntutkimusta. Ainoa ero kehittävän arvioinnin ja kehittämistutkimusotteiden välillä onkin se, että kehittämistutkimuksessa tuotetaan ilmiöitä muuttamalla tieteellistä tietoa (esim. siitä kuinka ilmiöt toimivat), kun taas kehittävässä arvioinnissa on tavoitteena ilmiöiden kehittäminen (kuinka ilmiö saadaan toimimaan paremmin).

Koska kehittävä arviointi on enemmän lähestymistapa kuin menetelmä, ei siinä oletusarvoisesti sitouduta tiettyihin metodologisiin periaatteisiin. Arviointimenetelmät ja aineistot voivat siten olla yhtä hyvin määrällisiä kuin laadullisiakin. Esimerkiksi tilastoaineistojen ja -analyysien avulla voidaan hyvin nostaa esille erilaisia ongelmakohtia, joiden tiedostaminen auttaa hankkeen kehittämistä (esim. osallistujarakenteiden vääristyminen, jolloin voidaan muuttaa markkinointi- ja rekrytointistrategiaa). Kehitettävät hankkeet ovat kuitenkin yleensä sentyyppisiä, että kehittävää arviointia on hyvin vaikea tehdä ilman laadullisia menetelmiä. Perinteisempien teemahaastattelujen rinnalla arvioinnin toteutus vaatii hyvin usein etnografista lähestymistapaa ja ns. illuminatiivista (Parlett & Hamilton 1976) arviointia.

Kehittävän arvioinnin perusluonteesta johtuen on olennaista, että arviointi on prosessiarviointia, eli etenee hankkeen alkuvaiheesta lähtien hankkeen loppuun ja vaikuttavuuden arviointiin asti. Hyvin toimiva on esimerkiksi CIPP-malliin (Context, Input, Process, Product; Stufflebeam 1985) pohjautuva lähestymistapa, jota kuvataan tarkemmin artikkelin lopussa.

Kuten edellä todettiin, etenkin monimutkaisemmissa ja useita toimijoita käsitävissä hankkeissa on olennaista myös toimijoiden osallistaminen, jolloin kehittävä arviointi lähenee osallistavaa (participatory; King 1998; Garaway 1995) arviointia. Vastaavasti voidaan käyttää erilaisia itsearviointimalleja ja työkaluja. Toimiva on myös sidosryhmäpohjainen arviointi (stakeholder evaluation; Ayers 1987). Vastaavantyyppiset periaatteet liitetään yleensä toimintatutkimukseen (ks. Heikkinen & Jyrkämä 1999, 36).

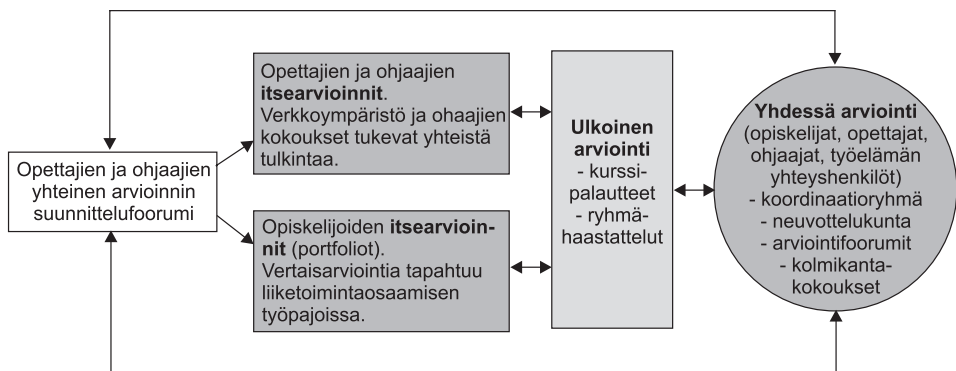
Esimerkki 1: kehittävä arviointi koulutusohjelman arvioinnissa

Helia Ammatillisessa opettajakorkeakoulussa kehittävällä arvioinnilla on viitattu arviointiin, jonka tarkoituksena on kehittää organisaation toimintaa ja sen tietoperustaa. Lähestymistapa on saanut vaikutteita useista eri arviointisuuntauksista ja -malleista (ks. Guba & Lincoln 1989; Raivola 1995; Valkama & Vuorela 1995; Hämäläinen 2002; Hakkarainen 1991; ks. Kauppi ym. 2003), ja siinä painottuvat erityisesti

- arvioinnin hahmottaminen yhteisön oppimisprosessina
- toimijoiden osallistuminen arviointiprosessiin
- moniarvoisuuden salliminen
- arvioinnin ja toiminnan integroituminen
- arvioinnin kontekstuaalisuus.

Helian kehittävän arvioinnin projektissa sovellettiin ja kehitettiin kehittävän arvioinnin malleja ja välineitä erilaisissa pilottiprojekteissa opettajien perus- ja täydennyskoulutuksessa vuosina 1999–2004. Ne liittyivät oppimistehtävien, ohjauksen ja koulutusohjelmien arviointeihin. Projektin tulema on koottu julkaisuun Kehittävä arviointi – periaatteita ja sovelluksia (Kauppi 2005).

Yhtenä projektin pilottina kehitettiin amk-jatkotutkintoa, erillisiä opintoja ja opinnäytetyötä itsearvioinnin ja sitä tukevan ulkoisen arvioinnin avulla (Noussiainen, Häyrynen & Kauppi 2004). Arvioinnin keskeisiä periaatteita olivat arvioinnin prosessinomaisuus ja jäsentyminen osaksi oppimis- ja opetustoimintaa toimijoiden itsearvioinnin kautta. Arviointitiedon hyödyntämiseksi toiminnan kehittämiseksi sidosryhmät osallistuvat arvioinnin suunnitteluun ja kerätyn tiedon tulkintaan yhteisillä foorumeilla. Arvioinnin luotettavuutta lisättiin systemaattisella tiedon keräämisellä ja dokumentoinnilla sekä näihin perustuvalla ulkoisella arvioinnilla. Seuraava kuvio kuvaa arvioinnin kokonaisuutta.



Kuvio 1. Amk-jatkotutkintokoulutuksen arviointiprosessi (Noussiainen, Häyrynen & Kauppi 2004, 82)

Arvioinnin kohteeksi rajattiin opinnäytetyöprosessi, koska se muodosti jatkokutkinto-opintojen punaisen langan. Arviointitiedon pääasiallisena keruutapana käytettiin ohjaajien ja opettajien tekemiä itsearviointeja ja opiskelijoiden portfolioita, haastatteluita sekä kurssipalautteita. Tärkeäksi osaksi arviointiprosessia määriteltiin arvioinnin eri sidosryhmien (opiskelijat, opettajat ja ohjaajat, työ-

elämän yhteyshenkilöt) yhteiset foorumit, joissa tulkittiin ja arvioitiin kerättyä tietoa ja tehtiin arvioinnin perusteella kehittämisehdotuksia. Hyödyntämällä mahdollisimman pitkälle jo olemassa olevia arvioinnin muotoja ja foorumeita rakennettiin arviointia aidosti osaksi toimintaa ja sen kehittämistä.

Esimerkki 2: kehittävä arviointi valtakunnallisissa koulujen kehittämishankkeissa

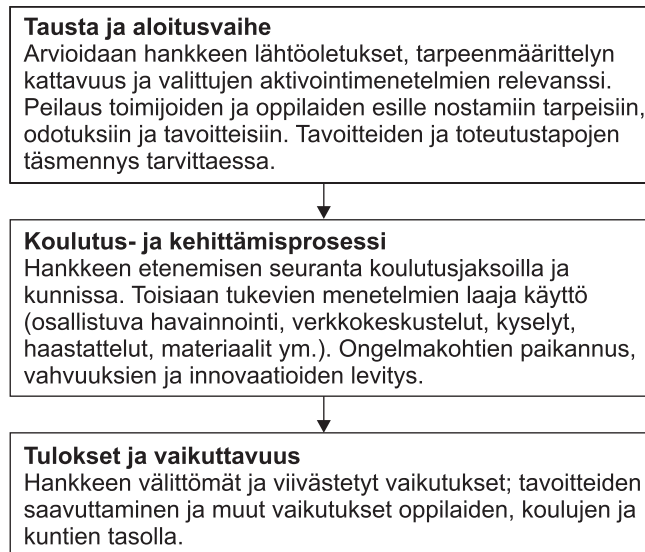
Helsingin yliopiston Koulutus- ja kehittämiskeskus Palmenia on ollut toteuttamassa kehittävä arviointia kahdessa laajassa OPM:n koulujen kehittämishankkeessa. Osallistuva oppilas – yhteisöllinen koulu -hankkeessa (2005–2007) tavoitteena on kehittää oppilaiden osallistumista ja osallisuutta omassa koulussa ja kunnassa. Joustava perusopetus -hankkeessa (2006–2008) kehitetään koulupudokkaiden ehkäisemiseksi yksilöllisiä ja joustavia opiskelutapoja perusopetukseen. Molemmissa hankkeissa on mukana yli sata opettajaa, rehtoria ja kuntien edustajaa joka puolelta Suomea. Hankkeiden perusmalli on molemmissa tapauksissa sama, eli paikallisia toimijoiden vetämiä kehittämishankkeita tuetaan alueellisilla koulutuksilla.

Hankkeiden laajuus, kesto, maantieteelliset etäisyydet ja toimijoiden suuri määrä asettavat erityisiä vaatimuksia kehittävä arvioinnille. Ratkaistavia ongelmia ovat erityisesti kerättävän arviointitiedon luonne, aineistonkeruumenetelmät, tulosten välittäminen toimijoille ja hankkeen vetäjille sekä ennen kaikkea toimijoiden osallistaminen arviointiin. Koulujen arjessa toimivien oto-kehittäjien aika on rajallinen resurssi, joten arviointi pitää toteuttaa mahdollisimman vähän lisätyötä tuottavalla tavalla, mutta kuitenkin siten, että osallistavan arvioinnin periaatteet toteutuvat ja arviointi nivoutuu saumattomasti kehittämiseen.

Aineistonkeruumenetelmiksi valittiin osallistuva havainnointi (koulutusjakso), verkkolomakkeet (koulutuspalautte, toimintatietojen keruu), teemahaastattelut, ryhmähaastattelut ja keskustelut koulutusjaksoilla sekä verkkopohjaisen projektitympäristön käyttö. Verkkopohjainen yhteistyöympäristö on käytännössä ainoa toimiva vaihtoehto, joka arvioinnin lisäksi palvelee myös toimijoiden välistä yhteydenpitoa ja vuorovaikutusta, hankkeen ohjausta ja erilaisten materiaalien jakamista. Verkkoympäristö on toteutettu Palmenian Studium-alustalle (www.studium.helsinki.fi), ja se sisältää keskustelualueet, sisäpostin, kalenterin, materiaalipankkeja ja erilaisia verkkokyselyjä tarpeen mukaan. Kouluissa kehittämishankkeista kerätään pääasiassa autenttista aineistoa eli toiminnan ja koulutuksen yhteydessä luonnostaan syntyvää aineistoa (esim. toimintasuunnitelmat). Lisäksi eri vaiheissa on kehitetty yhdessä kouluttajien kanssa tiedonkeruulomakkeita, joihin sisältyy toiminnan kehittämistä tukevia kysymyksiä (esi-

merkiksi kuntaryhmien yhteistyötä vaativia kysymyksiä, jolloin aineistonkeruu palvelee samalla yhteistyön laadun ja määrän reflektointia ryhmän sisällä).

Viitekehyksenä toimii prosessiarviointi. Esimerkiksi Joustava perusopetus -hankkeessa arviointi on toteutettu seuraavan CIPP-mallin vaiheistusta mukailevan kaavion mukaan.



Molemmissa hankkeissa kehittävän arvioinnin periaatteena on ollut arviointitulosten mahdollisimman nopea toimittaminen toimijoiden ja kehittäjien (projektipäällikkö, kouluttajat, rahoittaja, ohjausryhmä) tietoon. Käytännössä esimerkiksi koulutusjaksoilta kootut verkkopalautteet analysoidaan mahdollisimman pian ja käydään läpi yhdessä kouluttajien kanssa, jotta mahdollista hienosäätöä voidaan tehdä ennen seuraavaa koulutusjaksoa. Osallistuvan havainnoinnin kautta saadut tulokset käsitellään välittömästi paikan päällä tai heti koulutuksen jälkeen arviointipalaverissa. Tulokset ovat saatavissa myös verkkoalustalla, ja niistä käydään keskustelua keskustelualueilla ja koulutusjaksoilla. Yksi keskeinen kehittämisen työväline on hankkeen ohjausryhmä, joka määrääjain käsittelee arvioinnin väliraportit ja arviointitulokset ja ohjaa hankkeen etenemistä.

Johtopäätöksiä

Kehittävässä arvioinnissa keskeiseltä vaikuttaa käytännön toimijoiden osallistaminen arviointiin, mikä etenkin laajemmissa hankkeissa mahdollistaa arvioinnin ja kehittämisen saumattoman vuoropuhelun. Samalla toimijat ja eri sidosryhmi-

en edustajat voivat osallistua keskusteluun arvioinnin tavoitteista, menetelmistä ja johtopäätöksistä arvioinnin eri vaiheissa. Toimijoiden ja sidosryhmien osallistuminen parantaa myös sitoutumista arvioinnin hyödyntämiseen toiminnassa ja sen kehittämisessä.

Arvioinnin kytkeytyessä toimintaan ja sen kehittämiseen on tärkeää huomioida myös organisaatioiden arviointikulttuuri. Arviointia koskevat perinteiset odotukset ja tottumukset kytkeytyvät usein arvosteluun ja kontrollointiin. Avoimen ja kehittämishakuisen arviointikulttuurin rakentaminen on pitkäkestoinen yhteisöllinen prosessi, jossa luottamuksen kulttuurin rakentuminen kytkeytyy myös organisaatioiden johtamiseen.

Lähteet

- Ayers, T. 1987. Stakeholders as partners in Evaluation: A stakeholder-collaborative Approach. *Evaluation and Program Planning* 10 (3), 263–271.
- Brown, A. & Evans, K. 1990. Rolling feedback strategy: An approach to evaluation in dynamic contexts. *Evaluation and Research in Education* 4 (1), 129–145.
- Cousins, J. & Whitmore, E. 1998. Framing participatory evaluation. *New directions for evaluation* 80, 5–23.
- Engeström, Y. 1995. Kehittävä työntutkimus. Perusteita, tuloksia ja haasteita. Helsinki: Hallinnon kehittämiskeskus.
- Garaway, G. 1995. Participatory evaluation. *Studies in Educational Evaluation* 21, 85–102.
- Guba, E. G. & Lincoln, Y. S. 1989. *Fourth generation evaluation*. Newbury Park: Sage.
- Hakkarainen, P. 1991. Opetuksen ja oppimisen laadullinen arviointi korkeakouluissa. Jyväskylän yliopisto. Koulutuksen tutkimuslaitoksen julkaisusarja B71.
- Heikkinen, H. & Jyrämä, J. 1999. Mitä on toimintatutkimus? Teoksessa H. L. T. Heikkinen, R. Huttunen & P. Moilanen (toim.) *Siinä tutkija missä tekijä – toimintatutkimuksen perusteita ja näköaloja*. Jyväskylä: Atena kustannus.
- Heinonen, S. 2001. Policy perspective on programme evaluation. Case of a labour market oriented training programme in Finland. Teoksessa S. Kontiainen & K. Nurmi (toim.) *Effectiveness of adult education. Historical and evaluative studies in Finland*. University of Helsinki: Department of Education.
- Hämäläinen, K. 2002. Arvioinnin onnistumisen edellytyksiä. Teoksessa K. Hämäläinen & M. Kaartinen-Koutaniemi (toim.) *Benchmarking korkeakoulujen kehittämisvälineenä*. Korkeakoulujen arviointineuvoston julkaisuja 13/2002. Helsinki: Edita, 7–9.
- Hämäläinen, K. & Kauppi, A. 2000. Evaluation in today's education – From control to empowerment. *Lifelong Learning in Europe* V, 1, 68–75.
- Kauppi, A. 1992. Towards transformative evaluation of training. Teoksessa M. Parjanen (toim.) *Legitimation in adult education*. University of Tampere. Institute for Extension Studies. Publications A2/92.
- Kauppi, A. 1993. Mistä nousee oppimisen mieli? – Kontekstuaalisen oppimiskäsityksen perusteita. Teoksessa *Aikuisen oppimisen uudet muodot*. Helsinki: Aikuiskasvatuksen Tutkimusseura & Kansanvalistusseura.
- Kauppi, A. 1998. Uudistava oppiminen, koulutus ja työn kehittäminen. *Opetushallitus. Tutkimus 2/1998*. Helsinki.
- Kauppi, A. (toim.) 2005. Kehittävä arviointi – periaatteita ja sovelluksia. *Helian julkaisusarja C. Ammatillinen opettajakorkeakoulu* 12: 2005.

- Kauppi, A., Hanski, K., Nousiainen, R. & Lahtiranta, K. 2003. Arviointi pedagogisen kehittämisen välineenä. Teoksessa H. Kotila (toim.) Ammattikorkeakoulupedagogiikka. Helsinki: Edita.
- King, J. A. 1998. Making sense of participatory evaluation practice. *New directions for evaluation* 80, 57–67.
- Kuula, A. 1999. Toimintatutkimus: kenttätyötä ja muutospyrkimyksiä. Tampere: Vastapaino.
- Nousiainen, R., Häyrynen, A. & Kauppi, A. 2004. Amk-jatkotutkintokoulutuksen kehittävä arviointi. Teoksessa E. Okkonen (toim.) Ammattikorkeakoulun jatkotutkinto – toteutuksia ja kokemuksia. Hämeenlinna: Hämeen ammattikorkeakoulu.
- Nyyssölä, K. 1995. Korkeakoulutetut työllisyyskoulutuksessa. Työvoimapolitiittisen työllisyyskoulutuksen ensimmäinen vuosi. Turun yliopisto: Koulutussosiologian tutkimuskeskus. Raportteja 26.
- Parlett, M. & Hamilton, D. 1976. Evaluation as illumination: A new approach to the study of innovatory programs. Teoksessa *Evaluation Studies Review Annual I*, G. V. Glass (toim.) Beverly Hills, CA: Sage, 140–157.
- Popham, W. J. 1988. Educational evaluation. New Jersey: Prentice Hall.
- Räisänen, H. 1995. Työvoimakoulutuksen tuloksellisuus. Työllisyys, ammatillisuus ja kohdentuminen. Työministeriö. Työpoliittinen tutkimus 94. Helsinki.
- Stufflebeam, D. 1985. Stufflebeam's improvement-oriented evaluation. Teoksessa D. L. Stufflebeam & A. J. Shinkfield (toim.) *Systematic evaluation*. Boston: Kluwer-Nijhoff.
- Valkama, H. 2002. Kehittämistutkimus – idean kehittelyä ja ankkurointia. Metodologia-projekti, työpapereita 1/2002. Helsinki: Helia Ammatillinen opettajakorkeakoulu. Saatavilla [www-muodossa: <http://ope.helia.fi/~valha/kehittamistutkimus2.pdf>](http://ope.helia.fi/~valha/kehittamistutkimus2.pdf).
- Valkama, H. & Vuorela, T. 1995. Tulosajattelun umpikujasta kehittävään arviointiin. Teoksessa *aikuiskoulutuksen arviointi. Panoraamoja ja lähikuvia. Aikuiskasvatuksen 36. vuosikirja*. Helsinki: Aikuiskasvatuksen Tutkimusseura ja Kansanvalistusseura.



Itsearviointi kehittävän arvioinnin menetelmänä

Artikkelissa tarkastellaan itsearviointia yhtenä arvioinnin muotona sekä määritetään itsearvioinnin käsitettä. Itsearviointia lähestytään kehittävän arvioinnin menetelmänä. Sen lisäksi artikkelissa tuodaan esille myös sen muita ulottuvuuksia. Itsearvioinnissa keskeistä on omien arvojen esiin nostaminen ja niiden suhteuttaminen toimintaympäristöön. Itsearviointiin liitetään usein myös oman toiminnan reflektointi. Itsearvioinnista voi tulla voimaannuttavaa arviointia, jos yksikössä on halu kehittyä ja resursoida itsearvioinnin toteuttamiseen.

Asiasanat: itsearviointi, kehittävä arviointi

Johdanto

Koulutuksen arviointi on ollut viimeiset vuosikymmenet monen valtion ja kunnan keskeinen menetelmä määriteltäessä koulutuksen arvoa, tehokkuutta ja taloudellisuutta. Arviointi on pohjautunut eri sosio-poliittisissa toimintaympäristöissä ja aikoina erilaisiin menetelmiin. Koulutuksen arvioinnin toteutustavoissa näkyy myös arviointitapojen kehittyminen. Konttisen (1995) mukaan koulutuksen arviointi samaistettiin sen alkuvaiheessa ja 1980-luvulla Suomessa mittaamiseen. Tämän jälkeen sen on nähty toimivan päätöksenteon tukena ja opetussuunnitelmien toteutumisen arviointina sekä erilaisten koulutusohjelmien arviointeina.

Mitä lähemmäksi nykypäivää tulemme, sitä enemmän arvioinneissa korostetaan myös organisaation itsesääätelyä ja omavastuuta sekä paikallisen tason osuutta. (Konttinen 1995, 6.) Arviointitoiminta on yleistynyt niin paljon, että puhutaan arvioivan valtion (*evaluative state*, Neave 1988) tai auditoivan yhteiskunnan (*audit society*, Power 1999) synnystä. 2000-luvulla on kiihtynyt puhe arviointien liittymisestä managerialismiin ja uusliberalistiseen koulutuspolitiikkaan (ks. esim. Ball 2001; Kivirauma 2001).

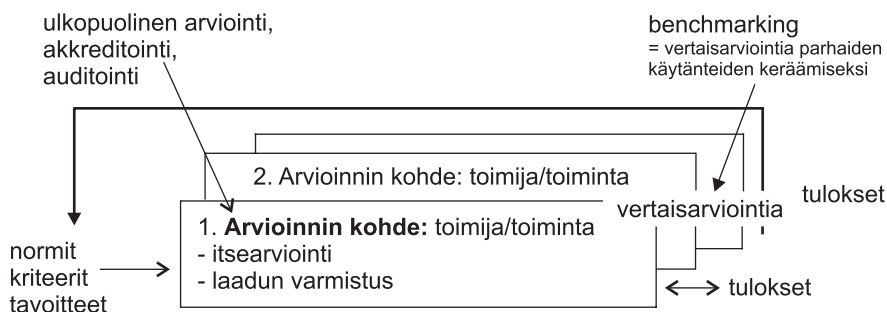
Scrivenin (1991, 139) mukaan arviointi (*evaluation*) tarkoittaa prosesseja, joilla erilaisten asioiden arvoa, ansiota tai hyötyä määritellään tai näiden prosessien tuloksia. Arviointien toteuttamiseen on kehitetty erilaisia keinoja, joista yksi on organisaation tai yksikön, kuten koulun, tekemä itsearviointi. Itsearviointeja tehdään monesta eri syystä, mutta tässä artikkelissa keskityn itsearviointien kehittämislouottuuteen.

Tavoitteenani on pohtia artikkelissa itsearviointien käsitettä sekä sitä, millaisia tehtäviä itsearviointiin liitetään. Itsearviointi sidotaan usein reflektioon ja oman toiminnan kriittiseen tarkasteluun. Artikkelin tarkoituksena on pohtia, mitä itsearviointi voi olla kehittämisen näkökulmasta ja voiko kehittävä arviointi johtaa voimaantumiseen.

Itsearviointien käsite meillä ja muualla

Arviointi on yläkäsite, jonka alla on alakäsitteitä, kuten itsearviointi, ulkoinen arviointi, seuranta (*monitoring*), benchmarking, akkreditointi ja auditointi (ks. kuvio 1).

Saarisen (1995, 20) mukaan itsearviointi tarkoittaa arvioitavan yksikön tai yleensä arvioitavan kohteen omaan toimintaan, sen edellytyksiin ja tuloksiin kohdistamaa tarkastelua. Jakku-Sihvonen ja Heinonen (2001) määrittelevät itse-



Kuvio 1. Itsearviointi arvioinnin kokonaisuudessa

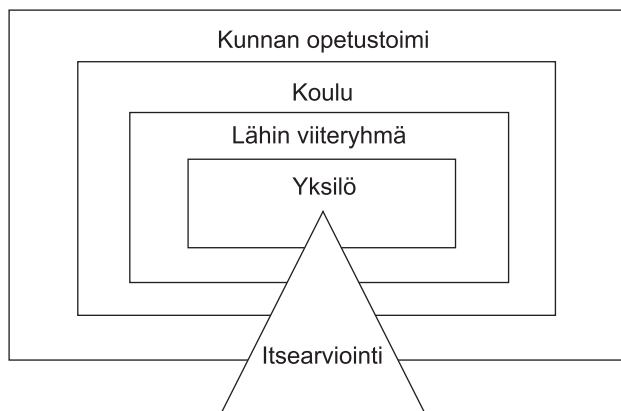
arvioinnin sellaiseksi tiedon hankkimiseksi, jota organisaatio itse kerää itsestään oman toimintansa kuvaamiseksi ja kehittämiseksi. Sen tarkoituksena on toiminnan kehittäminen ja oppimisen edellytysten parantaminen. (Jakku-Sihvonen & Heinonen 2001, 67.) Tästä määritelmästä ei käy kuitenkaan ilmi, kuka arvioinnin tekee. Ulkopuolisten suorittamat arvoinnit kuuluvat myös määritelmän sisään, sillä silloinkin yksikkö saa tietoa itsestään oman toimintansa kehittämiseksi. Tällöin ei kuitenkaan pitäisi puhua itsearviointista, vaan ulkopuolisten tekemät arvoinnit pitäisi selkeästi erottaa omiksi arvioinneikseen. Joskus itsearviointista käytetään termejä sisäinen arviointi ja oman toiminnan arviointi.

Itsearviointin ja sisäisen arvioinnin käsitteet eivät ole vakiintuneita, vaan ne pitää sitoa käyttöyhteyteensä. Englantilaisessa ja amerikkalaisessa arviointikäsitteistössä käsitteitä self-assessment ja self-evaluation käytetään toisinaan myös keskenään päinvastaisissa merkityksissä. Self-assessment on usein mittaamiseen perustuvaa arviointia ja self-evaluation laadullista arviointia. (Saari 2002, 74.) Itsearviointista käytetään Yhdysvalloissa self-study-termiä, joka tarkoittaa lähinnä toiminnan kuvausta. Euroopassa käytetyssä englanninkielisessä termistössä (self-evaluation, self-assessment) itsearviointin odotetaan sisältävän myös oman toiminnan arviointia. (Moitus & Saari 2004, 14.) Powell (2000, 40) jakaa erilaiset itsearviointit sen mukaan, keskittyvätkö ne tuloksiin vai prosesseihin ja onko kohteena organisaatio vai yksilöt. Lisäksi itsearviointit voidaan jakaa kolmeen ryhmään: self-review (laadullinen), self-evaluation (arvoihin liittyvää) ja self-assesment (mittaaminen) (Powell 2000, 40; Saari 2002, 77). Itsearviointien määrittelyyn liittyy se, millaisessa kontekstissa ne toteutetaan, millaisia arvoja niillä halutaan tukea ja millainen rooli reflektiolle annetaan.

Keskeistä useimmissa itsearviointin määritelmissä on, että jokin taho tai yksilö arvioi itse omaa toimintaansa sitä kehittääkseen. Harvey (2002, 7–8) mukaan sisäiseen käyttöön ja ulkoiseen käyttöön tehdyt itsearviointit pitäisi erottaa toisistaan. Saaren (2002) mukaan omaehtoinen itsearviointi johtaa positiivisiin muutoksiin helpommin kuin vain ulkoisten arviointien tulosten vastaanottaminen. Arvioinnin tavoitteet ovat erilaisia omistajuudesta riippuen. Onkin perusteltua olettaa, että itsearviointi kirjoitetaan sen mukaan, käytetäänkö sitä itse oman toiminnan kehittämiseksi vai tuleeko se ulkopuolisten arvioitsijoiden käyttöön. (Saari 2002, 76, 80.)

Itsearviointi on ensisijassa yksikön tai organisaation oman toiminnan arviointia. Itsearviointia tarkasteltaessa onkin keskeistä ja ongelmallista miettiä, kuka tai mikä itse on. (Hämäläinen & Kauppi 2000, 71; Saari 2002, 74.) Kuviossa 2 tarkastellaan, mihin esimerkiksi koulun viitekehyksessä itsearviointi voisi kohdistua.

Kouluympäristössä itse voi olla joko yhteisö- tai yksilötaso. Opettaja tai opilas voi itse arvioida omaa toimintaansa. He voivat tehdä sitä myös lähimmän viiteryhmänsä, kuten samassa asemassa olevien opettajien, muiden saman alan



Kuvio 2. *Itse-toimijat koulun viitekehyksessä Karjalaista (2005, 45) mukaillen*

opettajien tai muiden vertaisten kesken. Oppilaat voivat arvioida toimintaansa luokkana. Koulu voi tehdä itsearviointia organisaationa tai kunnan opetustoimi kaikista kunnan tai tietyn alueen kouluista.

Itsearviointin erilaisista muodoista keskeisimpiä ovat yksilön oman oppimisen tai kehittymisen itsearviointi ja henkilökunnan tekemä arviointi oman työnsä eri osa-alueista ja rakenteista (Scriven 1991, 327). Kellsin (1995, 28) mukaan itsearviointi on henkilökunnan toteuttamaa kuvailua tai analyysia omista suorituksista, organisaation tilasta, prosesseista ja tuloksista. Yleensä itsearviointin avulla tuotetaan raportti, joka helpottaa oman toiminnan kehittämistä tai mahdollisten ulkoisten arvioijien työtä. Kells kuitenkin painottaa, että keskeistä itsearviointinissa on prosessi eikä niinkään tuotettu raportti. (Kells 1995, 28; Hämäläinen & Kauppi 2000.) Itsearviointi voi supistua vain muodolliseksi itsearviointiraportiksi, vaikka se voi olla myös jatkuvaa arviointia, kuten toimintaprosessien näkyväksi tekemistä, palautteiden keräämistä ja erilaisia kehittämishankkeita. Näitä jatkuvan arvioinnin käytännön toimia voidaan hyödyntää itsearviointin kokonaisuudessa. Itsearviointin seuraukset liittyvät siihen, miksi sitä on tehty ja mitä sillä halutaan saavuttaa.

Itsearviointi tuo arvioitavan oikeudet ja arvot esille niin kuin yksikkö haluaa niitä itse painottaa. Kun yksilö, ryhmä tai organisaatio voi itse määritellä omat kehittämiskohteensa, niin toiminnan ongelmat ja edut tulevat esille. Itsearviointi voidaankin määritellä oman toiminnan arvon esiin nostamiseksi. Poikelan ja Poikelan (2006, 142) mukaan arvioinnin kontekstualismi merkitsee huomion kiinnittämistä toimintaprosessiin ja toimijoiden ratkaisuihin prosessin eri vaiheissa, jolloin analyysin perustekijöitä ovat konteksti, prosessit ja tulokset. Itsearviointi-käsitteen kontekstisidonnaisuuden vuoksi onkin selvitettävä, mitä itsearvioinneilla missäkin yhteydessä tarkoitetaan, mitä niillä arvioidaan ja mihin tarkoitukseen tuloksia käytetään.

Itsearviointin kehittämistehtävä

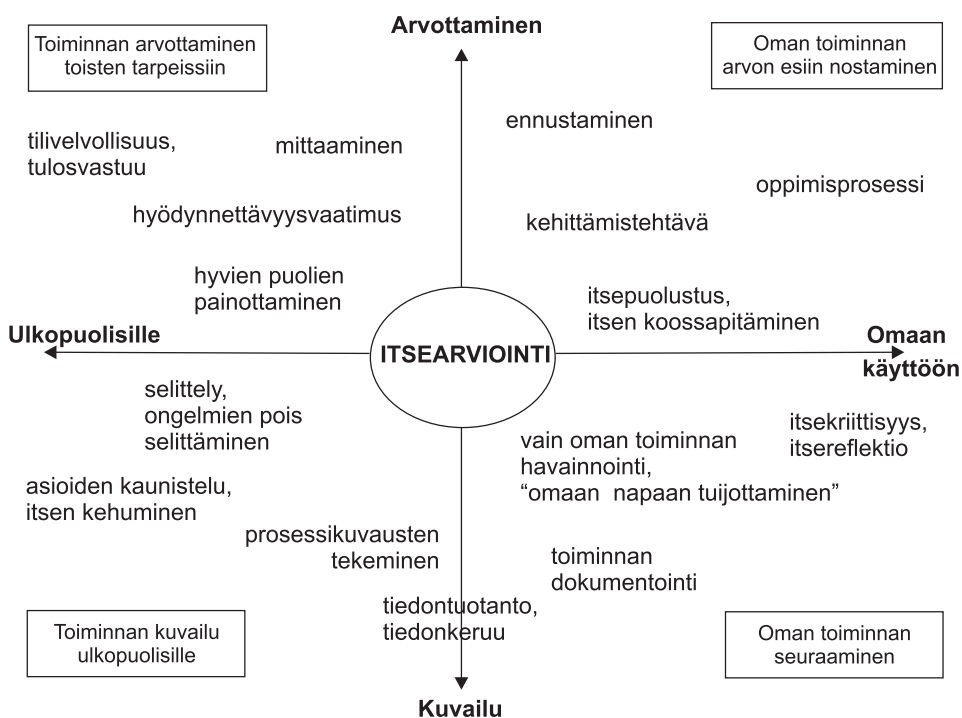
Chelimsky (1997) on tehnyt jaottelun, joka luokittelee arvioinnin tapoja ja tarkoituksia. Chelimsky jakaa arvioinnit kolmeen ryhmään, joita ovat tilivelvollisuusarviointi (evaluation for accountability), kehittämisarviointi (evaluation for development) ja tiedontuotantoarviointi (evaluation for knowledge). (Chelimsky 1997, 10.) Vaikka itsearviointin ajatellaan usein olevan kehittämiseen tähtävää arviointia, sitä voidaan käyttää myös tilivelvollisuus- ja tiedontuotantoarviointien muotona. Chelimskyn mukaan arviointityypin valinnalla on vaikutusta saatavan tiedon luonteeseen. Arvioinnin metodologia vaikuttaa arviointitiedon käyttöön ja hyödynnettävyyteen. Tietoa voidaan tuottaa myös tilivelvollisuus- ja kehittämisarvioinneilla, ja kehittämisarviointien tulosten perusteella voidaan käydä pedagogista tai poliittista keskustelua. (Chelimsky 1997, 22.)

Kehittävä arviointi on usein luonteeltaan formatiivista (Patton 1997, 68). Kehittävän arvioinnin tulos ei ole loppuraportti, vaan ensisijaisia tuloksia ovat oppiminen, kokemus ja oppimisen avulla syntyvä uusi ajattelu (Räisänen 2005, 110). Kehittämissuuntautunut arviointi on laaja-alaista, sillä se kerää tietoa sekä heikkouksista että vahvuuksista. Kehittämissuuntautunut arviointi olettaa, että kerättyä tietoa heikkouksista ja vahvuuksista voidaan käyttää apuna kehittämisessä ja toiminnan reflektoinnissa. (Patton 1997, 68.) Kehittävässä arvioinnissa ei korosteta arviointiin osallistuvien olemista arvioinnin kohteena. Sen sijaan korostetaan arviointikohteena olevien aktiivista osallistumista arvioinnin suunnitteluun ja toteuttamiseen sekä arviointitulosten perusteella tapahtuvaan kehittämistyöhön (Räisänen 2005, 111).

Räisänen (2005) painottaa, että arviointi on kehittävää, jos siihen osallistuvat tahot ja yksilöt oppivat prosessin aikana ja jos arvioinnista on hyötyä myös arvioinnin jälkeen. Arvioinnin vaikuttavuutta voidaan lisätä korostamalla arviointiin osallistuneiden mahdollisuutta saada itseään koskevat tulokset ja niihin liittyvät vertailutulokset, kiinnittämällä huomiota arviointiraporttien selkeyteen ja panostamalla arviointitulosten levittämiseen. Kehittävä arviointi ei Räisänen mukaan pohjautu ulkoiseen kontrolliin eikä siitä seuraa sanktioita. Kehittävän arvioinnin soveltamisen lähtökohta on tuottaa tietoa, jota koulutuksen järjestäjät, oppilaitokset ja opettajat voivat käyttää kehittäessään omaa toimintaansa. Kehittävän arvioinnin tehtävänä on myös välittää arvioinnin avulla esiin tulleita hyviä käytänteitä eri toimijoiden tarpeisiin. Tämän tavoitteen lisäksi kehittävä arviointi tuottaa tietoa myös koulutuspoliittisen päätöksenteon ja laajemman kehittämisen perustaksi. (Räisänen 2005, 115, 118.) Chelimsky (1997, 21) korostaa, että kehittämisarvioinnin tavoitteena on vahvistaa organisaatioita, kun taas tiedontuotantoarviointin tavoitteena on tuottaa tietoa päätöksenteon tueksi.

Itsearviointin muut tehtävät ja refleksiivisyys

Itsearviointinille annetut tehtävät voivat olla myös ristiriidassa keskenään, vaikka kehittämistehtävää usein painotetaan. Itsearvioinneissa sekoittuvatkin erilaiset funktiot ja tavoitteet, kuten tilivelvollisuuden täyttäminen, kehittämisspyrkimykset, itseriittoisuus, oman toiminnan kaunistelu, ulkopuolisten vakuuttelu ja oman erinomaisuuden osoittaminen. Toimintaa voidaan arvottaa tai sitä voidaan kuvailla itsearviointin nimissä. Molempia kutsutaan usein itsearviointiksi, vaikka vain arvottaminen sisältää arvon antamisen tai arvon esiintuomisen. Arvottamisella voidaan tarkoittaa myös arvojen pohtimista tai osallistumista arvojen kehittelyyn. Itsearviointia voidaan tehdä omaan käyttöön tai ulkopuolisille. Näistä neljästä ulottuvuudesta syntyy nelikenttä (ks. kuvio 3), jonka jokaisella osiolla on erityispiirteensä.



Kuvio 3. Itsearviointin ulottuvuuksia

Kuten kuvio 3 käy ilmi, itsearviointinissa korostetaan usein itsekriittisyyttä ja reflektiota, vaikka kaikissa itsearviointin muodoissa ne eivät toteudukaan. Itsearviointiin saattavat vaikuttaa myös ulkopuoliset arvostukset tai ulkopuolelta määritellyt kriteerit, jolloin itsearviointin ajatus toiminnan ja toimijoiden omien arvojen esiin nostamisesta hukkuu. Saari (2002) tähdentää, että itsearviointin

diskurssia kehystää arviointia tuottavan instituution tai ryhmän valta-asema sekä koulutus- ja arviointipoliittinen konteksti. Itsearviointien teksti voi olla konventionaalista, jolloin kehittämisajatus jää vähäiseksi. (Saari 2002, 7.) Vaikka itsearviointia tehdään kehittämisen nimissä, itsearvioinnin muotoon ja toteutus-tapoihin vaikuttaa eniten se, miksi ja kenelle itsearviointia tehdään. Hyvää tar-koittavasta kehittämisfunktioistakin saattaa tulla mantra, joka ei palvele ketään ja joka ei toteudu toiminnassa (ks. esim. Eskola 2006).

Arvioinneissa on usein ongelmana, että itsearvioinnit tuottavat paljon kuva-usta mutta vain vähän arvioimista ja arvottamista. Saaren (2002, 97) mukaan arviointitulosten käyttökelpoisuus kehittämisvälineinä on arviointien ja arvot-tamisten tuottamista, sillä itsekriittinen arviointi on koulutuksen kehittämi-sen kannalta hyödyllisempää kuin hyvän käsityksen tai kuvauksen antaminen toiminnasta (ks. myös Hämäläinen & Kauppi 2000). Arviointi voi muodostua pahimmillaan taitavasti kirjoitetuksi tekstiksi, joka on irrallaan kuvaamastaan toiminnasta ja toiminnan kehittämisestä. Koulutusyksiköt voivat myös palkata arviointitekstien kirjoittajia, jotka saattavat työskennellä organisaation muusta toiminnasta erillään. (Saari 2002, 79.) Arvioinnin toteuttamisessa voivat arvioi-jien omat ja ulkopuolelta tulevat arvot olla ristiriidassa, jolloin itsearvioinnin to-teuttajien pitää reflektoidusti miettiä, mitä he haluavat arvioinnillaan saavuttaa ja miksi arviointi ylipäätään toteutetaan.

Itsearvioinnissa korostetaan yksilön tai yhteisön tulemistä aiempaa tietoisem-maksi omasta toiminnastaan ja toimintansa muutoksista. Oleellinen väline on kriittinen reflektio. (Hämäläinen & Kauppi 2000, 71; Saari 2002, 79.) Itsearvioin-nin reflektiivisyydellä tarkoitetaan yleensä oman toiminnan kriittistä arviointia. Mezirow (1995a, 8; 1995b, 21) määrittelee reflektion omien uskomusten oikeu-tuksen tutkimiseksi ja tunteiden selvittämiseksi. Sen tavoitteena on uudenlai-nen ymmärrys, aikaisemmin tehtyjen arviointien muuttaminen sekä toiminnan suuntaaminen ja ongelmanratkaisuun käytettyjen strategioiden ja menettelyta-pojen toimivuuden uudelleenarviointi. Mezirow erottaa reflektiosta kriittisen reflektion, jolla hän tarkoittaa aikaisemmin opittujen ennako-oletusten kyseen-alaistamista sekä omien ongelmanasettamistapojen ja merkitysperspektiivien arvioimista. (Mezirow 1995a, 8; 1995b, 21, 28–29.)

Saari (2002) painottaa, että jos itsearviointi jää refleктоimattomaksi toimin-naksi, se tuottaa vain pintapuolisen muutoksen, jos sitäkään, eikä siinä nähdä yksikön uusiutumistarpeita. Defensiivinen suhtautuminen käsiteltäessä omia puutteita onkin itsearvioinnissa ongelmallista. Jos arvioinneista ja toiminnas-ta laajemminkin puuttuu reflektointi, toiminnan jatkuvaa kehittämistä tuskin tapahtuu. Jos omiin kehittämishaasteisiin suhtaudutaan defensiivisesti tai niitä ei nähdä, organisaation voimavarat ohjautuvat vallitsevan tilanteen puolustami-seen. Reflektiivinen itsearviointi edellyttää rohkeutta tarkastella oman toimin-nan lähtökohtia, arvoja, edellytyksiä ja muuntautumiskykyä. Saari väittääkin,

että laajojen kuvausten lisäksi itsearviointeja luonnehtivat dialogiton puhe ja yksilön tai ryhmän kannalta edullinen esitys. Kyse on harmoniaan pyrkivästä kirjoitustyylistä, joka vielä ajan myötä voi muuntua niin, että arviointitekstit eivät kerro todellisista asiantiloista vaan taitavasta kirjoitus- tai puhetavasta. Itsearviointi palvelee tällöin muita päämääriä kuin oman toiminnan kehittämisen tavoitetta. (Saari 2002, 77–79, 83.)

Itsearviointi voi uudistaa toimintaa, nostaa esiin omat arvostukset sekä myös muuttaa ulkoisia arvoperusteita. Tämä on mahdollista silloin, jos itsearviointin toteuttajat itse tietävät, mikä ovat parhaita välineitä ja tapoja heidän toimintansa kuvailuun ja kehittämiseen sekä ovat avoimia toiminnan kehittämiseksi ja uusille ideoille.

Kehittävästä arvioinnista voimaantumiseen

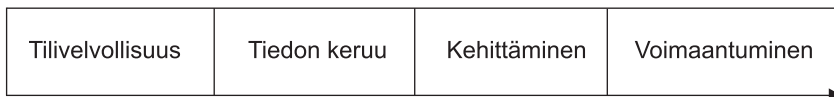
Kehittävän arvioinnin katsotaan kuuluvan niihin arviointeihin, joissa yksilöiden ja yhteisöjen voimaantuminen mahdollistuu. Fetterman (1997, 382) määrittelee voimaannuttavan arvioinnin arviointikäsitteiden ja -tekniikoiden sekä arviointitiedon käytöksi, jonka avulla voidaan edistää ja vahvistaa yksilön oikeutta päättää omista asioistaan. Voimaannuttavalla arvioinnilla halutaan korostaa vahvaa sitoutumista arviointiin, reflektioon perustuvaa oppimista, eri tahojen yhteistyötä, arviointiin osallistuvien oikeutta osallistua arviointia koskevaan päätöksentekoon ja arvioinnin toteuttamiseen sekä mahdollisuutta ulkopuolisten asiantuntijoiden apuun (Räisänen 2005, 111–112). Voimaannuttava arviointi hyödyntää niin laadullisia kuin määrällisiä menetelmiä. Sen tarkoituksena on auttaa ihmisiä auttamaan itse itseään. Voimaannuttavaa arviointia käytetään usein yhteistoiminnallisesti ryhmissä, jolloin arvioinnin päämääränä on olla osa jatkuvaa kehittämistä. Osallistujat oppivat arvioimaan omaa toimintaansa itse määrittelemillään tavoitteilla sekä muokkaamaan suunnitelmiaan ja tavoitteitaan arviointien pohjalta. (Fetterman 1997, 382–383, 386.)

Voimaannuttavassa arvioinnissa korostuvat arvioinnin ennustamattomat ulottuvuudet (Guba & Lincoln 1989): etukäteen ei voida tietää, miten prosessi etenee ja miten osallistujat ovat kokeneet voimaantuneensa prosessissa. Myös prosessi itsessään luo todellisuutta ja vaihtoehtoisia kehittämissuuntia. Arvioinnin keskiössä on osallistujien oma oppimisprosessi ja valmiudet oppia kehittämään omaa toimintaansa. Itsearviointissa ei ole ulkopuolista arvioijaa, vaan kohde arvioi itse itseään, ideaalitapauksessa myös omista lähtökohdistaan ja omin arvoperusteiden tuoden ne ilmi muille yhteisön jäsenille. Itsearviointi voi samalla voimaannuttaa tekijöitään.

Itsearviointin tarkastelu kehittämisen ja voimaantumisen näkökulmista tulee lähelle Habermasin tiedonintressejä. Habermasin tiedonintressit ovat tekninen,

praktinen ja emansipatorinen (Habermas 1977). Tekninen tiedonintressi hakee tietoa luonnon ja muun ympäristön ilmiöiden selittämistä ja kontrolloimista varten. Praktinen tiedonintressi pyrkii ymmärtävään tietoon ihmisestä tai jostain kulttuuri-ilmiöstä. Emansipatorinen tiedonintressi pyrkii tiedostamiseen, joka voi tuoda toimintaan vapautumisen kokemuksen. (Ks. esim. Leino & Leino 1991, 71.)

Habermasin tiedonintressejä voidaan muokata ja soveltaa myös arviointiin (ks. kuvio 4). Myös arviointi voi olla teknistä, praktista tai emansipatorista eli voimaannuttavaa. Teknisistä tilivelvollisuus- ja tiedonkeruuarvioinneista voidaan päästä praktiseen arviointiin ja kehittämiseen. Kehittäminen voi jalostua voimaannuttavaksi arvioinniksi, jolloin arvioija kokee uudistuvansa ja auttavansa itse itseään eli omaa organisaatiotaan. Keskeistä arvioinnissa ei olekaan enää saadut tulokset tai kerätty tieto, vaan toiminnan kehittäminen ja voimaantumisen kokemus.



Kuvio 4. *Tilivelvollisuusarvioinnista voimaannuttavaan arviointiin*

Voimaannuttava arviointi vahvistaa yksilön ja yhteisön omia arvolähtökohtia sekä auttaa kehittämään omaa arviointiajattelua omista lähtökohdistaan. Erityisesti itsearviointeissa korostuu voimaantumisen mahdollisuus. Yksilöt ja yhteisöt kokevat saavansa itsearviointista jotain sellaista, mitä ei muuten syntyisi toiminnasta ja toiminnan lomassa. Kun yksilöt ja yhteisöt pääsevät itse määrittelemään itseään, kehittämisalueitaan ja vahvuuksiaan, voi syntyä vahva sitoutumisen ja voimaantumisen kokemus. Tällöin yksilöillä ja yhteisöillä on myös voimaa kehittää toimintaansa jatkossakin.

Lopuksi

Itsearviointi voi kehittää yhteisöä tai yksikköä, jos sen avulla aidosti reflektoidaan toimintaa ja yhteisössä on halua kehittyä. Itsearvioinnissa on oleellista mieltä, kuka itse oikeastaan on ja miten itse missäkin arvioinnissa rajataan. Toisinaan itsearviointiin kuuluvat kiinteästi läheisimmät sidosryhmät, toisinaan sidosryhmät katsotaan ulkopuolisiksi arvioijiksi. Koulujen itsearviointeissa opettajat ja oppilaat voivat muodostaa arvioinnin kohteen ja toimijat, toisinaan itsearviointiin otetaan mukaan myös vanhemmat, kunnan ja työelämän edustajat.

Itsearviointi voi toteutua eri muodoissa. Oppijat voivat arvioida omaa oppimistaan tai saamaansa opetusta tai opetuskokonaisuuksia, opettajat voivat arvioida omaa tai työtovereidensa opetusta tai työyhteisöä. Työyhteisö voi arvioida omaa toimintaansa kokonaisuutena eritellen sen eri osa-alueita, kuten opetussuunnitelmaa, henkilöstön kehittämistä ja johtamista. Rehtori voi arvioida koulun toimintaa yhdessä opettajien kanssa tai arvioida muiden rehtoreiden kanssa laajemmin kunnan opetustoimea. Kaikki nämä eri toiminnot ovat itsearviointia eri painotuksin.

Itsearviointia voidaan käyttää myös kansallisten arviointien aineistoina, mutta tällöin pitää kiinnittää erityistä huomiota itsearviointeja määritteleviin ohjeistuksiin ja annettuihin sisältöalueisiin. Mahdollisimman vertailtavien tietojen saamiseksi ohjeistusten pitäisi olla mahdollisimman tarkat, vaikka tällöin yksiköiden omat arvot ja kehittämiskohteet saattavat jäädä huomiotta. Jos ohjeistukset tulevat yksikön ulkopuolelta, voidaan myös miettiä itsearviointia itseä: onko se enää paikallisen tason toimija vai onko itsenä eli varsinaisena toimijana ohjeiden antaja? Jos itsearviointia tuotettuja aineistoja käytetään osana laajempia arviointeja, on huomioitava, että itselle ja toisille kirjoitetut itsearviointit ovat tehty eri tarkoituksia varten ja ovat siten erilaisia. Ulkopuolisille tehdyissä itsearvioinneissa saattaa oman toiminnan reflektointi muuntua ulkopuolisten vaakuitteluksi oman toiminnan erinomaisuudesta.

Yleistysten tekeminen itsearviointien pohjalta voi olla vaarallista tai ainakin harhaanjohtavaa. Itsearviointia kehittävän arvioinnin menetelmänä ei ole tarkoitettu yleistyksiä varten, vaan niiden pääasiallinen tehtävä on paikallistason toiminnan kehittäminen esiintuotujen vahvuuksien ja kehittämisehdotusten avulla. Kehittävän arvioinnin ajatus katoaa tai ainakin sen rooli pienenee, jos tarkoituksena on tehdä esimerkiksi koko maan koululaitoksia koskevia yleistyksiä.

Itsearviointien validiutta voidaan myös tarkastella: kertooko se toiminnasta olennaista ja voivatko muut yhtyä tehtyihin johtopäätöksiin ja kehittämisehdotuksiin. Voidaan siis miettiä, miten paikkaansa pitävää itsearviointia tuottama tieto on. Jos yksikkö tai organisaatio tekee itsearviointia oman toimintansa kehittämiseksi, heille on totta se, minkä he kokevat todeksi. Jos yhteisössä koetaan tarvetta jonkin tietyn asian tai toiminnan esiin nostamiseksi tai kehittämiseksi, niin kysymys itsearviointia paikkaansa pitävyydestä on väärin asetettu. Validiteetin käsite sopiikin huonosti kehittävän arvioinnin näkökulmaan, sillä se on otettu määrällisen arvioinnin paradigmasta.

Itsearviointi on välttämätöntä kehittävän arvioinnin luotettavuuden kannalta, koska subjektilla ajatellaan olevan itsellään paras käsitys arvioitavasta toiminnasta ja siihen vaikuttavista tekijöistä. Itsearviointia voidaan parhaimmillaan läpivalaista toimintaympäristöä ja siinä olevaa hiljaista tietoa sekä nostaa ne arvioinnin keskiöön. Yksilöiden ja yhteisön tekemän refleksiivisen tarkastelun avulla voidaan myös tuottaa uutta tietoa ja kehittää uusia toimintamalleja.

Arviointitoiminnan laajentuessa myös arvioinnin haittavaikutukset ovat usein ilmeisiä, kuten arviointiväsymys, resurssien kuluminen ja hallinnoimisen tunteen lisääntyminen. Tästä huolimatta itsearviointi, kun se toteutetaan paikallisesti, oppimista korostavasti ja refleктоivasti, voi auttaa siihen osallistuneita kehittämään omaa työtään ja työyhteisöään. Kun yhteisö ja siihen kuuluvat yksilöt voivat itse määrittellä oman toimintansa arvoa ja arvostuksia, yhteisö voimaantuu. Kun itsearviointille resursoidaan tarpeeksi aikaa ja tehtyä työtä arvostetaan myös johdon ja ylläpitäjän tasolla, voimaantumisen seurauksena työyhteisö voi saada tarmoa huomata työyhteisönsä ongelmakohtat ja kehittää niitä toimivimmiksi. Voimaantunut työyhteisö jaksaa puhalttaa yhteen hiileen. Arviointi hyvin toteutettuna onkin yhteispeliä, jossa on vain voittajia.

Lähteet

- Ball, S. 2001. Globaalit toimintaperiaatteet ja kansalliset politiikat eurooppalaisessa koulutuksessa. Teoksessa A. Jauhiainen, R. Rinne & J. Tähtinen (toim.) Koulutuspolitiikka Suomessa ja ylikansalliset mallit. Suomen kasvatustieteellinen seura. Kasvatusalan tutkimuksia 1, 21–43.
- Chelimsky, E. 1997. The coming transformations in evaluation. Teoksessa E. Chelimsky & W. R. Shadish (toim.) Evaluation for the 21st century. A Handbook. Thousand Oaks: Sage, 1–26.
- Eskola, J. 2006. Pimp my ride. Pääkirjoitus. Kasvatus 37 (3), 217–221.
- Fetterman, D. M. 1997. Empowerment evaluation and accreditation in higher education. Teoksessa E. Chelimsky & W. R. Shadish (toim.) Evaluation for the 21st Century. A Handbook. Thousand Oaks: Sage, 381–395.
- Guba, E. G. & Lincoln, Y. S. 1989. Fourth generation evaluation. Newbury Park: Sage.
- Habermas, J. 1977. Theory and practice. London: Heinemann.
- Harvey, L. 2002. The end of quality? Quality in Higher Education 8 (1), 5–22.
- Hämäläinen, K. & Kauppi, A. 2000. Evaluation in today's education. From control to empowerment. Lifelong Learning in Europe V (2), 68–75.
- Jakku-Sihvonen, R. & Heinonen, S. 2001. Paikallistasolla tapahtuva koulutuksen arviointi. Teoksessa R. Jakku-Sihvonen & S. Heinonen, Johdatus koulutuksen uudistuvaan arviointikulttuuriin. Opetushallitus. Arviointi 2. Helsinki: Yliopistopaino, 64–77.
- Karjalainen, A. 2005. Pedagoginen muutos yliopisto-organisaatiossa – idealismia vaiko realismia? Teoksessa R. Jakku-Sihvonen (toim.) Uudenlaisia maistereita. Kasvatusalan koulutuksen kehittämislinjoja. Opetus 2000. Jyväskylä: PS-kustannus, 43–54.
- Kells, H. R. 1995. Self-study processes. A Guide to self-evaluation in higher education. American Council on Education. Series on Higher Education. Fourth Edition. Phoenix: Oryx Press.
- Kivirauma, J. 2001. Kansainvälistymisen pakot: kohti eriarvoistuvaa koulutuspolitiikkaa. Teoksessa A. Jauhiainen, R. Rinne & J. Tähtinen (toim.) Koulutuspolitiikka Suomessa ja ylikansalliset mallit. Suomen kasvatustieteellinen seura. Kasvatusalan tutkimuksia 1, 73–90.
- Konttinen, R. 1995. Arvostelusta näyttöön. Koulutuksen arvioinnin kehityspiirteitä Suomessa. Kasvatus 26 (1), 6–14.
- Leino, A.-L. & Leino, J. 1991. Kasvatustieteen perusteet. Helsinki: Kirjayhtymä.
- Mezirow, J. 1995a. Esipuhe. Teoksessa J. Mezirow. Uudistava oppiminen. Kriittinen reflektio aikuiskoulutuksessa. Helsingin yliopiston Lahden tutkimus- ja koulutuskeskus, 5–13.

- Mezirow, J. 1995b. Kriittinen reflektio uudistavan oppimisen käynnistäjänä. Teoksessa J. Mezirow. Uudistava oppiminen. Kriittinen reflektio aikuiskoulutuksessa. Helsingin yliopiston Lahden tutkimus- ja koulutuskeskus, 17–37.
- Moitus, S. & Saari, S. 2004. Menetelmistä kehittämiseen. Korkeakoulujen arviointineuvoston arviointimenetelmät vuosina 1996–2003. Korkeakoulujen arviointineuvoston julkaisuja 10. Helsinki.
- Neave, G. 1988. On the cultivation of quality, efficiency and enterprise: An overview of quality assurance in higher education in Western Europe 1986–1988. *European Journal of Education* 23 (1–2), 7–23.
- Patton, M. Q. 1997. *Utilization-focused evaluation*. The new century text. 3rd edition. Thousand Oaks: Sage.
- Poikela, E. & Poikela, S. 2006. Arviointi ongelmaperustaisen pedagogiikan kontekstissa. Teoreettisia lähtökohtia ja opettajien kokemuksia. Teoksessa A. R. Nummenmaa & J. Välijärvi (toim.) *Opettajan työ ja oppiminen*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 139–153.
- Powell, L. A. 2000. Realising the value of self-assessment: The influence of the Business Excellence Model on teacher professionalism. *European Journal of Teacher Education* 23 (1), 37–48.
- Power, M. 1999. *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Räisänen, A. 2005. Kehittävään arviointiin. Teoksessa H. K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnista*. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä, 109–127.
- Saari, S. 2002. Opettajankoulutuksen arviointi- ja kehittämisdiskurssi koulutuspolitiikassa kontekstissa. Tampereen yliopisto. *Acta universitatis tamperensis* 893.
- Saarinen, T. 1995. Nousukaudesta lamaan, määrästä laatuun. Korkeakouluarvioinnin käynnistyminen ja kokemukset laitoksilla. Turun yliopisto. Koulutussosiologian tutkimuskeskus. Raportti 32.
- Scriven, M. 1991. *Evaluation thesaurus*. Fourth edition. Newbury Park: Sage Publications.



Utvärdering i utvecklande syfte på nationell nivå

I denna artikel granskas vad det finländska ställningstagandet för utvärdering i utvecklande syfte innebär för organiserandet av utvärderingar på det praktiska planet. Inledningsvis diskuteras vad utvärdering innebär i utvecklingsprocessen under olika förhållanden. Effekterna av utvärderingar varierar med kontext och konsekvenserna är inte alltid de avsedda. En medveten utvecklingsorienterad inriktning i olika skeden av utvärderingsprocessen behövs. Då siktet är inställt på utveckling, blir vissa frågeställningar och vissa sidor av ett utvärderingsobjekt viktigare än andra. De olika tillvägagångssätt och modeller för utvärdering som har utarbetats lämpar sig i varierande grad, då man eftersträvar utveckling. Ytterligare tangeras frågan om vilken ställning i och nytta av utvärderingen de olika parter som berörs av utvecklingsarbetet har i olika sammanhang.

Nyckelord: utbildningsutvärdering, utvecklingsarbete, nationell utvärdering, utvärderingstyper

Rådet för utbildningsutvärdering, som grundades 2003 och som bland annat har som uppgift att stödja studerandenas inläring, undervisningspersonalens arbete och utvecklingen av läroanstalterna (Statsrådets förordning om utvärdering av utbildning 20.2.2003 /150), har i sin verksamhetsstrategi gått in för en utvecklande utvärdering (Utbildningsutvärderingens nya inriktning 2004) i de

utvärderingsprojekt som rådet organiserar. Därmed följer Rådet för utbildningsutvärdering de riktlinjer som tidigare dragits upp och tillämpats av Rådet för högskolornas utvärdering. Utvärdering har traditionellt kopplats samman med kontrollfunktioner och går naturligtvis aldrig helt fri från dem. Även om man utvärderar i syfte att utveckla något, ingår en fas av kontroll av var man står. Utvärdering i kontrollsyfte med befogenheter att sanktionera är dock något helt annat.

Utvärdering i akt och mening att utveckla verksamheten löper dock risk att i praktiken utnyttjas för kontrollsyften. Medierna t.ex. använder utvärderingar i kontrollsyfte, om de rangordnar skolor på basen av utvärderingsresultat som utbildningsanordnaren ursprungligen låtit sammanställa för att kunna inrikta sina stödåtgärder rätt. Utgående från utvärderingsresultaten stämplas skolorna då som goda och dåliga i den stora allmänhetens ögon, vilket kan motverka anordnarens syften. För att utvärderingarna skall kunna tjäna som redskap i utvecklingsarbetet kräver utvärderingsverksamheten klara etiska principer av samma art som forskningen tillämpar, samt medvetenhet om effekterna av olika tillvägagångssätt. I denna artikel granskas vad det finländska ställningstagandet för utvärdering i utvecklande syfte innebär för organiserandet av utvärderingar på det praktiska planet.

Vad innebär utvärdering och vad menas med utvecklande utvärdering?

Vad avses då egentligen med utvärdering och hur skiljer sig utvärdering från forskning? Utvärdering är en verksamhet som inbegriper att ge kunskap om något objekt och att bestämma i vilken mån det är värdefullt (Åberg 1997, 12). Utvärdering innebär alltså att man för en viss funktion producerar kunskap som man sedan underställer en bedömning. I grunden är det fråga om att ställa en observerad verklighet mot något ideal (Åberg 1997, 14). Kunskapsproduktionen kännetecknas i princip av samma krav på tillförlitlighet som i kvantitativ eller kvalitativ forskning. Medan forskningen nöjer sig med en tolkande beskrivning av verkligheten, ofta med inriktning på orsak och verkan, syftar utvärderingen däremot till att med utgångspunkt i vissa värden bedöma hur lyckad verkligheten är och identifiera var bristerna finns. I denna bedömningsprocess ställs empirin mot mål, intressen, teori (som antingen är etablerad och generell eller också specifik och lokal) eller andra explicita ideal (Åberg 1997, 95). En annan skillnad är att utvärdering sker inom en "uppdragsram", som ofta är snävare formulerad än vid forskning, som sker inom en "teoriram" formulerad av forskaren (Gunnarsson 1987, 118; se också Karlsson 1999, 19).

Med utvecklande utvärdering avses sådan utvärdering som har som mål att stöda organisationer och kollektiv, såsom läroanstalter, i utvecklingen av deras verksamhet. Enligt en tolkning av Rådet för utvärdering av högskolorna är utvecklande utvärdering en process som utgår från användaren och i den processen skraddarsys förfarandena med hänsyn till utvärderingens mål, temat som skall utvärderas och deltagarnas behov (Auditering av högskolornas kvalitetssäkringssystem 2005, 33). Utvecklande utvärdering kan alltså ta olika former. Rådet för utbildningsutvärdering har tillämpat en form av utvecklande utvärdering, men andra former är möjliga. I princip är syftet med utvecklande utvärdering på nationell nivå att ge aktörerna på fältet stöd och hjälp i deras eget utvecklingsarbete. Syftet är inte att ange utvecklingsnormer och tvinga fältet till utveckling. På såväl europeisk som nationell nivå finns det dock ett visst intresse också för styrning med hjälp av utvärdering.

Vad innebär utvärdering i utvecklingsprocessen under olika förhållanden?

I ett dynamiskt samhälle ingår utveckling som ideal i all verksamhet. Även om en verksamhet skulle uppfylla de mål som ställts upp för den, så medför förändringar i kontexten krav på förändringar i verksamheten för att de goda resultaten skall bestå. Också nya mål ställs upp för en verksamhet, då behoven förändras. En utvärdering i syfte att utveckla kräver därför förutom en målrelaterad bedömning av det objekt man fokuserar också att man beaktar behov och kontext samt förändringar i dessa.

Utvärderingen får olika roll och ställning, dvs. olika funktion, i organisationerna beroende på hur verksamheten i huvudsak styrs inom organisationerna. Den styrning som riktas/har riktats till anordnarna av utbildning kan indelas i målstyrning/insatsstyrning, marknadsstyrning och resultatstyrning (se t.ex. Tiisonen 2006). I organisationer med resultatstyrning innebär utveckling i praktiken främst en strävan till ökad effektivitet och lönsamhet. Effektivitet och lönsamhet uppnår man med ökad frekvens och ökad kvantitet av olika slag. I organisationer med resultatstyrning ingår utvärderingen som en central komponent i verksamheten. Uppnådda resultat av vikt vid resurstilldelningen dokumenteras fortlöpande. I organisationer med insatsstyrning satsar man inte lika stort på utvärdering. Utvärderingen kan minimeras till kontroll av att verksamhetens inriktning överensstämmer med de mål som uppifrån definierats för organisationen. Ett naturligt kriterium för utvärderingen är då ändamålsenlighet. Vid marknadsstyrning behövs också fortlöpande utvärdering. Det gäller att hålla sig ajour med marknadsbehoven, för att organisationen skall kunna svara mot dem.

Det som då utvärderas är behoven, behovsändringar och marknadens tillfredsställelse med organisationens verksamhet.

Den utbildningspolitiska kontexten på såväl nationell som regional och lokal nivå är likaså av betydelse för den funktion utvärderingen slutligen får i utvecklingsarbetet. Funktionen behöver ju inte vara densamma som det avsedda eller deklarerade syftet (Karlsson 1999, 31). Utvärderingen får t.ex. en annorlunda funktion i en miljö med konkurrensbetonat klimat jämfört med en miljö där gemensamma intressen betonas. I ett konkurrensbetonat klimat accepteras ständig utvärdering av resultaten (både kollektiva och individuella), men i genomsnitt fungerar utvärderingen sannolikt inte kraftingivande (se t.ex. Lyytinen & Nikkanen 2005 om kraftingivande utvärdering "empowerment"). I en miljö där gemenskap och gemensamma intressen värdesätts faller det sig mer naturligt att utvärdera ändamålsenligheten och göra det med avseende på hur verksamhetens olika delar fungerar i helhetsperspektiv. Ett konkurrensbetonat klimat förorsakar också lätt att verksamheten börjar styras av utvärderingen och de resultat som premieras i stället för av utbildningens egen interna logik utifrån utbildningens uppgift (se Moitus & Saari 2004, 74). Faran är att utvärderingen blir husbonde och utvecklingen blir dräng.

Likaså får utvärderingen och utvärderingsresultaten olika betydelse i olika typer av organisationer. I vissa organisationer utvecklar man aktivt verksamheten vartefter nya riktlinjer kan skönjas och tar gärna emot respons utifrån. I andra anpassar och integrerar man nya idéer i den mån de passar in i den egna verksamheten. I en tredje typ av organisationer ställer man sig kritiskt avaktande till nya uppgifter och till respons (jfr. Vaherva et al. 2007, 58). Dessa tre organisationstyper kunde benämnas prestationsinriktad utifrånstyrd organisationskultur, uppgiftsorienterad organisationskultur respektive självdefensiv organisationskultur.

Moitus & Saari (2004, 74) vid Rådet för högskolornas utvärdering frågar sig om utvärderingarna har fått en för stark ställning i utbildningen. De har noterat en tydlig utvärderingströtthet på högskolefältet och konstaterar att utvärderingarna inte har fungerat på avsett sätt, dvs. kraftingivande på organisationerna och personalen. Alternativt antar de, att en fortgående utvärderingsprocess i grund och botten är främmande för den vardagliga verksamheten. Också Kauppi (2005, 10) konstaterar att utvärderingen många gånger uppfattas som "extra" arbete, som egentligen är mera till skada än till nytta med tanke på undervisning och studier och utvecklandet av dem. Ur utvecklingsarbetets synvinkel utgör utvärderingen en begränsad fas, där förhandling om och (om)definiering av målen, planering och genomförande av en verksamhet tidsmässigt utgör merparten av arbetet. Den kritiska frågan blir då, hur många och omfattande utvärderingar av olika aspekter av utbildningen som det är till nytta att genomföra under en viss tid och hur stor del av verksamheten som behöver dokumenteras.

Nationella s.k. externa utvärderingar är problematiska genom att de är svåra att koppla till utbildningsorganisationernas egen utvecklingsrytm. Om utvärderingen åtskiljs från övrig verksamhet, är det svårt att nyttiggöra den vid utveckling av verksamheten (Kauppi 2005, 131). T.ex. frekventa eller omfattande externa utvärderingar av delar av utbildningen kan på grund av sina resurskrav till och med bli till skada för små anordnare med små resurser. För att nationella utvärderingar skall främja utvecklingen också på lokal nivå, måste de lokala förhållandena beaktas i planeringsskedet och utvärderingarna utformas enligt dem. Det behövs metodologiska koncept till stöd för alla typer av läroanstalter och för alla parter i utvärderingsprocessen. I en uppföljning av utbildningsinspektionernas effekter på kommunal nivå i Sverige framkom det t.ex. att lärarna upplever minst nytta av inspektionerna och administratörerna mest (Sandahl & Bringle 2006).

Huruvida resultaten av externa nationella utvärderingar används ens i det nationella utvecklingsarbetet beror på hur politiskt lämpliga de är (Moitus & Saari 2004, 78). Om resultaten står i konflikt med aktuell utbildningspolitik, nonchaleras de lätt. Det faktum att utvärderingar görs men inte används i beslutsprocessen, åtminstone inte på det sätt som man föreställt sig, har lett till att man frångått synen på utvärdering som ett tekniskt och neutralt instrument, som fungerar oberoende av de politiska processerna (Åberg 1997, 37). Utvärdering av offentlig verksamhet befinner sig i ett politiskt laddat spänningsfält, vilket gör att utvärdering som verksamhet inte kan betraktas fristående från politik (Karlsson 1999, 69).

Vad är värt att utvärdera i utvecklingssyfte och vilka utvärderande frågor stöder utveckling?

Om man ser på utvärderingsverksamheten ur historisk synvinkel, från 1950-talet fram till nuvarande sekel, kan man urskilja tre olika fokuseringspunkter för utvärdering, som Karlsson (1999, 17) kallar 1) mål-resultatutvärdering, 2) processutvärdering och 3) interaktiv utvärdering. Vid mål-resultatutvärdering mäter utvärderaren utfallet av en verksamhet utan intresse för hur resultatet uppnåtts. Vid processutvärdering görs en kvalitativ bedömning av hur verksamheten fungerar. Om man utvärderar i syfte att stöda utvecklandet av en verksamhet är det givet att både process och resultat bör fokuseras. Objektet för utvärderingen bör enligt Vedung (1998) analyseras på fyra områden: de insatser som har gjorts, genomförandet, resultaten (i relation till mål) och andra effekter. Vid interaktiv utvärdering, som är en nyare variant, inriktas intresset på det som olika intressegrupper ser som viktigt att utvärdera utifrån sina perspektiv. Orsakerna till att man alltmer betonar behovet av att olika intressegrupper deltar i definitionen av

utvärderingsobjekten är två (Karlsson, 1999,17). En orsak är att man på det sättet vill öka utvärderingarnas relevans och identifiera frågor som stöder utvecklingsarbetet på olika plan. En annan orsak är att man vill försäkra sig om att också svaga intressegruppers intressen tillvaratas

En grundläggande fråga i planeringen av utvärdering är vilka slag av utvärderingsresultat som eftersträvas. Utifrån typen av åsyftade resultat kan man utkristallisera tre typer av utvecklande utvärdering:

- Kvalitetsorienterad utvärdering, där den centrala frågan ur åtgärdssynvinkel är: Bör något göras?
- Problemorienterad utvärdering, där den centrala frågan ur åtgärdssynvinkel är: Vad bör göras?
- Prognostisk utvärdering, där den centrala frågan ur åtgärdssynvinkel är: Vad händer om man gör något visst?

Ett syfte med utvecklande utvärdering är alltså att samla in fakta som kan tjäna som beslutsunderlag för styrning av utvecklingen. Den vanligaste formen av utvecklande utvärdering är *kvalitetsorienterad*. Syftet är att granska om verksamheten och/eller resultaten är tillfredsställande eller om det finns anledning att skrida till åtgärder. I kvalitetsorienterad utvärdering kan verksamheten eller resultaten bedömas genom jämförelse med olika referenspunkter:

- med målet eller normen för verksamheten, om man är intresserad av att utvärdera "absolut" läge
- med andra kontrahenter eller objekt, om man är intresserad av att utvärdera relativt läge
- med utgångsläget, om man är intresserad av att utvärdera förändring/utveckling

Utvärderingarna av inlärningsresultaten är av denna typ. Hit kan också hänföras de utvärderingar av verksamheten som utnyttjar olika verksamhetsideal som jämförelsepunkt (EFQM, CAF och andra modeller för analys av en verksamhets kvalitet, vilka till största delen har en rationell, hierarkisk organisationsmodell som grund och utvärderar hur väl en organisation överensstämmer med detta ideal).

I de utvärderingar av utvecklande typ som utförts av Rådet för utbildningsutvärdering har kvalitetsorienteringen kompletterats med en form av *problemorientering*. Samtidigt som man bedömt kvaliteten har man identifierat problemområden genom att definiera kritiska punkter, starka sidor och förbättringsområden. Utvärderarens uppgift är då att söka lösningar på problemen och ge förslag till åtgärder som antas främja utvecklingen. Inom utvärderingsforskningen har man

olika uppfattningar om huruvida rekommendationer hör till utvärderingen. En del anser att externa utvärderare endast bör ha befogenheter att avgöra vilka områden som kräver utvecklingsinsatser, dvs. att definiera problemen, och att det hör till de parter som har ansvar för verksamheten att dryfta åtgärder, dvs. lösa problemen. Problemorienterad utvecklande utvärdering innebär dock att man också gör försök att finna lösningar på de problem som identifieras eller på de problem som eventuellt utgjorde utgångspunkt för utvärderingen. Renodlad problemorienterad utvärdering kräver fokus på orsakerna till olika resultat och på förhållandena under vilka olika resultat uppstår.

Prognostisk utvärdering fokuserar i sin tur på följderna av olika åtgärder. Innan man fattar beslut om tillvägagångssätt för utveckling av verksamheten, kan det ibland vara skäl att i förväg bedöma följderna av alternativa åtgärder. Denna typ av utvärdering är sällsynt. Den prognostiska utvärderingens metodiska hemvist finns att söka i framtidsforskningens paradig. Det gäller att mångsidigt och mot bakgrund av teoretisk kunskap och empiriska analyser bedöma konsekvenserna av ett flertal alternativa åtgärder. Resultatet av denna utvärdering blir utvecklingsförslag som prioriteras med hänsyn till vissa kontextuella villkor.

Beroende på utvärderingens allmänna syfte kan man välja en av dessa tre utvärderingstyper, men en grundlig utvärdering i syfte att stöda utvecklingsarbetet kan i princip omfatta alla tre frågeställningar. De kan då ses som delar i en utvärderingssekvens.

Vad krävs av utvärderingsdesignen om man siktar på utveckling?

När utvärderingens syfte är att stöda utvecklingsarbetet kan man antingen försöka stöda alla nivåer inom utbildningen i samma utsträckning eller satsa på stöd för en enskild nivå. Konceptet utvecklande utvärdering så som det tillämpats inom Rådet för utbildningsutvärdering har gått ut på att aktivera de parter vars verksamhet utvärderas, dvs. utbildningsanordnarna och undervisningspersonalen, samt att samla information för att kunna identifiera problem och komma med utvecklingsförslag dels åt dem som varit föremål för utvärderingen och dels åt de styrande parterna inom utbildningen, främst undervisningsministeriet, Utbildningsstyrelsen och lärarutbildningsenheterna.

De nationella utvärderingarna kan aktivera utbildningsanordnarna och andra berörda genom sin blotta existens. Utvärderingarna innebär information om vad som på nationell nivå anses viktigt att utveckla. Enbart det att man ställer frågor till en part kan fungera som impuls till utvecklingsarbete. Frågornas utformning är avgörande för vilket stöd för vidareutveckling av fenomenet som besvarandet av frågorna ger. Frågorna kan stöda insikten om att något bör göras, de kan ge

riktlinjer för lösning av problem eller de kan ge idéer till olika åtgärder som kunde vidtas. Utvärdering i utvecklande syfte kan vara mer eller mindre styrande. Syftet kan vara att styra, men också oberoende av syfte kan utvärderingen fungera mer eller mindre styrande.

Ett effektivare sätt att aktivera är att be anordnarna genomföra en självutvärdering i anslutning till det fenomen som utvärderas för att på så sätt i bästa fall sätta igång en utvecklingsprocess på lokal nivå (se t.ex. Mäensivu et al. 2007). Självutvärderingarna kan sedan användas som material för slutledningar om läget på nationell nivå och om nationella utvecklingsbehov. Det ligger dock en konflikt i att styra självutvärdering av enskilda fenomen utifrån, eftersom självutvärdering i första hand är ett redskap för intern utveckling utifrån egna behov.

För att stimulera till utveckling på det nationella planet krävs det att utvärderingen och de resultat man kommit fram till förankras i den rådande utbildningspolitiska kontexten. Den kvalitetsorienterade utvärdering som tillämpar en utprövande och mätande strategi, och som endast uttalar sig om huruvida något bör göras, lämnar på sätt och vis de nationella utvecklingsinstanserna i sticket. Det krävs åtminstone en upplysande och förklaringsinriktad strategi för att ge en mera handfast väggkost i utvecklingsarbetet, om inte en interaktiv, samarbets- och förändringsinriktad strategi (för indelningen av strategierna se Shadish et al. 1991, 472).

Till den kvalitetsorienterade utvärderingen, där man utprövar och mäter kvaliteten, hör också identifiering av goda lösningar, god praxis. Dessa goda lösningar har dock i praktiken visat sig vara svåra att utnyttja som utvecklingsredskap och svåra att överföra som sådana till andra kontexter. Kritikerna framhåller att sociala förändringsprocesser inte är så lättstyrda att man automatiskt skulle använda modelllösningar (Karlsson 1999, 54).

Den upplysande och förklaringsinriktade strategin går ut på att utforska processer, hur de går till, hur de kan utvecklas och vad de betyder för de inblandade. Den centrala frågan är varför olika lösningar fungerar i vissa sammanhang men inte i andra. Målet är en nyanserad bild av verkligheten och information om vilka slags lösningar som verkar fungera i vilka kontexter. För utveckling på nationell nivå, där variationen mellan olika organisationer och miljöer kan vara stor, ger denna strategi beslutsfattarna ett betydligt bättre utgångsläge. Strategin är dock mer krävande i och med att den kräver en mer gedigen teoretisk förankring. Faran med denna strategi är att man uppnår generell kunskap som inte ger tillräckligt konkret vägledning (Karlsson 1999, 54).

När man siktar på utveckling, ger allmän information om fenomenet inte mycket i väggkost, även om det i vissa fall kan vara helt tillräckligt. Det krävs i allmänhet att man når fram till en djupare förståelse av fenomenet i hela dess bredd, vilket också innefattar en kritisk granskning av utvärderingsobjektet och tom. utvärderingsuppdraget. En utvärdering som parallellt bearbetar olika perspektiv

på fenomenet kan ses som den högststående formen av utvärdering. Av det sagda framgår att ett utvärderingsprojekt med seriöst syfte att stödja utvecklingsarbetet kräver tid och resurser, tid för såväl planering som tolkande analys och bedömning. Färdiga modeller kan vara till hjälp, men endast om de används som alternativa redskap och om de väljs med hänsyn till de krav utvärderingsobjektet ställer och underställs en kritisk granskning av i vilka avseenden utvärderingen låses genom modellvalet.

Vad bör beaktas vid bedömningen när utvärderingen syftar till utveckling?

Bedömningskriterierna är de måttstockar som värderingen sker utifrån. När måttstocken väl är bestämd, så anger riktvärden vilket mått som kan anses vara bra eller dåligt, normalt eller onormalt. Till grund för bedömningen kan ligga olika referenspunkter, såsom mål, det förgångna, liknande verksamhet inom landet eller internationellt, bästa existerande praxis, intressenternas förväntningar, förvaltningspolitiska värden, professionella normer, kostnader, intäkter.

Valet av begrepp, synvinklar och frågeställningar för utvärderingen styr valet av bedömningsgrunder. Därför är det skäl att kritiskt granska utgångsläget och medvetet överväga valet av begrepp och frågeställningar. Rådet för utbildningsutvärdering har som lagstadgad uppgift att organisera utvärderingar på nätverksbasis, dvs. att utnyttja en mångsidig expertis. Detta är ägnat att främja att olika värderingar och synvinklar beaktas. I synnerhet om värdekonflikter förekommer bland experterna blir grunderna för valet genomskinliggjorda.

De nationella utvärderingarna av utbildningen och dess olika aspekter kan fokuseras på två olika sätt. Det ena sättet innebär att man utvärderar utbildningsenheter, dvs. man betygsätter utbildningsenheterna med hänsyn till deras kvalitet i fråga om det fenomen som utvärderas. Det andra sättet innebär att utvärderingen fokuserar fenomenet som sådant, dvs. att man betygsätter det nationella tillståndet i fråga om fenomenets olika aspekter. Utbildningsenheterna granskas då inte enskilt utan endast med hänsyn till vissa kontextuella drag.

När enskilda utbildningsenheter betygsätts, definieras kriterierna i förväg och enheternas verksamhet och resultat bedöms med hänsyn till dessa kriterier. Fastställandet av kriterier utgör i detta fall en viktig fas. Den centrala utmaningen är: Vilka kriterier bör väljas, så att utvärderingen beaktar olika utbildningsenheters varierande förutsättningar att bedriva en verksamhet/uppnå ett gott resultat. Faran finns att kriterierna gynnar vissa typer av utbildningsenheter. Om man t.ex. bedömer skolors pedagogiska kvalitet enbart genom att bedöma elevernas kunskaper efter genomgången utbildning, missgynnas de skolor som utbildar elever med svaga initiala kunskaper eller elever med problem av olika slag som

stör deras studier. Om man bedömer skolors kvalitet t.ex. genom hur systematiskt organiserad och dokumenterad enhetens verksamhet är, missgynnas små skolor vilkas styrka kanske är fungerande relationer mellan personalen med möjlighet till snabbt ingripande då problem uppstår.

Det finns en tendens inom utvärderingen att utveckla allmänna bedömningsgrunder eller kriteriesystem som kan utnyttjas oberoende av fenomen (CQAF, CAF o. dyl.). En rutinmässig användning av samma bedömningsgrunder oberoende av fenomen riskerar dock att alltid missgynna samma typer av enheter. Detta i sin tur kan leda till t.ex. utslagning och/eller att den nationella mångsidigheten minskar, när organisationerna tar bedömningskriterierna till riktlinjer för sitt utvecklingsarbete.

När det nationella tillståndet i fråga om ett fenomen betygssätts, är kriterierna likaså viktiga. Kriterierna kan dock formuleras i ett senare skede av utvärderingsprocessen. Det väsentliga är först att analysera och beskriva fenomenet på ett mångsidigt sätt för att finna olika typer av utslag hos fenomenet samt analysera kontextens betydelse för variationen. Bedömningen kan göras sist, speglad mot ett urval av olika kriterier. Denna inriktning på utvärderingen möjliggör bedömning ur olika intressenters synvinklar och ger flera svar, som inte nödvändigtvis behöver sammanfattas till en konsensusuppfattning. Inriktningen ger därför en mer mångsidig bild av verkligheten till stöd för utvecklingsarbetet på nationell nivå. För utvecklingsarbetet på lokal nivå däremot ger bedömning av enskilda utbildningsenheter ett mer direkt och konkret stöd, ifall enheterna informeras om resultatet av utvärderarnas bedömning av dem.

Mot en optimal utvärderingsprocess i utvecklande syfte

Utvärdering är en process som har effekter av många slag. Det uttalade syftet med en utvärdering kan vara gott, men utvärderingen kan ändå komma att få också negativa konsekvenser, om man ser till det övergripande målet för den verksamhet som utvärderas. En noggrann planering med beaktande av kontexten och analys av pedagogiska effekter, såväl primära som sekundära effekter, är därför viktig.

Av etiska orsaker och i syfte att engagera till utvecklingsarbete anses det viktigt att försäkra sig om att de som blir föremål för en utvärdering också själva kan inverka på frågeställningarna, utvärderingsmetoderna och framför allt på kriteriedefinitionerna och att de kan delta i tolkningen av resultaten. Denna interaktiva samarbets- och förändringsinriktade strategi för utvärdering utgör ett viktigt komplement till den upplysnings- och förklaringsinriktade strategin, när man åsyftar utveckling.

De fenomen som utvärderas är ofta komplexa och kräver mångsidiga analyser för att resultaten skall vara etiskt hållbara. Vid organiseringen av utvärderingar bör därför tillräckligt med tid reserveras. Betoningen måste ligga på utvecklingsbara fenomen och konstruktioner för att utvärderingen skall ge positiv energi till de berörda aktörerna och för att utvärderingen skall få en utvecklingsfrämjande funktion.

Litteratur

- Auditering av högskolornas kvalitetssäkringssystem. 2005. Auditeringshandbok för åren 2005–2007. Publikationer från rådet för utvärdering av högskolorna 5:2005. Helsinki.
- Gunnarsson, L. 1987. När måttstocken är problemet. Kritik av utvärderingsforskning. I C. Edlund, C. Gunnarsson, L. Gunnarsson, G. Hermerén, K. Moberg & T. Nilstun, Utvärderingsforskning. Kartläggning av forskningsläget. MURA-projektet. Lund: Studentlitteratur.
- Karlsson, O. 1999. Utvärdering mer än metod. Tankar och synsätt i utvärderingsforskning. En serie kunskapsöversikter från Svenska kommunförbundet 3. Saatavilla www.muodossa.se: <www.mdh.se/isb/mea/rapport_kom_forb_99.pdf>. (Luettu 20.4.07.)
- Kauppi, A. (toim.) 2005. Kehittävä arviointi – periaatteita ja sovelluksia. Helian julkaisusarja C. Ammatillinen opettajakorkeakoulu 12.
- Lyytinen, H. K. & Nikkanen, P. 2005. Kehittämiskoivoo arvioinnista. Teoksessa H. K. Lyytinen & A. Räisänen (toim.) Kehittämissuuntaa arvioinnista. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä.
- Moitus, S. & Saari, S. 2004. Menetelmistä kehittämiseen. Korkeakoulujen arviointineuvoston arviointimenetelmät vuosina 1996–2003. Korkeakoulujen arviointineuvoston julkaisu 10:2004. Helsinki.
- Mäensivu, K., Mäenpää, H., Määttä, M., Volanen, M-V., Knubb-Manninen, G., Mehtäläinen, J. & Räisänen, A. 2007. Lukio- ja ammatillisen koulutuksen yhteistyö opetuksen järjestämisessä. Koulutuksen arviointineuvoston julkaisuja 23. Jyväskylä.
- Sandahl, R. & Bringle, S. 2006. Skolverkets utbildningsinspektion – ger den några effekter? Sverige: Skolverket.
- Shadish, W. R. Jr., Cook, T. D. & Leviton, L. C. 1991. Foundations of program evaluation: Theories of practice. Newbury Park, Calif: Sage.
- Tiihonen, T. 2006. Yhteistyö ja tulosojaus eli miksi syöttää jos vain maalit lasketaan. Esitelmä SVY-päivillä 28-29.3.2006 Jyväskylän yliopistossa.
- Utbildningsutvärderingens nya inriktning. 2004. Utvärderingsprogram för perioden 2004–2007. Publikationer från Rådet för utbildningsutvärdering 2. Jyväskylä.
- Vaherva, T., Malinen, A., Moisio, A., Raivola, R., Salo, P., Kuusipalo, P., Silvennoinen, H. & Vaahtera, K. 2007. Vapaan sivistystyön vaikuttavuus ja suuntaviivaopinnot. Koulutuksen arviointineuvoston julkaisuja 25. Jyväskylä.
- Vedung, E. 1998. Utvärdering i politik och förvaltning. Lund: Studentlitteratur.
- Åberg, J-O. 1997. Det rationella och det legitima. En studie av utvärderingars teori och praktik. Göteborgs universitet: Sociologiska institutionen.

III

Arvioinnin kohteet menetelmällisten valintojen lähtökohtina



Koulutuspoliittinen arviointi

Koulutuspolitiikka rakentaa ja tukee ohjelmia ja interventioita, joiden tarkoitus on lisätä koulutuksen avulla ihmisten, yhteiskunnan ja erilaisten osaryhmien hyvinvointia. Koulutuspoliittinen arviointi on näin makrotason arviointia ohjelmien toteutumisesta ja hyvinvoinnin lisääntymisestä – politiikkakysymysten kääntämistä arviointikysymyksiksi. Metodeinaan se käyttää kustannus-tehokkuus-, kustannus-hyöty- ja kustannus-vaikuttavuusanalyysien menetelmiä. Kansalaisten hyvinvointikokemuksia voidaan lähestyä ymmärtäen käyttäen laadullisia menetelmiä. Artikkelin pohtii tiedonhankinnan eri tasoja, korostaa informaation tiivistämistä indikaattoreiksi, erityisesti aikasarjoiksi ja esittelee erilaisia arviointimalleja. Kuten kaiken poliittisen toiminnan arvottamisessa, koulutuspoliittisen arvioinnin vaikeus tai perimmiltään mahdottomuus on arvojen yksimielisen preferenssijärjestyksen työstäminen arviointipäätöksen kriteeriksi: mikä kunkin mielestä on onnistunutta koulutusta.

Asiasanat: koulutuspolitiikka, arviointimallit, tuotantofunktio, indikaattori

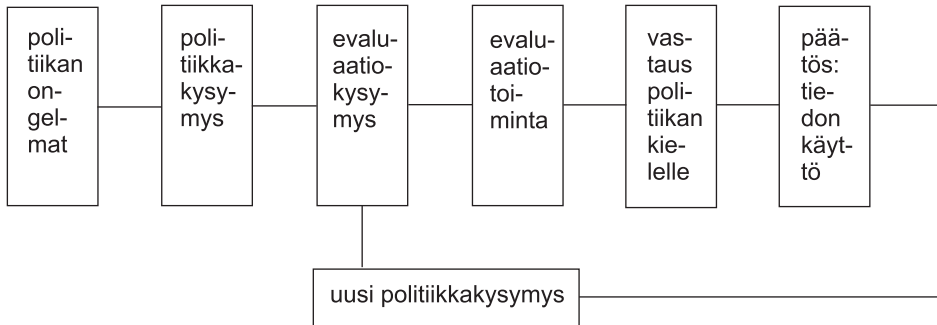
Mitä on koulutuspoliittinen arviointi?

Koulutuksen arviointispektri on laaja: yksittäisen oppilaan koulusuorituksista aina koulutuksen eri alat ja tasot integroivan kokonaisjärjestelmän ja sen ohja-

uksen arviointiin saakka. Mihin tässä kirjossa sijoittuu koulutuspoliittinen arviointi? Rationaalisen koulutuksen ennakoinnin ja suunnittelun edellytyksenä on luottamus siihen, että koulutuksella voidaan saada aikaan sekä yksilöissä että yhteisöissä hallittuja sosiaalisia ja taloudellisia muutoksia, toisin sanoen hyvinvoinnin lisääntymistä. Koulutuspolitiikan arviointi kohdistuu näin harjoitettuun ja suunniteltuun (ex ante) koulutuspolitiikkaan, eli onko luottamus ollut perusteltua. Koulutuspolitiikaksi määrittelemme (Lehtisalo & Raivola 1999, 32) toimet ja pyrkimykset, joilla eri intressiryhmät ja julkisen vallan poliittiset ja hallinnolliset elimet niiden koosteena yrittävät vaikuttaa koulutuksen suuntaamiseen, voimavaroihin, rakenteisiin, prosesseihin, sisältöön ja kaikkien näiden avulla tuotoksiin ja tuloksiin. Kohde sisältää siis yhtä hyvin lainsäädäntöprosessin ja vallankäyttöä korostavan jakopolitiikan (politics) kuin niiden seurauksena syntyneet suunnitelmat ja toiminnot (policy). Arviointi liikkuu tällöin makrotasolla ja rajat suunnitteluun ja tutkimukseen ovat usein sangen joustavat. Hyvän esimerkkinä on lainsäädännön valmistelua varten asetettu asiantuntijakomitea, joka analysoi aikaisempia käytänteitä ja niiden menestymistä, tuottaa vaihtoehtoisia lähestymistapoja esille nostettuun ongelmaan, arvioi niiden kustannusvaikutuksia ja tekee ratkaisuehdotuksia. Arvioinnin tarkoituksena on tuottaa tieteellisin tiedonhankinta- ja analyysimenetelmin uskottavaa ja luotettavaa informaatiota sekä kohteen ympärillä käytävän keskustelun (neuvottelun) että päätöksenteon tarpeisiin. Koulutuspoliittisella arvioinnilla ja arvioijalla on siten selkeä rooli poliittisessa prosessissa. Keskeisiä arvioinnin kohteita ovat kansallisen koulutuksen taso, opetuksen laatu, koulutusjärjestelmän ja sen osien johtaminen ja poliittinen ohjaus, koulutusinvestointien määrä, jakautuminen, osuvuus ja tosiasiallinen kohdentuminen, koulutusvirrat ja koulutusvaranto, järjestelmän eri osien toimiminen yhdessä kokonaisjärjestelmänä (integraatio) sekä yhä useammin organisaatioiden toiminnan tehokkuus ja tuloksellisuus. Käsitys hyvästä yhteiskunnasta ja hyvästä ihmisestä tiivistyy opetussuunnitelmaan tai sen perusteisiin ja tämän käsityksen perusteella esimerkiksi perusopetuksen tuntijakoon. Asiasta käyty keskustelu on uudistusten yhteydessä ollut kiivastakin, mikä osoittaa sen poliittisen merkityksen arvioinnin kohteena.

Yhdysvaltain General Accounting Officen johtaja Eleanor Chelimsky on tehnyt pitkään arvokasta esityötä palvellessaan kongressin ja senaatin lainvalmistelu- ja säätämistyötä koulutuspoliittisen arvioinnin ja ennakoinnin avulla. Hän esittää politiikan ja arvioinnin suhteen kuusiaskelisenä mallina, jota hän sanoo laitoksensa soveltaneen (Chelimsky 1987). Kyse on edestakaisesta käänösprosessista politiikan kielen ja tarpeiden sekä evaluoijan kielen ja asiantuntemuksen välillä.

Politiikan kysymys johdetaan reaalipolitiikan ongelmista ja muotoillaan sel-laiseksi, että evaluaation käyttö ja hyöty voidaan maksimoida. Evaluoija ei siis itse valikoi kohdettaan. Tämä piirre on yksi arviointia ja tutkimusta erottava



Kuvio 1. Koulutuspoliittinen evaluaatio käänösprosessina

tekijä. Evaluoija saa kohteen toimeksiantona tehtävänänsä tuottaa parasta mahdollista informaatiota. Ongelmana saattaa olla, että politiikan tekijä ei osaa tai halua kysyä kaikkein relevanteimpia kysymyksiä. Tässä vaiheessa evaluoija selvittää aiempia vastaavien prosessien kokemuksia (Chelimsky kutsuu menettelyä nimellä Evaluation Planning Review) seuraavien vaiheiden pohjaksi.

Evaluaatiokysymyksen muotoilun lähtökohtana on selvittää, voidaanko politiikkakysymykseen tutkimuksen tai arvioinnin keinoin lainkaan vastata. Esimerkiksi useinkaan ei voida antaa selkeää vastausta syy- ja seurauskysymykseen tai selkeitä arvovalintoja koskevaan kysymykseen. Yleistettävän tiedon saannin mahdollisuus, sen hankkimisen kustannukset ja käytettävissä oleva aika sanelevat kysymyksen muotoilun. Ristiriitaisissa ja herkissä kysymyksissä asetelman ja metodologian on oltava erityisen vahva ja uskottava, minkä vuoksi tapaustutkimus tai pelkän kvalitatiivisen aineiston käyttö harvoin tulee kysymykseen. Kysymys on muotoiltava poliittisen kentän ilmastoa tunnustellen, jotta vastausta ylipäätään voidaan käyttää keskustelussa.

Myös arviointisuunnitelma laaditaan poliittinen konteksti mielessä. Arvioija katsoo useimmiten taaksepäin, poliitikko normatiivisesti ja preskriptiivisesti tulevaisuuteen. Pelkkä palaute toteutettavasta ohjelmasta ei siis riitä. Poliitikko odottaa myös arviota ohjelman toimimisesta tulevaisuudessa. Aineisto kerätään useimmiten kansallisella tasolla ja yhä tavallisempaa on, että se yhdistetään kansainväliseen aineistoon. Arviointisuunnitelmasta on hyvä keskustella tilaajan kanssa tulosten käytön varmistamiseksi. Chelimsky varoittaa välitulosten raportoinnista, koska ne saatetaan tulkita ilman riittävää kontekstia väärin tai ne ovat virheellisiä, jos tiedot ovat vielä tarkistamatta.

Tulokset on myös tehokkaasti viestittävä tilaajalle ja sitä edustavalle kuulijakunnalle. Tutkija ei ole tottunut priorisoimaan tutkimustuloksiaan. Hallinnossa toimivat henkilöt ja poliitikot odottavat selkeää sanomaa: mikä on tärkeää, mitkä ovat järjestykset ja voiko esitetyille ongelmille tehdä jotakin. Raportin on

oltava tiivis ja rönsyilemätön. Kuulijakunnan mukaan räätälöity suullinen raportointi vie viestin tehokkaasti perille. Arviointiraportista järjestetyt kuulemiset ja kommentoinnit ja niiden perusteella tehdyt muutokset suunnitelmaan generoivat uusia politiikkakysymyksiä, joiden esiin nostamisessa arvioijakin voi olla aktiivinen.

Koulutuspoliittisen tutkimuksen vaikeus ja haaste on osoittaa selvä väylä, joka johtaa tehdyistä toimenpiteistä (politiikasta) havaittuihin tuloksiin. Syyseuraus-suhteen osoittamisessa on monia vaiheita, ja arviointi saattaa seurata vain osaa eri mahdollisuuspoluista. Toisaalta se voi tarttua vaikutusketjussa liian varhaisiin välivaikutuksiin. Usein on vaikea mitata relevantteja seuraustekijöitä ja niiden pysyvyyttä. Poliittikaohjelman vaikutuksia on vaikea eristää muista tuloksiin positiivisesti tai negatiivisesti vaikututtavista tekijöistä, koska toimijoilla on elämää toimenpideohjelman ulkopuolellakin ja vaikutusten esiin nouseminen voi kestää kauan.

Useimmat uudistukset tai ohjelmat pyritään toteuttamaan hierarkkisesti ylhäältä alas -mallilla, jolloin organisaatioteoreettisesti korostetaan sosiaalisia rakenteita, yksimielisyyttä, tavoitteiden jakamista ja homogeenista toimeenpanijajoukkoa. Mutta ihmisillä näissä prosesseissa on erilainen tausta ja erilaiset perspektiivit, ja he toimivat eri asemissa sosiaalisissa instituutioissa. Valtaa, ideologista moninaisuutta ja poliittista toimintaa korostava mikropoliittinen näkökulma on silloin tarpeen (Datnow 2000). Tämän ilmaston kuvaamisessa laadulliset tiedonhankintamenetelmät saattavat olla vahvoja.

Julkinen politiikan prosessi koostuu sarjasta toimintoja: työjärjestyksen (agendan) laatiminen, valittavien vaihtoehtojen spesifiointi, poliittinen (autoratiivinen) valinta ja päätöksen toimeenpano. Toiminta tapahtuu kolmena erillisenä ja osittain riippumattomana virtana. 1) Ongelmavirrassa eri asiat kilpailevat poliitikkojen huomiosta ja paikasta agendalla. 2) Ratkaisuvirrassa eri vaihtoehtoja tuodaan esiin, niistä keskustellaan, niitä kokeillaan, korjataan, yhdistellään ja lopulta muotoillaan toteutuspaketti. 3) Poliittikavirrassa lainsäätäjät ja toimeenpanijat tuovat esiin, väittelevät, edistävät tai vastustavat aloitteita. Monivirtamallissa ”politiikkaa tehdään”, kun kolme virtaa yhtyy suotuisassa sosiopoliittisessa tilanteessa. ”Mahdollisuuden ikkunan” avaa joko politiikka- (esim. hallitusohjelman toteuttaminen) tai ongelmavirta: asiat ovat edenneet pisteeseen, joissa niihin on puututtava. Poliitiikan tekeminen ei etene lineaarisesti vaiheesta toiseen, vaan jokainen virta on jatkuvasti läsnä toisistaan riippumatta ja yhtyy vain, kun mahdollisuusikkuna avautuu. Näin ratkaisua saatetaan tarjota ennen kuin ongelma on täsmäntynyt tai esitetään syitä ilmiölle, jota ei riittävästi tunneta. Evaluaatio on läsnä kaikissa näissä virroissa. Poliittista ohjelmaa tai ratkaisua voidaan tämän mukaan tutkia Chelimskyn (edellä) mallin ja matriisikehyksen avulla. Esimerkiksi identifioidaan (I) ongelma, ratkaisuehdotukset, harjoitettu tai suunniteltu politiikka, median osuus siihen, toimenpiteiden ajoitus, (II) kenttä,

joka joutuu ratkaisun kanssa elämään (edistäjät, vastustajat, neutraali osapuoli) ja (III) kontekstoidaan tietolähteet tutkittavaan ilmiöön. (Stout & Stevens 2000.)

Arviointitiedon mittaamisen tasot

Pedagogisten tai didaktisten toimien tehokkuutta arvioidaan tavallisimmin prosessi-tuotos- malleilla. Koulutuksen tuottavuus- tai tuotantofunktioarviot puolestaan ovat tyypillistä koulutuspoliittista arviointityötä. Ne pyrkivät selvittämään, mikä yhteys erilaisilla panostekijöillä (useimmiten fyysiset voimavarat) on koulutuksen välittömiä oppimistuloksia laajempien tavoitteiden saavuttamiseen. Koulutuspoliittisten ohjelmien hyödyllisyyden arviointi (policy analysis) on esimerkinä tästä. Kuitenkin jos koulutus nähdään panos-prosessi-tuotos-tulosmallin mukaisesti, jolloin eri vaiheet rakentavat mahdollisuuden (kapasitaatio) seuraavan vaiheen onnistumiselle, koulutuspoliittinen arviointi ei voi välttää mittaamasta oppimistuloksia, joiden sitten katsotaan voivan vaikuttaa asetettujen päämäärien saavuttamiseen tai niitä kohti etenemiseen. Oppimissaavutuksiakin on vaikea luotettavasti mitata. Kennedy (1999) puhuukin eri tarkkuustasoilla tehdyistä arvioista (approksimaatiosta).

Ensimmäistä tarkkuustasoa edustavat suora havainnointi ja standardoidut testit. Jälkimmäisiä on helppo soveltaa suurellekin joukolle. Havainnointia on kuitenkin mahdoton standardoida. Erilaisia menetelmiä havainnointiin on runsaasti. Eri havainnointitutkimusten tuottamia tuloksia on kuitenkin vaikea vertailla: näin on jopa samasta luokasta tai koulusta saatujen eri ajankohtina tehtyjen havaintojen kohdalla. Havainnointi tuottaa suuren ja vaikeasti hallittavan aineiston, mikä vaikeuttaa sen laajaa soveltamista koulutuspoliittiseen arviointiin.

Toisen tason muodostaa toimijoiden, useimmiten opettajien, tuottama aineisto omista teoistaan, toiminnastaan tai ympäristön tapahtumista reaaliaikaisesti. Kennedy (emt.) kutsuu näitä situaationaaliksi toiminnan kuvauksiksi, koska tarkoituksena on tuottaa tarkka kuvaus tapahtuneesta. Sen sijaan tarkoituksena ei ole tuottaa asenteita peilaavaa yleistä pohdintaa siitä, miten asioiden tulisi olla. Toimijoiden pitämät lokikirjat tai päiväkirjat tai hypoteettisiin tilanteisiin (vignetit) tuotetut reaktiot tuottavat myös rikkaan mutta vaikeasti hallittavan aineistomassan. Usein opettajilta kysytään myös koulun talouteen tai hallinnolliseen ja poliittiseen ohjaukseen kuuluvia asioita, joista heillä ei kuitenkaan välttämättä ole faktatietoa.

Kolmannen tason arviot poikkeavat edellisistä siten, että ne ovat eräänlaisia viivästettyä arviointia, tapahtuneen jälkeen muistelua. Kyselylomakkeilla tai haastatteluilla tuotettu aineisto on tavallinen aineistotyyppi erillisissä arviointihankkeissa.

Neljäs taso on etäännytynyt kaikkein kauimmaksi muutokseen pyrkivästä toiminnasta. Siinä vastaajat pohtivat toimenpiteiden, ohjelmien tms. hyödyllisyyttä, käyttökelpoisuutta, kustannuksia, vaikutuksia jne. – toisin sanoen yleistä tyytyväisyyttään tai tyytymättömyyttään ohjelmaan. Selville saadaan vastaajien arvoja ja asenteita.

Arvioinnissa ja arviointitutkimuksessa voidaan kerätä neljänlaista kvantitatiivista aineistoa (ks. esim. Fitz-Gibbon 1996). Aineistotyyppit on syytä pitää erillään, koska niiden perusteella tehtävät johtopäätökset ovat hyvin eri vahvuisia. Numeromuodossa oleva data ei ilman muuta ole objektiivista ja tulkittavissa. Raaka-aineisto, esimerkiksi koulusaavutuspiistemäärät, poissaolotunnit, oppilaskohtaiset menot tai rehtorin kokemus virkavuosina, muodostavat ainoastaan pohjan jatkoanalyysille. Hyvä, tai pitäisikö sanoa huono, esimerkki ovat Englannissa käytetyt ”liigataulukot”, joissa koulut pantiin paremmuusjärjestykseen koulukokeiden koulukohtaisten keskiarvopistemäärien mukaan. Suuri yleisö ja jopa jotkut hallintoviranomaiset ja poliitikot lukivat listat osoituksena kunkin koulun laadusta. Raaka-aineisto on kuitenkin epäkontekstuaalista, toisin sanoen se on irrotettu kaikista siihen yhteydessä olevista tekijöistä. Ainoa johtopäätös, joka siitä voidaan tehdä, on tieto mitatun ominaisuuden jakaumasta, hajonnasta ja vaihteluvälistä. Edes koulupiirin tai muun alueen keskiarvotieto ei ilman vertailukohtaa sisällä mielekästä informaatiota.

Minimivaatimus tulkintaa varten on vertailuaineiston olemassaolo. Vertailuaineisto voidaan muodostaa myös perusaineistosta jakamalla se osa-aineistoiksi jonkin luokitteluperusteen mukaan. Esimerkiksi oppilaskohtaisia menoja voidaan tarkastella kunnan asukaskohtaisen verotulokertymän, kunnan sijainnin, elinkeinorakenteen tms. mukaisissa luokissa tai koulujen saavutuspiistemääriä maahanmuuttajaoppilaiden tai työttömien vanhempien suhteellisen osuuden mukaisissa luokissa. Voimme päätellä erojen olemassaolon ja kenties sen, että ne ovat systemaattisia. Emme sen sijaan voi päätellä, ovatko erot odotettuja tai ovatko ne yhteydessä muihin luokitteluperusteisiin tai vielä vähemmän, mistä erot johtuvat. Raportoitavaksi jää karkea eri tavoin vertailuryhmittäin luokiteltu kuvaus raaka-aineistosta. Se ei vielä tarjoa riittävästi välineitä suunnittelun avuksi tai havaitun epätyytyttävän asiantilan korjaamiseksi, ainoastaan sen havaitsemiseksi. Vertailuaineiston arvo on suurimmillaan aikasarjoissa, joissa kohde on itsensä vertailuryhmä: samoja muuttujia ja samoja luokitteluperusteita käyttämällä voi havaita systemaattisia muutossuuntia, trendejä. Tämä toimii herätyskellona asiaan puuttumiseksi.

Matalin taso, jonka perusteella esimerkiksi resurssien allokointipäätöksiä, organisaatioiden tai henkilöiden sanktiointipäätöksiä, normatiivisten ohjeiden antamista tai toiminnan radikaalia uudelleen suuntaamista voi perustella, on sovitettun eli residuaaliaineiston taso. Sen soveltamiseksi on oltava teoreettinen tai käytännön kokemuksesta johdettu malli, joka jakaa mitatut muuttujat selitet-

täviin ja selittäviin muuttujiin ja määrittelee niiden keskinäiset suhteet. Usein vertailukohtana on ns. perustaso eli ajankohtana 1 mitattu taso ja lisäarvona suoritustaso ajankohtana 2. Residuaaliksi tilastotiede määrittelee erotuksen mitatun ja tilastomallista ennustetun tuloksen välillä. Nimitys tulee siitä, että termi kuvaa selitettävän muuttujan, vaikkapa ylioppilaskokeiden pistemäärien (vaihtelusta), sitä osuutta, mikä jää jäljelle, kun on otettu huomioon muut mallissa osoitetut siihen vaikuttavat tekijät. Puhutaan myös lisäarvoaineistosta. Tämä nimitys on oikeutettu esimerkiksi silloin, kun tutkitaan, paljonko oppilaat ovat oppineet lisää edelliseen mittauskertaan verrattuna. Kun käytännössä on mahdoton kontrolloida kaikkia toimenpideohjelman ulkopuolisia tekijöitä, paljon käytetty tapa onkin lisäarvon malli: paljonko edistystä on tapahtunut riippuvassa muuttujassa? Ajattelu perustuu seuraavaan malliin (ks. Ludwig & Bassi 1999):

$$Y_{iT} = \alpha_0 + \alpha_1 T S_{iT} + \alpha_2 T F_{iT} + \sum \alpha_{1t} T S_{it} + \sum \alpha_{2t} T F_{it} + \epsilon_i + \epsilon_{iT} + \sum \epsilon_{it} \quad (t = 1, 2, \dots, T-1)$$

Oppilaan (i) testipistemäärä (Y) mittausajankohtana (T) on koulu- (S) ja perhetekijöiden (F), niiden aikaisempien vuosien vaikutusten sekä mittaamatta jääneiden tekijöiden (regressioyhtälöissä ns. virhetermi) mittausajankohdan ja edeltävien vuosien vaikutusten funktio. Ongelmaksi tietenkin jäivät mittauksetta jääneet tuntemattomat tekijät.

Paneeliaineisto antaa mahdollisuuden puuttua ongelmaan oppimissaavutusten lisäarvon avulla. Tässä mallissa aikaisempien vuosien testipistemääriä käytetään kovariaattina kontrolloimaan tuntemattomia tekijöitä:

$$Y_{iT} = b_0 + b_1 T S_{iT} + b_2 T F_{iT} + b_3 T Y_{i,T-1} + v_{iT}$$

Malli on kuitenkin harhaton vain siinä tapauksessa, että koulutekijät eivät korreloi virhetermin (v) kanssa ja että mittaamatta jääneet ja viivästyneet perhe- ja oppilastekijät ovat läsnä myös aikaisemmissa mittauksissa ja testipistemääriissä ja että eri vuosien satunnaisvaikutukset eivät korreloi keskenään. Malleja voidaan toki käyttää muidenkin asioiden kuin koulusuoritusten arviointiin.

Residuaaliaineiston tilastollinen käyttö perustuu mitattujen muuttujien väliin korrelaatioihin. Kun koulutuksessa, niin kuin kaikessa sosiaalisessa elämässä, lähes kaikki ”hyvät” asiat korreloivat keskenään (samoin kuin ”huonot” asiat), käytettyihin malleihin voidaan ottaa vain osa relevanteista tekijöistä. Residuaali sisältää aina tätä tietämättömyyttä, jonka analyysimalli ilmoittaa selitysasteena (100 % -selitysaste). Tilastotieteen peruskurssilla korostetaan, että korrelaatio ei kerro itsessään vaikutuksen suuntaa. Korrelaatiosta ei voi johtaa syy- ja seuraussuhteita eli kausaalisuhteita. Hienostuneimmat monimuuttuja-analyysitkaan, joista tavallisimmin käytetään arviointitutkimuksessa regressioanalyysin sovel-

lutuksia, eivät aukottomasti paljasta syitä ja seurauksia, ainoastaan muuttujien yhteisvaihtelun ja eri muuttujien suhteellisen osuuden tästä yhteisvaihtelusta.

Ainoa keino, jonka tilastotiede tarjoaa tietämättömyystekijän poistamiseksi residuaalista, on kokeellinen asetelma satunnaisotantoineen. Pedagogisessakin arvioinnissa ja tutkimuksessa koejärjestelyt ovat mahdollisia. Kun alussa koulutuspoliittinen arviointi määriteltiin makrotason arvioinniksi, sillä tarkoitettiin, että arviointi kohdistuu koko koulutusjärjestelmään, sen määriteltyyn osaan tai osan toimintaan. Kokonaisilla järjestelmillä on hankala, ellei mahdoton, kokeilla, sen poliittiset järjestelmäkokeilut ovat osoittaneet. Suomessa on tavallista ennen järjestelmätason muutoksia turvautua erilaiseen kokeilutoimintaan. Nämä eivät kuitenkaan ole tieteellisiä kokeita, ne kestävät yleensä liian lyhyen ajan ja niiden tuottaman aineiston perusteella tehtävät johtopäätökset saattavat siis olla harhaisia. Useimmiten poliittinen päätös toiminnan omaksumisesta on jo tehty ennen kokeilua. Jos kokeilu ei tuota selkeästi odotustenvastaisia tai muuten huomiota herättävän negatiivisia tuloksia, sen katsotaan vahvistaneen toimintapolitiikkaa. Kun koulutusjärjestelmää koskevan tiedon kuitenkin tulisi olla yleistettävää, merkitsee se, että arviointi- tai tutkimustietoa perusteenaan käyttävän poliittisen ratkaisun tulisi perustua residuaalitaso aineistoihin.

Kaikki sosiaaliset järjestelmät prosessoivat ihmisiä. Muutoksen onnistuminen onkin aina alimman toimijatason tuote. Tavaksi on tullut puhua katutason byrokraatista. Hänen merkityksenantoaan tai mielekkyyskokemuksiaan ei kova kvantitatiivinen mittaus tavoita. Hänen maailmansa ei myöskään muodostu pelkästään koulutuksen avaamasta maisemasta. Koulutus instituutiona on osa sosiaalista järjestelmää, jonka osat vaikuttavat koulutukseen ja joihin koulutus vaikuttaa. Voivatko koulutuksen mekanismit vähentää yhteiskunnallisia jännitteitä ja lisätä koheesiota yhteiskunnissa, joissa sosiaaliset ristiriidat (esimerkiksi laaja maahanmuutto) aiheuttavat vastakkainasettelua? Laajakaan survey-tutkimus ei tavoita tällaisissa tapauksissa olennaisia tekijöitä, eikä pelkästään kouluun keskittyminen paljasta, mitkä nämä mekanismit ovat. On näet kahdenlaisia elementtejä, jotka mahdollistavat ihmisten välisen yhteistyön ja ymmärryksen: 1) institutionaaliset säännöt ja 2) vakaannuttavat traditiot. Näiden rajaamassa todellisuudessa on neljänlaisia toimijoita koheesiota tuottamassa: poliittiset elimet, taloudelliset toimijat, sosiaaliset toimijat (esim. kirkko) ja kasvatukselliset toimijat. Koulu vaikuttaa kolmella tavalla: se välittää tietoa sosiaalisista sopimuksista, tuottaa sopimusten mukaista käyttäytymistä ja auttaa ymmärtämään, mitä seuraamuksia sopimusten rikkomisesta on (Heyneman 2000). Poliittisessa evaluaatiossa ei näin ollen voi nostaa kohdetta irti kontekstistaan eikä ihmisistä, jotka antavat kohteelle mielen.

Perinteistä ekonometrasta tai survey-tyypin aineistoa voidaan näin tarkentaa ja paremmin ymmärtää kvalitatiivisella aineistolla, mutta mielestäni high stake-päätöksiä ei voi perustaa puhtaasti kvalitatiiviselle aineistolle. Tällä tarkoitetaan

päätöksiä, joilla on merkitys suurelle joukolle ihmisiä tai suuri merkitys pienellekin joukolle ihmisiä. Poliittikka-analyysin ja järjestelmätason arvioinnin perusteella vedetyt johtopäätökset eivät voi perustua uniikkeihin tapauksiin.

Indikaattorit

Monimuotoisen todellisuuden tiivistämisessä ja hallitsemisessa on tarpeellista indikaattoreiden käyttö. Poliittisen tai hallinnollisen ohjauksen mahdollistamiseksi tarvitaan arvokartoitusten ohella aikasarjaindikaattoreita, jotka kertovat kohde-toiminnan sekä tilassa että ajassa tapahtuneista muutoksista. Indikaattoriajattelu lähtee organisaation suunnitellun tuotoksen määrittelystä ja mittaamisesta. Sen jälkeen on kysyttävä, millä panoksilla nämä suoritukset on saatu aikaan. Tyyppilliselle kustannus-hyöty-analyysille riittääkin näiden kahden indikaattoriryhmän muodostaminen. Vertailevia toiminnan tehokkuuspäätelmiä voidaan tehdä niiden perusteella, mutta vielä ei voida pohtia syitä tehokkuuseroille. Seuraavaksi on käännäyttävä prosessien puoleen: minkälaiseksi toiminnaksi organisaatio muuntaa resurssit ja miten se allokoii resursseja eri toiminnan kohteisiin?

Oikeudenmukaiselle indikaattorien käytölle on esitetty useita kriteereitä, mm. seuraavia (Fitz-Gibbon 1996):

1. Moniportaisessa hallinnossa organisaation eri tasot ovat vastuussa eri tuotoksista tai eri määrin samastakin tuotoksesta. On siis määriteltävä toimintayksikkö, jota indikaattori koskee ja jonka suoritteita se indikoi.
2. Tulosindikaattoriksi voidaan valita vain sellaista toimintaa kuvaava mitattain, johon arvioitava yksikkö voi vaikuttaa. Tulosvastuuta ei voi vaatia ilman kausaalisuuden osoittamista tai olettamista toiminnan ja tulosten välillä.
3. Tulosindikaattorit ovat kontekstuaalisia, riippuvat monesta tekijästä. Siksi ne on muotoiltava suhteelliseksi, usein suhdeluvuiksi.
4. Arviointi- ja indikaattoritieto on palautteena jaettava aina pienimmälle toiminta- tai hallintoyksikölle saakka.
5. Indikaattoreilla on oltava ilmeisvaliditeetti, laaja yhteisymmärrys siitä, että ne ovat oikeudenmukaisia ja kohdistuvat tärkeisiin asioihin.
6. Indikaattorien käyttäjien ja niiden perusteella toimenpiteiden kohteeksi joutuneiden on ymmärrettävä indikaattoreiden sisältö ja niiden muodostamistapa. Niiden tulee olla siis viestittävässä.
7. Indikaattorien ja ne tuottaneiden mittausten tulee olla harhattomia.
8. Edellinen merkitsee sitä, että ne on voitava tarkistaa alkuperäisestä aineistosta.

9. Aikasarjoina käytettyjen indikaattoreiden on säilytettävä tulkittavuutensa ja alkuperäinen tarkoituksensa vuodesta toiseen. Jos indikaattoreilla seurattu toiminta paranee, sen tulee näkyä indikaattorin arvossakin.
10. Indikaattoreiden tulee olla hyödyksi enemmän kuin haitaksi, ja niiden rakentamisen sekä tietokannan ylläpitämisen tulee olla kustannustehokasta.

Koska mittaamisen kohteeksi joutuminen saattaa johtaa reaktiiviseen käyttäytymiseen, mittaluvuista muodostetut indikaattorit saattavat olla harhaisia kahdella tavalla. Niitä pyritään väärentämään tai muodostamaan omaa käyttäytymistä suosivalla tavalla. Vuotuinen keskimääräinen opiskelijamäärä ilmoitetaan yläkanttiin pitämällä pitkään kirjoissa opiskelunsa keskeyttäneitä tai niitä, jotka eivät aloittaneetkaan opiskelua tai oppilaan aiheettomille poissaoloille pyritään löytämään jokin muu luokka. Toisaalta organisaatio ja siinä toimivat ihmiset pyrkivät parantamaan indikaattorien arvoja turvautumalla käyttäytymiseen, joka saattaa olla vaikutuksiltaan halutun tehostamista vaativan toiminnan vastaista. Koulutuksen laatu saattaa huonontua, jos opetuksen tehostamisen sijasta lasketaankin tavoitteiden tasoa, esimerkiksi määrittelemällä hyväksyn kurssisuorituksen alaraja entistä alhaisemmaksi. Indikaattoreiden tarkoitus on tehostaa ja parantaa toimintaa, ei johtaa sitä epätoivottuun suuntaan (Fitz-Gibbon 1996.)

Kritiikki väittää makrotason arvioinnin kutistavan monimuotoisen todellisuuden mitattavaksi todellisuudeksi ja instituutioiden hallintapyrkimysten kutistavan mitattavan todellisuuden edelleen indikaattoritodellisuudeksi. Keeves ym. (2005) tiivistävät indikaattorihjauksen ongelmat teknis-analyttisiksi ja käytettävyysongelmiksi. Edellisistä he pitävät vakavimpina validiteettiongelmia: indikaattori kuvaa hyvin kapeasti tarkoitettua, indikaattori ei kuvaa tärkeää aspektia vaan helposti mitattavaa tai indikaattori on irrotettu kontekstista, joka vasta antaisi sille merkityksen ja tulkintamahdollisuuden. Makrotason mittaukset perustuvat useimmiten otantaan ja tuottavat mitatuissa muuttujissa suuren hajonnan, mikä merkitsee tunnusluvuissa suurten luottamusvälien tarvetta. Koulusaavutusindikaattorit perustuvat kohorttitietoihin oppilaista, jotka tulivat kouluun ja joiden koostumus on muuttoliikkeen vaikutuksesta suurestikin vaihdellut vuosia aikaisemmin. Opettajatkin muuttavat ja rehtorit vaihtuvat. Koulutuspoliittisessa tutkimuksessa ja arvioinnissa on usein tarkastelu- ja tilastoyksikkönä koulu. Se ymmärretään pelkistetysti organisaatiotasoksi, vaikka koulu on sangen monikerroksinen ja monimuotoinen kooste yksilöitä. Se on sosiaalinen ja pedagoginen yhteisö, tuotantolaitos ja julkishallinnon yksikkö. Siellä opetetaan eri aineita eri tavalla. Siellä on monenikäisiä oppilaita, joita ryhmitellään eri tavoin. Tätä moniaineisuusutta on vaikea vangita indikaattorein. Keskiarvosuorituksina samantyyppisiksi arvioitujen koulujen tulokset ovat saattaneet syntyä hyvinkin erilaisten prosessien ja organisatoristen ratkaisujen tuloksena.

Käytettävyysongelma näkyy indikaattorin oletetun hyödyntäjän näkökulmasta. Sama informaatio samassa muodossa ei välttämättä palvele kaikkia osapuolia: poliitikkoja, hallintoa, vanhempia, oppilaita, opettajia ja kiinnostunutta yleisöä. Tietojen saatavuus ja ymmärrettävyys vaihtelee käyttäjien välillä. Indikaattorien harkitsematon käyttö ja julkistaminen saattaa johtaa tuhoisaan noidankehään itseään toteuttavana huonon koulun ennusteena. Eettiset ongelmat ja moraaliset vaikutukset kuuluvat käytettävyysongelmiin pohdintaan. Koska indikaattori vain kuvaa ja paljastaa eikä kerro mitään syistä eikä anna selityksiä, siitä ei voi lukea, mitä ja miten toimintaa tulisi muuttaa ja onko kohteella edellytyksiä tehdä tarpeellisia muutoksia. Indikaattorit ovat tarpeen herättäjinä, yleiskuvan antajina ja aikasarjojen mahdollistajina, mutta niitä ei pidä käyttää harkitsemattomasti. Edellä mainittuja ongelmia pitäisi pyrkiä vähentämään jo etukäteen.

Arviointimallit

Meillä laiminlyöty arvioinnin laji on päätösten ja suunnitelmien toimeenpanon seuraaminen, implementaatiotutkimus tai -arviointi. Aivan kuten ensimmäisen polven tuotantofunktiosovellutuksissa, joissa panokset muuntuivat mustassa laatikossa tuotoksiksi, nykyään uskotaan, että hyvät suunnitelmat tai poliittiset päätökset virkamiesten ja etulinjan toimijoiden toteuttamana sinänsä ja itsestään joko muuttuvat tai eivät muutu halutuksi tavoitteeksi. Arvioidaan siis vain lähtötilanne ja tulokset. Implementaatio vastaa panos-tuotos-mallin prosessitekijöitä, sitä dynamiikkaa, joka konkretisoi tavoitteet havaittuna muutoksena. Implementaation onnistumiseen vaikuttavat kohdejärjestelmän kompleksisuus ja koko (koululuokka tai koko koulutusjärjestelmä), eri organisaatioiden väliset suhteet, toimijoiden sitoutuminen ohjelmaan ja edellytykset (kapasiteetti) toimia odotetulla tavalla. Pysyvän muutoksen saa aikaan ainoastaan pienimmän yksikön, yksilön, sitoutuminen tavoitteen saavuttamiseen, hänen motivoitumisensa: miten mielekkäinä ja merkityksellisinä itselleen hän pitää ohjelmalle asetettuja tavoitteita? Implementaatioarvioinnissa on näin välttämätöntä kerätä informaatiota sekä makro- että mikrotason tekijöistä, kepeistä ja porkkanoista. Poliitikko voi asettaa tavoitteita ja tukea niiden saavuttamista, mutta se voi sangen rajallisesti vaikuttaa yksilöiden mielekkyyskokemuksiin ja merkityksenantoon. Siksi implementaatio on nähtävä prosessina ja omana järjestelmänään, jossa on meneillään jatkuva neuvottelu osapuolten välillä ja sen seurauksena hidas ja vaiheittainen muutos. Muutoksia odotetaan liian nopeasti ja siksi turvaudutaan liian varhaisessa vaiheessa summatiiviseen tulosarviointiin, kun sen sijaan pitäisi arvioida implementaatioprosessia. (McLaughlin 1987.)

Monet politiikkakysymyksistä johdetut arviointikysymykset voidaan nähdä taloustieteestä omaksutun tuotantofunktioajattelun sovellutuksena. Tähän on jo

edellä viitattukin. Järjestelmän sisäistä prosessia, joka muuntaa koulutuksen panostekijät tavoitteiden mukaiseksi koulutustuotokseksi määrätyillä hinnoilla ja käytössä olevalla tuotantovälineillä, kutsutaan koulutuksen tuotantofunktioksi (Raivola 2000). Käytetyt lähestymistavat voidaan luokitella kustannus-hyöty-, kustannus-vaikuttavuus- tai kustannus-käyttökelpoisuus-analyysiksi. Kaikki edellyttävät lähtökohdakseen panostekijöiden identifiointia ja kustannusanalyysia. Kaikki pyrkivät auttamaan kustannustehokkuudessa: suotuiset tulokset pyritään saamaan aikaan taloudellisesti edullisilla panoksilla. Erot syntyvät siitä, miten suhteen jälkimmäinen osa ymmärretään. Kustannus-hyöty-tarkastelu laskee hyödyt rahallisina arvoina: ovatko toiminnasta saadut hyödyt kustannuksia suuremmat? Kustannus-käyttökelpoisuus (tai hyödyllisyys) -analyysi taas arvioi tuloksia sen mukaan, mikä subjektiivinen arvo niillä on päätöksentekijälle tai esimerkiksi poliittiselle ohjelmalle. Kustannus-vaikuttavuus-analyysi ymmärtää puolestaan tulokset pedagogisin määrein ja laajemmin sosiaalisina impakteina yhteiskuntaan. Siihen sisältyy vertaileva tarkastelu eri investointikohteiden välillä, mikä on välttämätöntä, kun suunnitellaan ja arvioidaan vaihtoehtoisia julkisia investointeja ja etsitään kustannus/vaikutus-suhteeltaan parhaita. Taloustieteellisen ajattelun soveltamisen koulutukseen tekee kuitenkin vaikeaksi se, että psykologisia ja sosiaalisia hyötyjä ja kustannuksia on paljon vaikeampi mitata kuin vastaavia fyysisiä suureita. Koulutus on ”monituoteyritys”, joka saa aikaan monenlaisia vaikutuksia, myös odottamattomia ja negatiivisia, joita on vaikea sijoittaa määrätyn ohjelman tai koulutuksen ansioksi tai vaikutukset realisoituvat vasta pitkän ajan kuluttua. Koulutuksessa on useita sisäisiä ja peräkkäisiä tuotantofunktioita, joiden omavaikutuksia on vaikea mitata. (Levin 1988; Rice 1997.)

Ehkä pahiten laiminlyöty alue koulutuspoliittisessa arvioinnissa on kustannusanalyysi, vaikka se on kaikkien edellä mainittujen analyysien välttämätön ehto. Kaikki menoerät tulisi identifioida, erityisesti henkilöstökustannukset, koska koulutus on yksi työvoimavaltaisimpia aloja. Myös vapaaehtoistyö ja eri tavoin koulutettujen työntekijöiden koulutuskustannusten ero tulee huomata. Tavallisesti suurin menoerä koulutusajalta ovat ansaitsematta jääneet tulot ja yhteiskunnalta saamatta jääneet verotulot sekä näille laskettu korko. Ero on tehtävä vaihtuvien kulujen ja kiinteiden kulujen välillä, joissa esimerkiksi rakennusten ja laitteiden osalta on kirjattava vuotuiset arvonalennukset. Toiseksi, jokaiselle menoerälle on määriteltävä hinta, markkinoilta ostettaville markkinahinta ja muille arvioitava varjohinta. Kaikki kuluerät lasketaan yhteen. Kolmanneksi analysoidaan yksikkökustannukset arviointitehtävän ja arvioinnin tulkintatarpeen mukaan. Kustannukset voidaan määritellä oppilaskohtaisina, mutta yhtä hyvin koulukohtaisina, kuntakohtaisina, koulutuslakohtaisina, oppilaitosmuodon mukaisesti jne. Analyysiin kuuluu myös, kuka kustannukset kattaa, missä suhteessa eri maksajien välillä ja mistä lähteestä. Oikeudenmukaisuuskysy-

mykseksi pohdinta muodostuu, kun tarkastellaan, kuka korjaa hyödyt ja kantaa kustannukset koulutuksesta tai interventioista. Lisäksi tärkeä kysymys on, onko ohjelman rakentamiseen ollut varaa ja miten pysyvän toiminnan kustannukset pitkällä tähtäyksellä katetaan. Siltana edellä mainittuun kolmeen analyysityyppiin palvelee vaihtoehtoiskustannusten käsite: mitä muita ihmisten tarpeita yhtä suurilla kustannuksilla olisi voitu tyydyttää sen sijaan, että on valittu määrätyn ohjelman toteuttaminen? Tai voidaan myös kysyä, olisiko sama hyöty tai vaikutus saatu aikaan erilaisella kustannusrakenteella. (Levin 1988; Rice 1997.) Colemanin raportista alkaen on kiistelty resurssien merkityksestä koulun tuloksille. Esimerkiksi oppilaskohtaisia menoja on käytetty ennustamaan oppimistuloksia. Harvemmin on oivallettu, että kustannusanalyyseissa voidaan käyttää näitä menoja riippuvana muuttujana.

Kustannusten hallinnassa on kysymys toiminnan tehokkuuden ja tuottavuuden parantamisesta. Tuore OECD:n työpaperi näkee julkisen sektorin tehokkuuden muodostuvan kolmesta tekijästä: (1) kyvystä priorisoida tarpeet ja allokoida voimavarat tehokkaasti, (2) sen jälkeen kun prioriteetit on määrätty, kyvystä hallita ja ohjata kuluvirtoja tehokkaasti ja (3) kyvystä organisoida palvelujen tuottaminen tehokkaasti (markkina- tai tuotantotehokkuus) (Gonand ym. 2007).

Yhdysvaltalaisessa arviointityössä tuotantofunktio malli on irrotettu taloustieteellisestä kytkennästään, ja sitä on alettu kutsua loogiseksi ohjelmamalliksi. Looginen ohjelmamalli jäsentää prosessit, toimijat ja kontekstin selkeäksi kokonaisuudeksi ja ohjaa ohjelman toimeenpanoa ja arviointia. Se identifioi ja esittää eri tekijät usein kaaviona vaikutussuuntanuolineen:

- kohderyhmät: koulut, yksilöt, ryhmät, toiminnot
- resurssit: ihmiset, fyysiset resurssit, vapaaehtoiset, tarvekartoitukset, rahoituksen, mutta myös oppilaat ja opettajat, käytetyn ajan, säännökset, ideat, toimintaperiaatteet
- tavoitteellinen toiminta mutta myös häiriötoiminta
- osajärjestelmät
- välittömät ja viivästyneet tulokset, sekä konkreettiset ja havaittavat (output) että oppiminen, ymmärtäminen, taidot, käyttäytyminen, muutos (outcome)
- mittausjärjestelmän ja indikaattorit.

Toista, kustannustietoisuutta korostavaa mallia, kutsutaan tulosorientoituneeksi tai tulosrahoitusmalliksi. Se korostaa koulutusta investointina ja vaihtaa palveluaktiiviteetin korostamisen (mitä ja miten tehdään) suoritusavoitteisiin (millä vaikutuksella, mitä muutoksia). Malli käyttää ohjauskeinonaan välitavoitteiden saavuttamisen tarkastamista, mikä sallii korjaustoimiin ryhtymisen ajoissa. Mallin muunnos on tulosvastuun malli, joka yhdistää suunnittelun ja arvioin-

nin samaksi saumattomaksi prosessiksi. Ne vain kulkevat mallissa vastakkaisiin suuntiin. Malli kysyy, mitä halutaan saada aikaan, miten tavoitteen saavuttamisen tunnistaa ja mitä on tehtävä tavoitteen saavuttamiseksi. Tulostavuuksissa erotetaan tekijöiden ja ohjelman tulosvastuu toisistaan. Tarkentuva ohjelman tulosmalli perustuu hierarkkisiin ja peräkkäisiin askeliin ja vastaa kysymyksiin, miksi ohjelma on olemassa, miten se toteutetaan, millä resursseilla, mitkä ovat sen tulokset ja mitä vaikutuksia ohjelmalla on ollut. (Penna & Phillips 2005.)

Mallien taustalla on muutoksen teoriaksi kutsuttu näkökulma. Se nähdään keinona kehittää ratkaisuja monimutkaisiin yhteiskunnallisiin ongelmiin. Ideana on tunnistaa ja mitata aikaisia ja keskipitkän aikavälin saavutuksia, jotka raivaavat tietä pitkän aikavälin tavoitteiden saavuttamiselle. (Vrt. tulevaisuuden tutkimuksen feed forward -käsite.) Lähestymistapa ohjaa määrittelemään edellytykset kunkin vaiheen tavoitteiden saavuttamiselle ja luomaan toimenpiteet näiden edellytysten rakentamiselle. Myös indikaattorit, joiden avulla seurataan interventioiden onnistumista, ovat tarpeen. Pitkän aikavälin tavoitteiden asettamien pakottaa arvopohdintaan ja priorisointikeskusteluun. Harva ohjelma yksinään pystyy toteuttamaan laajoja sosiaali- ja koulutuspoliittisia tavoitteita. Muutoksen teoria kutsuu sektorirajat ylittävään integroivaan suunnitteluun ja luo edellytyksiä onnistuneelle monitasosuunnittelulle. (Anderson 2005.)

Lopuksi

Edellä on korostettu teknis-rationaalista näkökulmaa koulutuspoliittiseen arviointiin sillä perusteella, että se on suunnitteluyhteiskunnan tavallisin tapa nähdä se. Näkökulma ei kuitenkaan ole ainoa eikä aina riittävä. Instituutio- ja tulkinallinen näkökulma näkevät arvioinnin tehtävän eri tavalla, kuten seuraava taulukko osoittaa. Käytännössä näkökulmat lainaavat aineksia toisiltaan.

Taulukko 1. *Koulutuspoliittisen arvioinnin näkökulmat*

Ulottuvuus	Instituutioteoria	Rationaalinen suunnitteluteoria	Tulkinnalliset teoriat
Avaintoimijat	asianosaiset	politiikan tekijät, asiantuntijat, hallinto	paikalliset yhteisöt
Funktio	legitimoiva	tekninen käyttö	sosiokulttuurinen
Yhteys toimeenpanoon	julkisen tuen säilyttäminen	instrumentaalinen tavoitteille	arvojen ja merkitysten muokkaaminen
Peruskäyttö	neuvoteltu konsensus politiikan alasta ja resursseista	politiikan hienosäätö, tavoitteet–keinot-suhteiden osoittaminen	politiikan merkityksen ymmärtäminen ja toimijoiden identifiointi

Koulutuksen, kuten muidenkin sosiaalisten interventioiden, tarkoituksena on kansalaisten ja yhteisöjen hyvinvoinnin lisääminen. Perimmäiset tuloksellisuusindikaattorit on siis etsittävä hyvinvoinnin käsitteestä, jolloin on mahdollista sitoa yhteen politiikan arvoalinnoista johdetut ohjelmat, hallinnon ja toteuttajien tekemät toimet ja ohjelmien kohderyhmän (kansalaisten) palvelukokemukset.

Lähteet

- Anderson, A. 2005. An Introduction to Theory of Change. The Evaluation Exchange XI.
- Chelimsky, E. 1987. What have we learnt about the politics of program evaluation. *Educational Evaluation and Policy Analysis* 9 (3), 199–214.
- Datnow, A. 2000. Power and politics in the adoption of school reform models. *Educational Evaluation and Policy Analysis* 22 (4), 357–374
- Fitz-Gibbon, C. 1996. *Monitoring education*. London – New York: Cassell.
- Gonand, F., Joumard, I. & Price, R. 2007. Public spending efficiency: Institutional indicators in primary and secondary education. OECD Economics Department Working Papers 543. OECD Publishing.
- Heyneman, S. 2000. From the party/state to multiethnic democracy: Education and social cohesion in Europe and Central Asia. *Educational Evaluation and Policy Analysis* 22 (2), 173–191.
- Keeves, J., Hungi, N. & Afrasssa, T. 2005. Measuring value added effects across schools. Should schools be compared in performance. *Studies in Educational Evaluation* 31, 2–3, 247–266.
- Kennedy, M. 1999. Approximations to indicators of student outcomes. *Educational Evaluation and Policy Analysis* 21 (4), 345–363.
- Lehtisalo, L. & Raivola, R. 1999. *Koulutus ja koulutuspolitiikka 2000-luvulle*. Helsinki: WSOY.
- Levin, H. 1988. Cost-effectiveness and educational policy. *Educational Evaluation and Policy Analysis* 10 (1), 51–69.
- Ludwig, J. & Bassi, L. 1999. The puzzling case of school resources and student achievement. *Educational Evaluation and Policy Analysis* 21 (4), 385–403.
- McLaughlin, M. 1987. Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis* 9 (2), 171–178.
- Penna, R. & Phillips, W. 2005. Eight outcome models. The Evaluation Exchange XI, 2.
- Raivola, R. 2000. *Tehoa vai laatua koulutukseen*. Helsinki: WSOY.
- Rice, J. 1997. Cost analysis in education: Paradox and possibility. *Educational Evaluation and Policy Analysis* 19 (4), 309–317.
- Stout, K. & Stevens, B. 2000. The case of the failed diversity rule: A multiple streams analysis. *Educational Evaluation and Policy Analysis* 22 (4), 341–355.



Koulutuksen talouden arvioinnin menetelmiä

Monilla koulutusjärjestelmään liittyvillä päätöksillä on muiden vaikutusten lisäksi taloudellisia vaikutuksia. Ne saattavat esimerkiksi lisätä oppilaitosten tai niiden ylläpitäjien kustannuksia. Tutkintojen sisältöihin ja rakenteeseen tehdyillä muutoksilla voi olla vaikutuksia myös yksilön työmarkkina-asemaan, palkkaan ja viime kädessä koko yhteiskunnan hyvinvointiin. Arvioitaessa koulutuksen taloudellisia vaikutuksia ollaan kiinnostuneita siitä, mitkä ovat koulutuksen taloudelliset hyödyt toisaalta yksilölle ja toisaalta yhteiskunnalle suhteessa siihen sijoitettuihin voimavaroihin. Tämän lisäksi koulutuksen taloudessa ollaan myös kiinnostuneita itse prosessin taloudellisuudesta eli esimerkiksi siitä, miten resurssit vaikuttavat koulutusprosessin onnistumiseen tai kuinka tehokkaasti ja tuottavasti prosessi toimii. Yksi vaihtoehto arvioinnissa on ottaa lähtökohdaksi ja koulutuksen tulokseksi koulutuksen avulla hankittu työmarkkina-asema. Yleisempää ja menetelmällisesti yksinkertaisempaa on kuitenkin tarkastella koulutusprosessin tuloksia lähempänä itse koulutustapahtumaa. Tällöin selvitetään sitä, mikä vaikutus resursseilla on oppimiseen, jota mitataan yleensä jollakin standardoidulla testillä. Tämän lisäksi voidaan tarkastella itse koulutusprosessin tehokkuutta. Tässä artikkelissa tarkastellaan koulutuksen prosessien taloudellista arviointia ja siinä käytettäviä yleisimpiä menetelmiä nimenomaan tästä jälkimmäisestä näkökulmasta käsin. Lopussa keskustellaan myös taloudellisen arvioinnin tarvitsemista yksilöpohjaisista seuranta-aineistoista, joita suomalaisesta kou-

lujärjestelmästä ei ole toistaiseksi juuri ollut saatavilla. Artikkelin esimerkit kuvaavat perusopetusta ja lukiokoulutusta, mutta samojen käsitteiden ja menetelmien avulla on mahdollista tarkastella myös muita kouluasteita.

Asiasanat: Koulutuksen taloudellisuuden arviointi, tehokkuus, tuottavuus

Koulutuksen tuotantofunktio

Onko parempi opettaa lapsia pienemmissä opetusryhmissä oppimistulosten parantamiseksi? Entä onko opettajan koulutuksella merkitystä oppimiselle ja kuinka paljon opettajissa on lopulta eroja, jotka heijastuvat oppilaiden suoritustasoon? Kaikilla näillä koulutuksen prosesseihin liittyvillä kysymyksillä on myös taloudellista merkitystä. Jos esimerkiksi ryhmäkoolla ei ole mitään vaikutusta vaan lapset oppivat ja sosiaalistuvat samalla tavalla niin suurissa kuin pienissäkin ryhmissä, olisi luonnollista suosia suurempia ryhmiä ja käyttää säästyvät rahat jollakin hyödyllisemmällä tavalla.

Koulutuksen prosessien arviointi talouden näkökulmasta perustuu tuotanto- ja kustannusfunktioiden käsitteisiin. Tuotantofunktiolla kuvataan taloustieteissä tuotantoteknologian fyysisiä ja teknisiä riippuvuussuhteita. Se ilmaisee eri panosmäärillä ja teknologian tasolla saavutettavan maksimaalisen tuotannon määrän. Koulutuksen tuotantofunktion historia on jäljitettävissä Colemanin raporttiin (Coleman ym. 1966). Raportin jälkeen maailmalla on ilmestynyt satoja tutkimuksia, joissa on selvitetty luokkakoon, koulujen koon ja opettajien ominaisuuksien vaikutusta oppimiseen. Joissakin tutkimuksissa on myös selvitetty näiden tekijöiden vaikutusta opintojen jälkeisiin palkkoihin.¹ Tuotantofunktioiden avulla voidaan arvioida myös koulujen tehokkuutta.

Koulutuksen tuotoksena on näissä tutkimuksissa useimmiten käytetty erilaisia standardoituja testituloksia. Koulutuksen panokset on yleensä jaoteltu opiskelijoiden ominaisuuksia, vanhempien sosioekonomista taustaa, opiskelijoiden ystäväpiiriä ja koulujen resurssipanostusta kuvaaviin tekijöihin (ks. esim. Hanushek 2003). Opiskelijoiden ominaisuuksia ovat motivaatio, kyvykkyys ja aiempi osaaminen. Kun tuotoksena käytetään oppimistuloksia, on tärkeää ottaa huomioon opiskelijoiden aiempi osaaminen. Tavallista onkin, että opiskelijoiden taidot testataan keskeisissä aineissa opintojen alussa ja niiden päättyessä, ja alkuhetken suoriutumista käytetään yhtenä panoksena (ns. value-added -malli). Sosioekonomista taustaa mitataan usein vanhempien koulutuksella, tuloilla, etnisellä taustalla tai muilla vastaavilla tekijöillä. Näiden ohella tuotantofunktios-

¹ Näiden tutkimusten tuloksia ovat tarkastelleet mm. Hanushek (2003), Krueger (2003) ja Hedges ja Greenwald (1996).

sa kuvataan oppilaitosten resurssipanostusta, johon liittyvät opetuksen määrä, opettajat, tilat ja laitteet sekä muut oppilaitosten tarjoamat opetusta tukevat palvelut. Seuraavassa tarkastellaan lähemmin sitä, kuinka resurssien ja erityisesti kustannusten, luokkakoon ja opettajien vaikutusta suoritustasoon voidaan arvioida eri menetelmin ja aineistoin.

Resurssien, lähinnä kustannusten sekä luokkien ja koulujen koon, vaikutusta oppilaiden suoritustasoon arvioidaan tavallisimmin tilastollisesti estimoimalla pienimmän neliösumman (PNS) regressioyhtälöitä ottamalla huomioon oppilaiden taustatekijät. Suurimmassa osassa tutkimuksia on käytetty koulu- tai kuntatason aineistoja. Myös Suomessa toteutetuissa tutkimuksissa on luokkakoon tai kustannusten vaikutusta suoritustasoon muutamaa poikkeusta lukuun ottamatta arvioitu koulutason aineistoilla.

Tilastollisiin regressiomalleihin liittyy joitakin ongelmia. Näitä ovat mallista pois jääneiden muuttujien aiheuttama harha ja endogeenisuus. Endogeenisuudella tarkoitetaan sitä, että käytetty malli selittää jonkun tai joidenkin muuttujien saamia arvoja.² Koulujen välille endogeenisuutta aiheutuu siitä, että korkeammin koulutettujen lasten vanhemmat saattavat valita lapsilleen koulun, jossa luokkakoko on pienempi tai kustannukset korkeammat. Koulujen sisällä endogeenisuutta voi aiheuttaa se, että oppilaat jaetaan luokkiin heidän taitojensa perusteella.

Puuttuvien muuttujien aiheuttamaa harhaa ja endogeenisuutta on pyritty poistamaan usealla eri tavalla. Kokeellisen tutkimusasetelman käyttö on luonnollisesti paras vaihtoehto. Näin tehtiin ainakin Yhdysvalloissa, jossa Project STAR-hankkeessa ala-asteen oppilaat sijoitettiin satunnaisesti luokkakooltaan erilaisiin luokkiin ja tämän jälkeen tutkittiin oppimisen lisäystä erikokoisissa ryhmissä (ks. esim. Mosteller 1995 ja Krueger 1999).

Kokeellisen asetelman luominen on kallista ja työlästä, joten yleensä arvioinnissa ja tutkimuksessa käytetään rekisteripohjaista aineistoa. Tällöin ei-satunnaisesta valikoitumisesta aiheutuvaa harhaa voidaan yrittää poistaa esimerkiksi käyttämällä instrumenttimuuttujamenetelmää. Siinä perusideana on löytää muuttuja tai muuttujia, joiden vaihtelu on endogeenisten muuttujien kanssa samanlaista, mutta poikkeaa mallin virhetermin vaihtelusta. Luokkakoon vaikutusta tutkittaessa on joissakin tutkimuksissa instrumenttimuuttujana käytetty luokkakoolle asetettua lainsäädännöllistä ylärajaa (ks. esim. Angrist ja Lavy 1999 ja Hoxby 2000). Menetelmän ongelmana on helposti se, että tarpeeksi hyvää instrumenttia voi olla vaikea löytää. Analyysin validiteetti on kuitenkin täysin riippuvainen siitä.

Mallista puuttuvien muuttujien aiheuttamaa harhaa on mahdollista vähentää myös yksilöpohjaisella paneeliaineistolla, jossa on tietoa samoista kouluista ja

² Tavallisesti tilastollisissa malleissa oletetaan, että muuttujat ovat toisistaan ja mallin virhetermistä riippumattomia eli eksogeenisiä.

heidän oppilaista usealta vuodelta. Oppilaiden suoritustasoon on yhdistetty tietoa heidän taustastaan. Lisäksi mukana on tietoa heitä opettaneista opettajista ja koulujen resursseista. Tilastollisilla menetelmillä voidaan tällaisesta aineistosta erotella ajassa muuttumattomat koulukohtaiset vaikutukset, joita mallin muut muuttujat eivät mittaa. Näitä koulukohtaisia vaikutuksia nimitetään kiinteiksi vaikutuksiksi ja ne pienentävät mallista puuttuvista muuttujista aiheutuvaa harhaa.

Suomessa on tällä tavoin arvioitu lukioiden kustannusten vaikutusta opiskelijoiden ylioppilaskirjoitusmenestykseen (Häkkinen ym. 2003). Tässä tutkimuksessa käytettiin yksilötason aineistoa, jossa oli tietoa opiskelijoiden ylioppilaskirjoitusmenestyksestä, peruskoulun päättötodistuksen lukuaineiden keskiarvosta, työssä käynnistä ja vanhempien sosioekonomisesta taustasta vuosilta 1990–1998. Lisäksi mukana oli tietoa lukioiden kustannuksista samalta ajanjaksolta. Tutkimuksen tulosten mukaan kustannuksilla ei ollut vaikutusta ylioppilaskirjoitusmenestykseen siitäkään huolimatta, että ajassa pysyvät kiinteät vaikutukset otettiin malleissa huomioon. Eniten tuloksiin vaikutti aiempi koulumenestys ja vanhempien sosioekonominen tausta. Työssäkäynti heikensi menestystä ylioppilaskirjoituksissa.

Opettajien vaikutusta oppimiseen kuvataan tuotantofunktion avulla käyttämällä selittävinä tekijöinä erityisesti sellaisia opettajien ominaisuuksia kuin koulutus, kokemus ja testimenestys. Tavallisimmin tällainen tuotantofunktio estimoidaan koulutason aineistolla käyttämällä koulutason keskiarvoja. Myös tässä tarkastelutavassa on ongelmansa. Tutkimuksissa on esimerkiksi näyttöä siitä, että opettajat valitsevat työpaikkansa siten, että pidempään opettajana toimineet hakeutuvat korkeamman suoritustason kouluihin, joten malleissa on valikoitumisesta aiheutuvaa harhaa. Tämän lisäksi arvioijalla on yleensä käytettävissään vain pieni määrä opettajien ominaisuuksia kuvaavia muuttujia. Varsin tavallista on myös, että opettajia kuvaavat muuttujat eivät lopulta olekaan tilastollisessa yhteydessä oppilaiden suoritustasoon. Yksi syy siihen on, että niissä ei ole riittävästi vaihtelua.³

Yksi keino välttää osa edellä kuvatuista ongelmista on tarkastella opettajien ns. kokonaisvaikutusta. Lähestymistapa perustuu oppimisen lisäyksen mallintamiseen opettajien välillä. Tällaisen tarkastelutavan perusteella hyvä opettaja saavuttaa jatkuvasti suuren lisäyksen oppimisessa ja huono opettaja vastaavasti pienen lisäyksen. Menetelmällisesti opettajan kokonaisvaikutus identifioidaan malleissa opettajakohtaisen kiinteän vaikutuksen avulla. Menetelmän käyttö edellyttää aineistoa, jossa opettajia ja heidän oppilaitaan on seurattu useita vuo-

³ Esimerkkinä voidaan mainita opettajien koulutuksen mittaaminen korkeakoulututkinnon suorittaneiden osuudella. Suomessa suurin osa opettajista on suorittanut korkeakoulututkinnon, joten muuttujan vaihtelu on hyvin pientä.

sia. Tällaisia aineistoja ei Suomessa ole toistaiseksi saatavilla. Yleensä arvoinnit on toteutettu siten, että tietoja kyllä kerätään myös opettajista ja heidän toiminnastaan, mutta ainoastaan koulutasolla yhdistämättä tietoja opettajista ja opiskelijoiden oppimistuloksista.

Opettajan vaikutusta ei pidä aliarvioida, sillä Yhdysvalloissa tehdyissä tutkimuksissa se on osoittautunut merkittäväksi. Hyvän opettajan oppilaat olivat opinnoissaan edellä keskimäärin vuoden verrattuna huonojen opettajien oppilaisiin (Hanushek 1992). Konservatiivisemmankin arvion mukaan, jos oppilaita opettaa keskiverta opettajan sijasta opettaja, jonka taidot ovat paremmat kuin 85 prosentilla opettajia (mitattuna oppilaiden oppimisen lisäyksellä), on vaikutus samansuuruinen kuin luokkakoon pienentäminen 10 oppilaalla (Hanushek ja Rivkin 2006). Huomattavaa on, että eroa oppimisessa ei pystytä selittämään vaihtelulla esimerkiksi opettajien kokemuksessa tai koulutuksessa. Opettajakohtaisen kiinteän vaikutuksen avulla on myös mahdollista identifioida esimerkiksi jatkuvasti hyvään tulokseen ylttäviä opettajia ja tämän jälkeen jollakin toisella tutkimusotteella selvittää, kuinka he toimivat luokassa oppilaidensa kanssa.

Koulutuksen kustannusfunktio

Resurssien ja suoritustason välisen yhteyden lisäksi taloustieteilijät ovat kiinnostuneita koulutuksen järjestäjien välisiin kustannusvaihteluihin vaikuttavista tekijöistä. Nämä tarkastelut perustuvat kustannusfunktion käsitteeseen, jolla kuvataan panoshintojen ja tuotusmäärien yhteyttä kustannuksiin. Kustannusfunktio liittyy läheisesti tuotantofunktion käsitteeseen. Siinä tarvitaan tuotantofunktiossa käytettävän tiedon lisäksi tietoa arvioitavien yksiköiden panoshinnoista. Koska opetus on hyvin työvoimaintensiivistä toimintaa, tärkeimpiä panoshintoja koulutuksessa ovat opettajien ja muun henkilökunnan palkat. Näiden lisäksi kustannuksia aiheutuu hallinnosta, kiinteistöjen ylläpidosta, ruokailusta, oppilashuollosta ja koulukuljetuksista.

Kustannusfunktioiden avulla siis selitetään koulujen tai koulutuksen järjestäjien välisiä kustannusvaihteluita. Niiden avulla voidaan arvioida joidenkin mitakaavatekijöiden kuten luokkakoon, koulujen koon tai koulutuksen järjestäjän koon vaikutuksia kustannuksiin silloin, kun oppilaiden suoritustaso, sosioekonominen tausta ja panoshinnat on otettu huomioon. Siten kustannusfunktioiden avulla saadaan arvioinnin kannalta hyödyllistä tietoa, joka palvelee kouluja ja koulutuksen järjestäjiä heidän suunnitellessaan resurssien jakoa. Suomessa kustannusfunktioiden avulla on selvitetty perusopetuksen kustannuksiin vaikuttavia tekijöitä (Aaltonen ym. 2006) ja lukioiden opetuksen menojen vaihtelua (Kirjavainen 2007).

Tehokkuuden mittauksen menetelmiä

Edellä kuvattujen tuotanto- ja kustannusfunktioiden avulla on myös mahdollista vertailla kouluja tai koulutuksen järjestäjiä tehokkuuden näkökulmasta. Tuottavuudella tällaisessa tarkastelussa tarkoitetaan tuotoksen suhdetta panoksiin. Kysymys on toisin sanoen siitä, että tuotoksen mittalukua käytetään osoittajana ja panosten mittalukua nimittäjänä. Tehokkuus voidaan laskea suhteuttamalla kunkin yksikön tuottavuus kaikkein tuottavimman yksikön tai yksiköiden tuottavuuteen. Tehokkaan yksikön tuotos on maksimaalinen käytettävissä olevilla panoksilla. Tehokkaimmat yksiköt muodostavat ns. tehokkuusrintaman⁴, jonka alapuolella muut, tehottomat, yksiköt ovat. Laskemalla kaikkien yksiköiden keskimääräinen tehokkuus, voidaan arvioida keskimääräistä tehostamismahdollisuutta.

Tehokkuutta voidaan kuvata eri näkökulmista. Teknisellä tehokkuudella tarkoitetaan sitä, että annetulla panosmäärällä valmistetaan mahdollisimman suuri tuotos tai annettu tuotospäästö valmistetaan mahdollisimman pienin panoksin.

Kustannustehokkuudella tarkoitetaan sen sijaan tarkasteltavan yksikön kykyä tuottaa tuotos mahdollisimman alhaisin kustannuksin. Kustannustehokkuus voidaan jakaa kahteen osaan: tekniseen tehokkuuteen ja panosten käytön allokaatiiviseen tehokkuuteen. Panosten käytön allokaatiivinen tehokkuus kuvaa, kuinka lähellä panosten käytön rakenne on optimaalista ottaen huomioon panosten hinnat. Ollakseen kustannustehokas yksikön on oltava teknisesti tehokas ja käytettävä kustannukset minimoivaa panossuhdetta.

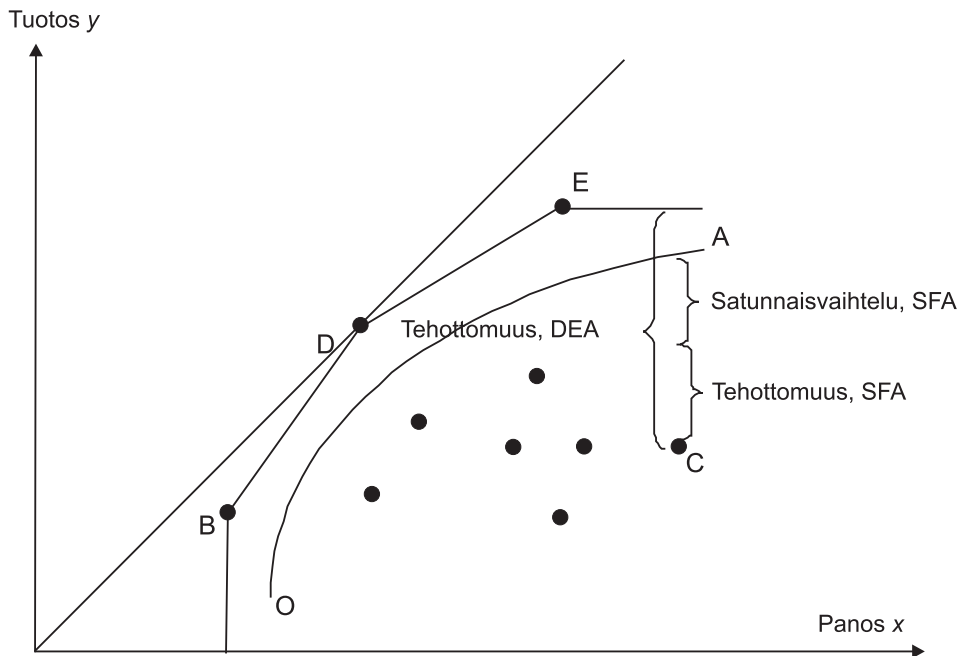
Tuottavuuden ja tehokkuuden mittaamisen menetelmät voidaan jakaa tilastollisiin ja lineaariseen optimointiin perustuviin menetelmiin.⁵ Seuraavassa kuvaillaan lyhyesti niistä tärkeimpiä, tulosten tulkintaa ja kuhunkin menetelmään liittyviä etuja ja haittoja. Tehokkuuden mittauksen menetelmiä ovat kuvanneet laajemmin mm. Coelli ym. (1999) sekä koulutukseen liittyen Johnes (2004), Aaltonen ym. (2007) ja Kangasharju ym. (2007).

⁴ Riippuen menetelmästä sitä kutsutaan myös tuotanto- tai kustannusrintamaksi.

⁵ Lineaarinen optimointi on matemaattinen menetelmä, jonka avulla etsitään lineaarisen tavoitefunktion suurinta tai pienintä arvoa tiettyjen rajoitteiden ollessa voimassa. Lineaarisen optimointiongelman tavoitefunktio voi esittää esimerkiksi minimoitavia tuotantokustannuksia.

Stokastinen rintama-analyysi

Stokastinen rintama-analyysi (Stochastic Frontier Analysis, SFA) on tilastollinen menetelmä tehokkuuden arviointiin.⁶ Kunkin tarkasteltavan yksikön tehottomuus lasketaan siinä tuotanto- tai kustannusfunktiolla estimoidun mallin virhetermistä jakamalla se kahteen osaan: satunnaisvaihteluun ja tehottomuutta kuvaavaan virhetermiin (ks. kuvio 1). Tehottomuutta kuvaavan virhetermin avulla yksiköille lasketaan tehottomuusluku, joka saa arvoja 0:n ja 1:n välillä (tai vaihtoehtoisesti 0–100 prosenttia). Tehottomuusluvun suuruus kuvaa yksikön etäisyyttä tehokkuusrintamasta. Kuviossa 1, jossa x-akselilla on panoksen x määrä ja y-akselilla tuotoksen y määrä, stokastisen rintama-analyysin määrittämää tehokkuusrintamaa ja samalla tehokasta tuotantotapaa kuvaa käyrä OA. Koulun C tehottomuus lasketaan etäisyydestä AC, josta osa on satunnaisvaihtelua ja osa tehottomuutta. Mitä lähempänä yksikkö on tehokkuusrintamaa, sitä lähempänä nollaa tehottomuusluvun arvo on, ja mitä lähempänä luku on ykköistä (100 pro-



Kuvio 1. Tehottomuuden määrittely stokastisella rintama-analyysillä ja Data Envelopment-analyysillä

⁶ Kattava esitys stokastisesta rintama-analyysistä löytyy teoksesta Kumbhakar & Lovell (2000). Paneeliaineistojen käyttöä stokastisessa rintama-analyysissä on tarkastellut Greene (2005a ja 2005b).

senttia), sitä kauempana yksikkö on tehokkuusrintamasta ja sitä tehottomampaa yksikön toiminta on.

Koska stokastinen rintama-analyysi on tilastollinen menetelmä, saadaan sen avulla tehokkuuden lisäksi tietoa tuotosten ja panosten välisestä tilastollisesta riippuvuudesta, kun arvioidaan teknistä tehokkuutta. Samoin saadaan tietoa panoshintojen ja tuotosten vaikutuksista kustannuksiin arvioitaessa kustannustehokkuutta. Arvioinnin kannalta tällainen tieto on hyödyllistä. Riippuen käytetyistä muuttujista ja malleista stokastisen rintama-analyysin kerroinestimaattien avulla voidaan kuvata esimerkiksi resurssien vaikutusta opiskelijoiden suoritus-tasoon tai koulujen koon tai koulutuksen järjestäjän koon vaikutusta kustannuksiin. Jos malleissa mukana olevat muut muuttujat kuvaavat riittäväällä tavalla tarkasteltavan yksikön toimintaa, on mallin tulosten perusteella mahdollista tehdä laskelmia esimerkiksi koulujen keskikoon muutoksen kustannusvaikutuksista. Mallit ovat myös viime vuosina kehittyneet niin, että niiden avulla voidaan ottaa huomioon tehottomuuden lisäksi myös ajassa pysyviä satunnaisia tai kiinteitä vaikutuksia käytettäessä paneeliaineistoja (ks. esim. Greene 2005a ja 2005b). Suomessa stokastista rintama-analyysia ovat käyttäneet Aaltonen ym. (2006) arvioidessaan perusopetuksen tehokkuutta ja Kirjavainen (2007) arvioidessaan nuorten lukiokoulutuksen tehokkuutta.

Taulukossa 1 on esimerkki stokastisella rintama-analyysilla lasketuista kustannustehottomuusluvusta, joita verrataan opetuksen opiskelijakohtaisiin kustannuksiin. Luvut kuvaavat lukioiden prosentuaalista kustannustehottomuutta (Kirjavainen 2007). Ne on laskettu kustannusfunktioimallilla, jossa seuraavat tekijät selittävät kunnallisten lukioiden opetuksen opiskelijakohtaisia menoja: opetushenkilökunnan palkat, ylioppilaskirjoitusmenestys, opiskelijoiden peruskoulumenestys, vanhempien sosioekonominen tausta, ruotsinkielisten opiskelijoiden osuus, opiskelijoiden ja opetushenkilökunnan välinen suhdeluku, lukioiden keskimääräinen koko, sijainti ja kunnan verotettavat tulot asukasta kohti. Tarkastelu on tehty viiden vuoden paneeliaineistolla (vuodet 2000–2004), ja siinä on otettu huomioon kuntien ajassa pysyvät kiinteät vaikutukset. Niitä voivat olla esimerkiksi mallissa kontrolloimattomat epäsuotuisat olosuhteet tai jokin ajassa muuttumaton toimintatapa.⁷ Tehottomuuslukua tulkitaan siten, että mitä lähempänä se on nollaa, sitä tehokkaampi kunta on.

Taulukosta 1 nähdään, että opetuksen yksikkökustannuksiltaan halvimmissa kunnissa opetuksen kustannustehokkuus voisi olla vielä parempi. Esimerkiksi Kalajoen kunnassa lukion opetus olisi voitu järjestää noin 5 prosenttia alhaisemmin kustannuksin. Sen sijaan kaksi kalleinta kuntaa, Utsjoki ja Savukoski, ovat kustannustehokkaimpien kuntien joukossa. Toisin sanoen näissä kunnissa

⁷ Tarkempi selostus käytetyistä malleista ks. Kirjavainen (2007).

Taulukko 1. *Esimerkki kustannustehottomuuden ja opiskelijakohtaisten opetuksen kustannusten välisestä yhteydestä lukiokoulutuksessa vuonna 2004*

Kymmenen kustannuksiltaan halvinta kuntaa v. 2004 (kolmen edeltävän vuoden keskiarvo, vuoden 2003 hinnoin)			Kymmenen kustannuksiltaan kalleinta kuntaa v. 2004 (kolmen edeltävän vuoden keskiarvo, vuoden 2003 hinnoin)		
Kunnan nimi	Opetuksen kustannukset / opiskelija	Kustannustehottomuus (%) v. 2004	Kunnan nimi	Opetuksen kustannukset / opiskelija	Kustannustehottomuus (%) v. 2004
Kerava	2 350	1,60	Maalahti	5 539	3,11
Huittinen	2 453	2,61	Puumala	5 631	1,72
Turku	2 546	2,56	Valtimo	5 734	3,32
Loimaa	2 709	2,23	Kärsämäki	5 754	4,32
Kalajoki	2 735	5,12	Vuolijoki	5 990	5,29
Vantaa	2 757	2,02	Kittilä	6 581	6,28
Siilinjärvi	2 820	6,81	Enontekiö	6 985	3,23
Hämeenlinna	2 827	3,66	Muonio	7 124	1,58
Joensuu	2 849	4,07	Savukoski	9 018	0,71
Ylivieska	2 849	3,88	Utsjoki	9 231	0,56

lukio-opetuksen järjestelyt ovat olosuhteisiin, opiskelijoiden taustoihin ja tuloksiin nähden tehokkuusmielessä hyvät.

Taulukossa 2 on puolestaan esitetty perusopetuksen kustannustehottomuus niissä 10 kunnassa, joissa opiskelijakohtaiset käyttömenot olivat vuonna 2004 alhaisimmat ja korkeimmat. Kustannustehottomuus on laskettu stokastista rintama-analyysia käyttäen siten, että opiskelijakohtaisia käyttömenoja on selitetty

Taulukko 2. *Esimerkki kustannustehottomuuden ja opiskelijakohtaisten käyttömenojen välisestä yhteydestä perusopetuksessa vuonna 2004*

Kymmenen kustannuksiltaan halvinta kuntaa v. 2004			Kymmenen kustannuksiltaan kalleinta kuntaa v. 2004		
Kunnan nimi	Käyttömenot / opiskelija	Kustannustehottomuus (%) v. 2004	Kunnan nimi	Käyttömenot / opiskelija	Kustannustehottomuus (%) v. 2004
Kerava	4 304	4,0	Maalahti	7 944	20,0
Huittinen	4 348	4,0	Puumala	8 178	15,0
Turku	4 351	5,0	Valtimo	8 291	29,0
Loimaa	4 373	3,0	Kärsämäki	8 964	23,0
Kalajoki	4 417	4,0	Vuolijoki	9 080	28,0
Vantaa	4 421	6,0	Kittilä	9 181	15,0
Siilinjärvi	4 425	1,0	Enontekiö	10 154	7,0
Hämeenlinna	4 446	3,0	Muonio	10 715	22,0
Joensuu	4 456	6,0	Savukoski	12 049	39,0
Ylivieska	4 488	5,0	Utsjoki	13 385	2,0

opetus- ja muun henkilökunnan keskipalkalla, perusopetuksen päättötodistuksen keskiarvolla, jatko-opintoihin siirtyneiden prosentuaalisella osuudella, erityisoppilaiden osuudella, muiden kuin suomenkielisten oppilaiden osuudella, kuljetusta saavien oppilaiden osuudella, kunnan oppilasmäärällä, keskimääräisellä koulujen koolla, koulutustasolla, verotettavilla tuloilla ja joillakin muilla kunnan ominaisuuksiin liittyvillä tekijöillä. Nämä kustannustehottomuusluvut on laskettu vuoden 2004 poikkileikkausaineistolla, joten niissä ei ole otettu huomioon mallissa kontrolloimattomia ajassa pysyviä vaikutuksia. Todennäköisesti sen vuoksi joidenkin Lapin kuntien kustannustehottomuus on selvästi korkeampaa kuin lukiokoulutuksessa. Myös perusopetuksen osalta voidaan kuitenkin todeta, että kustannuksiltaan alhaisimmissakin kunnissa on usein tehostamisvara useita prosentteja. Taulukosta nähdään myös, että kustannuksiltaan kalleimmassa kunnassa Utsjoella perusopetus olisi voitu toteuttaa vain kaksi prosenttia alhaisemmin kustannuksin.

Data Envelopment -analyysi, DEA

Data Envelopment -analyysi (DEA) on lineaariseen optimointiin perustuva tehokkuuden mittaamisessa kehitetty menetelmä. Sen käyttö on ollut suosittua julkisten palveluiden tehokkuuden arvioinnissa. Sitä on käytetty paljon myös koulutuksen tehokkuuden arvioinnissa sekä Suomessa että ulkomailla (ks. esim. Worthington 2001 ja Aaltonen ym. 2005).

DEA:ssa tavoitteena on maksimoida yksikön tehokkuus vertailuryhmän sisällä siten, että verrattavista yksiköistä muodostetaan lineaarikombinaationa ns. tehokkuusrintama. Kuviossa 1 tehokkuusrintaman muodostaa pisteiden B, D ja E kautta kulkeva käyrä (kun oletetaan muuttuvat skaalatuotot⁸). Menetelmässä kaikkien analysoitavien yksiköiden tulee sijaita joko tehokkuusrintamalla tai sen sisäpuolella. Tehokkuusrintamalla sijaitsevat yksiköiden tehottomuus on nolla, ja muiden yksiköiden tehottomuus määräytyy niiden suhteellisesta etäisyydestä tehokkuusrintamaan nähden. Mitä kauempana ne sijaitsevat rintamasta, sitä lähempänä ykköstä ovat niiden tehottomuusluvut. Kuviossa 1 koulun C tehottomuus määräytyy sen etäisyyden perusteella tehokkuusrintamasta. Toisin kuin stokastisessa rintama-analyysissä, DEA:ssa yksikön poikkeama tehokkuusrintamasta tulkitaan kokonaisuudessaan tehottomuudeksi. DEA sopii sekä teknisen

⁸ Tuotannon mittakaavaetuja kuvaavat joko vakioiset tai muuttuvat skaalatuotot. Vakioisilla skaalatuotoilla panoksen määrän kaksinkertaistaminen kaksinkertaistaa myös tuotoksen määrän. Muuttuvat skaalatuotot ovat joko nousevia tai laskevia. Jos skaalatuotot ovat laskevia, panoksen määrän kaksinkertaistuksessa tuotoksen määrä ei kaksinkertaistu. Nousevilla skaalatuotoilla tuotoksen määrä enemmän kuin kaksinkertaistuu.

tehokkuuden että kustannustehokkuuden arviointiin. Sen avulla voidaan myös ottaa huomioon mittakaavavaikutuksia.

Menetelmien vertailua

Stokastisen rintama-analyysin etuna erityisesti verrattuna DEA:iin voidaan pitää sitä, että se erottelee satunnaisvaihtelun tehottomuudesta ja mahdollistaa tilastollisen päättelyn. DEA:ssa jokaisen yksikön poikkeama tehokkuusrintamasta tulkitaan kokonaisuudessaan tehottomuudeksi. Stokastisessa rintama-analyysissa otetaan huomioon se, että osa tästä poikkeamasta voi olla satunnaisvaihtelua. Tilastollisen menetelmän etuihin kuuluu myös, että se on vähemmän herkkä aineistossa oleville virheille. Stokastisen rintama-analyysin erottelukyky on myös parempi kuin DEA:n, sillä jos DEA:ssa on täysin tehokkaita yksiköitä enemmän kuin yksi, menetelmä ei pysty tekemään eroa näiden välille. Stokastisessa rintama-analyysissa ei täysin tehokkaita yksiköitä ole välttämättä ollenkaan.

Stokastisen rintama-analyysin yksi ongelma on, että tehottomuudelle joudutaan oletamaan jokin jakauma, eikä sen valinta perustu talusteoriaan. Jakumaoletus vaikuttaa havaittavan tehokkuuden suuruuteen, joten väärin spesifioituna sillä on merkitystä jokaiselle arvioitavalle yksikölle. Myös tuotantofunktion muodosta eli tuotantoteknologiasta joudutaan tekemään etukäteen oletus, jolle ei ole talusteoreettisia perusteluja.

DEA:n etuna vaihtoehtoihin menetelmiin nähden pidetään sen joustavuutta. Sen avulla tutkimuksen kohteena olevia yksiköitä on mahdollista arvioida, vaikka tiedossa ei olisi panos- ja tuotoshintoja. Menetelmää on myös yksinkertaista soveltaa silloin, kun yksiköt tuottavat useita tuotoksia ja käyttävät useita panoksia. Lisäksi useiden tuotosten välille ei välttämättä tarvitse määrittää niiden suhteellisen tärkeyden osoittavaa painokerrointa,⁹ vaan laskentamalli tuottaa kullekin yksikölle automaattisesti sille edullisimmat painot. Myöskään oletusta panosten ja tuotosten välillä vallitsevasta tuotantoteknologiasta ei tarvitse tehdä. Riittää, että panosten ja tuotosten välillä on riippuvuussuhde. Yksi DEA:n ongelma on, että panos- ja tuotosmuuttujien määrän noustessa tehokkaiden yksiköiden määrä nousee. Tästä seuraa se, että menetelmän erottelukyky heikkenee. DEA on myös herkkä aineistovirheille.

Molemmilla menetelmillä saadaan hyödyllistä tietoa koulutuksen tehokkuudesta niin kansalliseen päätöksentekoon kuin oppilaitosten ja koulutuksen järjestäjien käyttöön. Kansallisella tasolla menetelmät tuottavat tietoa keskimääräisestä tehottomuudesta ja tehottomuuseroista. Kouluille ja koulutuksen järjestäjille menetelmät antavat vertailutietoa omasta tehottomuudesta suhteessa muiden

⁹ Tarvittaessa painokertoimet voidaan määrittää.

yksiköiden tehottomuuteen. Näiden lisäksi, kuten jo edellä mainittiin, stokastisen rintama-analyysin avulla saadaan hyödyllistä arviointitietoa eri tekijöiden välisistä riippuvuussuhteista. Tätä tietoa voivat käyttää hyväksi päätöksentekijät sekä kunnissa että ministeriöissä. Lisäksi voidaan vielä todeta, että tehokkuuden tarkastelu on monipuolisempaa kuin esimerkiksi pelkkä yksikkökustannuserojen tai oppimistulosten vertailu, koska siinä otetaan huomioon molemmat tekijät samanaikaisesti.

Kaksivaiheinen lähestymistapa: DEA ja tilastollinen selitysmalli

Osa koulujen panoksista on luonteeltaan sellaisia, että esimerkiksi koulu ei voi niihin vaikuttaa mutta ne vaikuttavat kuitenkin toimintaan ja esimerkiksi resursien käyttöön. Yksi tällainen tekijä on opiskelijoiden sosioekonominen tausta. Tehokkuutta arvioitaessa tällaisten tekijöiden huomioimisessa on kaksi tapaa. Ensimmäisessä vaihtoehdossa kaikki panokset ovat mukana tehokkuusanalyysissa. Ongelmaksi saattaa tällöin nousta se, että niiden koulujen tehokkuus saatetaan aliarvioida, jotka toimivat epäedullisemmissä olosuhteissa. Vaihtoehtoisesti voidaan käyttää kaksivaiheista lähestymistapaa siten, että koulujen tehokkuusluvut lasketaan DEA:lla (tai muulla menetelmällä) käyttämällä niitä panoksia, joihin koulu voi toiminnallaan vaikuttaa. Tämän jälkeen toisessa vaiheessa näitä tehokkuuslukuja selitetään tilastollisesti tekijöillä, joihin koulu ei itse voi vaikuttaa. Kaksivaiheinen lähestymistapa on ollut melko yleinen koulutuksen tehokkuutta arvioitaessa. Suomessa sitä ovat soveltaneet Kirjavainen ja Loikkanen (1998) arvioidessaan lukioiden tehokkuutta, Aaltonen ym. (2005, 2006) perusopetuksen tehokkuuden arvioinnissa ja Ollikainen ammatillisen peruskoulutuksen tehokkuuden arvioinnissa. Tätä lähestymistapaa on tarkastellut mm. Johnes (2004).

Monitasomallinnus ja tehokkuus

Monitasomallinnus ja yksilöpohjaisten aineistojen käyttö on yleistynyt viime vuosikymmeninä kasvatus- ja taloustieteissä.¹⁰ Edellä kuvatuissa menetelmissä tarkastelu tehdään yleensä aineistoilla, joissa käytetään joko koulu- tai kuntatason tietoja.¹¹ Koulu- tai kuntatason keskiarvoihin perustuviin aineistoihin liittyy

¹⁰ Perusteoksia tilastollisesta monitasomallinnuksesta ovat esimerkiksi Goldstein (2003) ja Raudenbush ja Bryk (2002).

¹¹ DEA:lla on viime vuosina tehty myös muutamia tarkasteluja yksilöaineistolla (ks. esim. Portela ja Thanassoulis 2001 ja Johnes 2006).

kuitenkin muutamia ongelmia, joiden vuoksi yksilöpohjaisten aineistojen käyttö olisi myös tehokkuutta arvioitaessa perustellumpaa.

Ensinnäkin koulu- tai järjestäjätason aineisto ei ota huomioon yksilöiden välistä vaihtelua. Tästä johtuen analyysissa ei oteta huomioon esimerkiksi sitä, että jossakin koulussa opinnot heikoilla lähtötiedoilla aloittavat oppilaat oppivat selvästi enemmän kuin jossakin toisessa koulussa. Näissä kahdessa koulussa on toisin sanoen toisistaan poikkeavat regressiosuoran kertoimet, kun selitetään opiskelun päättövaiheen opintomenestystä opintojen aloitusvaiheen opintomenestyksellä. Ideaalitapauksessa tällainen tulisi ottaa huomioon myös tehokkuusmittauksessa.

Toiseksi, tehokkuuden mittaaminen regressiopohjaisilla malleilla perustuu siihen, että jäännöstermit tulkitaan tehottomuudeksi. Koulu- tai kuntatason aineistossa jäännöstermi on kuitenkin herkkä mallissa mukana oleville muuttujille. Sen vuoksi tehokkuusjärjestys voi muuttua paljonkin mallien muuttujien vaihdellessa. Yksilöaineistoihin perustuvat tulokset ovat tilastollisesti harhattomia ja vähemmän herkkiä muuttujien määrän vaihteluille.

Tilastollisella monitasomallinnuksella tehtävässä tehokkuusanalyysissa malliin lisätään koulu- tai järjestäjäkohtainen satunnaistermi, joka tulkitaan tehottomuudeksi.¹² Monitasomallinnuksen ongelma on, että kuvattaessa monimutkaisia rakenteita myös tilastollisista malleista tulee monimutkaisia. Sen vuoksi tulokset voivat olla vaikeasti ymmärrettävissä. Lisäksi mallien monimutkaisuuden kasvessa ne muuttuvat myös laskennallisesti vaikeammiksi.

Tapaustutkimus

Tilastollinen analyysi on yksi vaihtoehto selvittää tehokkuuteen yhteydessä olevia tekijöitä. Se edellyttää kuitenkin suhteellisen paljon tietoa tarkasteltavista yksiköistä. Tällaisen tiedon saanti voi olla hyvin työlästä ja resursseja vaativaa. Tilastollisen analyysin sijaan tai sen rinnalla voidaan kvalitatiivisen tapaustutkimuksen avulla selvittää, mitä asioita tehokkaat koulut tai koulutuksen järjestäjät tekevät eri tavoin kuin tehottomat koulut tai koulutuksen järjestäjät.

Tapaustutkimuksen tavoitteena on saada syvällisempi käsitys tutkittavan yksikön toiminnasta. Usein tutkimusasetelma on vertaileva: tehokkaimpia kouluja verrataan esimerkiksi tehottomimpiin tai keskinkertaisesti suoriutuviin kouluihin joidenkin toiminnan kannalta keskeisten tekijöiden suhteen. Tutkittavien yksiköiden toimintatapojen yksityiskohtainen analysointi ja kuvailu tuottavat

¹² Tilastollisen monitasomallinnuksen käyttöä tehokkuusanalyysissa on tarkastellut Johnes (2004).

käytännönläheisempää ja konkreettisempaa tietoa ja auttavat hyvin toimivien käytäntöjen levittämistä laajemmalle.

Tapaustutkimuksissa aineistoa kerätään runsaasti vain muutamista yksiköistä haastattelemalla keskeisiä henkilöitä, havainnoimalla toimintaa ja tutustumalla organisaation toiminnasta tehtyihin keskeisiin dokumentteihin. Koulujen osalta tarkastelun kohteena voivat olla opetusjärjestelyt, pedagogiikka, ilmapiiri, johtaminen tai resurssien käyttö.

Tällaista lähestymistapaa ei ole muutamaa poikkeusta lukuun ottamatta sovellettu koulujen tehokkuutta selvittäneissä tutkimuksissa. Kasvatustieteen piirissä on kuitenkin tehty jonkin verran tapaustutkimuksia, joissa tarkastellaan ns. ylisuoriutuvia kouluja. Tätä tutkimustraditiota nimitetään ns. tuloksellisten koulujen tutkimukseksi (effective schools tai school effectiveness research). Myös tehokkuuden arvioinnissa tällaista tutkimusta voidaan pitää esimerkkinä.

Yksi harvoista tehokkuusanalyysin tuloksiin perustuvista tapaustutkimus-esimerkeistä löytyy Englannista. Siellä on ollut käynnissä hanke, jossa arvioitiin yläasteen koulujen tehokkuutta ja tehokkuuteen yhteydessä olleita tekijöitä (Dodd 2006). Tutkimuksessa valittiin tapaustutkimuksen kohteeksi DEA:n tulosten perusteella 38 tehokasta koulua, jotka toimivat ns. verrokkiyksiköinä tehottomammille kouluille. Näistä kouluista kerättiin aineistoa haastattelemalla koulujen johtajia, rajattua määrää opettajia ja joissakin kouluissa myös oppilaiden vanhempia. Lisäksi analysoitiin koulujen opetussuunnitelmia ja resurssien käyttöön liittyviä asiakirjoja.

Lopuksi

Mikään menetelmä ei täysin voi ratkaista arviointiaineistojen puutteellisuuksia. Yksi koulutuksen taloudellisten vaikutusten arvioinnin suurimpia esteitä on Suomessa ollut riittävän tietopohjan puuttuminen. Esimerkiksi koulujen resursseja kuvaavat tiedot ovat toistaiseksi olleet puutteellisia ja aiheuttaneet arviointitiedon luotettavuuteen epävarmuutta. Myös yksilön tietosuojan tulkinta on ollut tiukkaa. Silloinkin, kun rekisteritietoja yhdistelemällä olisi ollut mahdollista koota suhteellisen korkealuokkaisia aineistoja arviointitarkoituksiin, hyvin tiukasti tulkittu yksilön tietosuoja on ollut esteenä tälle työlle.

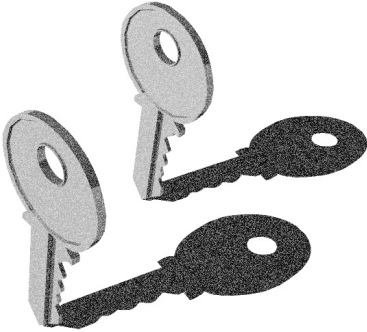
Laajempien yksilötason seuranta-aineistojen kokoaminen yksittäisten otosten sijaan voisi olla yksi keino parantaa myös koulutuksen taloudellisen arvioinnin laatua tulevaisuudessa. Tämä edellyttäisi sitä, että yksittäisten henkilöiden koulutusuria seurattaisiin perusopetuksesta lähtien läpi toisen ja kolmannen asteen koulutuksen aina työelämään asti. Matkan varrelta kerättäisiin laajasti tietoa sekä kouluista että opettajista. Vaihtoehtoisesti tai sen rinnalla seurattaisiin yksittäisten oppilaitosten toimintaa pidemmän ajanjakson ajan keräämällä tietoa

opiskelijoista, heitä opettaneista opettajista sekä oppilaitosten toimintatavoista. Jotta tällaisten aineistojen hyödyntäminen olisi mahdollisimman monipuolista, niiden suunnittelu vaatisi luonnollisesti monitieteellistä otetta niin, että myös muut kuin taloudellisen arvioinnin näkökulmat tulisivat otetuksi huomioon. Lähestymistavan etuna olisi, että taloudellinen arviointi integroituisi luontevasti osaksi kansallista koulutuksen prosessien ja tulosten arviointia.

Lähteet

- Aaltonen, J., Kirjavainen, T. & Moisio, A. 2005. Kuntien perusopetuksen tehokkuuserot ja tuottavuus 1998–2003. Valtion taloudellinen tutkimuskeskus. VATT-keskustelualoitteita 374. Helsinki.
- Aaltonen, J., Kirjavainen, T. & Moisio, A. 2006. Efficiency and productivity in Finnish comprehensive schooling 1998–2004. Valtion taloudellinen tutkimuskeskus. VATT-tutkimuksia 127. Helsinki.
- Aaltonen, J., Kirjavainen, T., Moisio, A. & Ollikainen, V. 2007. Perusopetuksen, lukioiden ja ammatillisen peruskoulutuksen tuottavuus ja tehokkuus. Loppuraportti. Valtion taloudellinen tutkimuskeskus. VATT-tutkimuksia 135. Helsinki.
- Angrist, J. D. & Lavy, V. 1999. Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* 114, 533–575.
- Coelli, T., Rao, P. D. S. & Battese, G. E. 1999. An introduction to efficiency and productivity analyses. Boston: Kluwer academic publishers.
- Coleman, J. S., Campbell, E. Q., Hobson, C. F., McPartland, J., Mood, A. M., Weifeld, F. D., & York, R. L. 1966. Equality of educational opportunity. Washington: U.S. Government Printing Office.
- Cooper, W., Seiford, L. & Tone, K. 2000. Data envelopment analysis. A comprehensive text with models, applications, references and DEA-solver software. Boston: Kluwer Academic Publishers.
- Dodd, A. 2006. Investigating the effective use of resources in secondary schools. Research report to DfES research. Report 779. London: Hedra plc.
- Goldstein, H. 2003. Multilevel statistical models. 3rd edition. London: Edward Arnold.
- Greene, W. 2005a. Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23, 7–32.
- Greene, W. 2005b. Reconsidering heterogeneity in panel data estimators of the stochastic frontier model. *Journal of Econometrics* 126, 269–303.
- Hanushek, E. 1992. The trade-off between child quantity and quality. *Journal of Political Economy* 100, 84–117.
- Hanushek, E. A. 2003. The failure of input-based schooling policies. *The Economic Journal* 113, F64–F98.
- Hanushek, E. & Rivkin, S. 2006. Teacher quality. Teoksessa E. Hanushek & F. Welch (toim.) *Handbook of the economics of education*. Elsevier.
- Hedges, L. & Greenwald, R. 1996. Have times changed? The relation between school resources and student performance. Teoksessa G. Burtless (toim.) *Does money matter? The effect of school resources on student achievement and adult success*. Washington DC: Brookings Institution Press.
- Hoxby, C. M. 2000. The effects of class size and composition of student achievement: New evidence from natural population variation. *Quarterly Journal of Economics* 115, 1239–1285.

- Häkkinen, I., Kirjavainen, T. & Uusitalo, R. 2003. School resources and student achievement revisited: New evidence from panel data. *Economics of Education Review* 22, 329–335.
- Johnes, J. 2004. Efficiency measurement. Teoksessa G. Johnes & J. Johnes (toim.) *International handbook on the economics of education*. Northampton: Edward Elgar Publishing.
- Johnes, J. 2006. Measuring efficiency: A comparison of multilevel modeling and data envelopment analysis in the context of higher education. *Bulletin of Economic Research* 58, 75–104.
- Kangasharju, A., Kirjavainen, T., Luoma, K. & Rätty, T. 2007. Koulutuspalvelujen tuottavuuden ja tehokkuuden mittaaminen. Teoksessa A. Kangasharju (toim.) *Hyvinvointipalvelujen tuottavuus. Tuloksia opintien varrelta*. Valtion taloudellinen tutkimuskeskus. VATT-julkaisu 46. Helsinki.
- Kirjavainen, T. 2007. Nuorten lukiokoulutuksen tehokkuus. Valtion taloudellinen tutkimuskeskus. VATT-tutkimuksia 131. Helsinki.
- Kirjavainen, T. & Loikkanen, H. A. 1998. Efficiency differences of Finnish senior secondary schools: An application of DEA and Tobit-analysis. *Economics of Education Review* 17, 377–394.
- Krueger, A. 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* 114, 497–532.
- Krueger, A. 2003. Economic considerations and class size. *The Economic Journal* 113, F34–F63.
- Kumbhakar, S. C. & Lovell, K. C. A. 2000. *Stochastic frontier analysis*. Cambridge: Cambridge University Press.
- Mosteller, F. 1995. The Tennessee study of class size in the early grades. *The Future of Children* 5, 113–127.
- Ollikainen, V. 2007. Ammatillisen peruskoulutuksen kustannustehokkuus 2001–2003. Valtion taloudellinen tutkimuskeskus. VATT-tutkimuksia 132. Helsinki.
- Portela, M. C. A. S & Thanassoulis, E. 2001. Decomposing school and school-type efficiency. *European Journal of Operational Research* 132, 357–372.
- Raudenbush, S. & Bryk, A. 2002. *Hierarchical linear models: Application and data analysis methods*. 2nd edition. Newbury Park: Sage Publications.
- Wooldridge, J. M. 2002. *Econometric analysis of cross section and panel data*. The MIT Press.
- Worthington, A. C. 2001. An empirical survey of frontier efficiency measurement techniques in education. *Education Economics* 9, 245–268.



Ammatillisen osaamisen arviointi

Arvioinnin rooli on ollut koulutuksessa aina keskeinen. Koulutuksen muotojen ja oppimiskontekstien monipuolistuessa myös suoritusten ja osaamisen arviointi on joutunut uusien haasteiden kohteeksi. Työpaikalla tapahtuvan oppimisen tullessa yhä tavanomaisemmaksi on alettu etsiä vaihtoehtoisia lähestymistapoja käytännön osaamisen arvioimiseksi. Mittaamisen sijasta voidaan pikemminkin puhua arviointikulttuurista, joka korostaa opetuksen ja arvioinnin yhdistämistä. Tästä arvioinnista käytetään myös nimitystä autenttinen arviointi, suoritusten arviointi, vaihtoehtoinen arviointi tai käytännön osaamisen arviointi. Tämän artikkelin tarkoituksena on kuvata ammatillisen osaamisen arviointia, joka tässä rinnastetaan käsitteeseen käytännön osaamisen arviointi. Termiä käytettiin vuosina 2003–2006 toteutetussa Leonardo da Vinci -projektissa ”QUAL-PRAXIS” (Quality Assurance and Practice-Oriented Assessment in Vocational Education). Suomen osalta tutkimuksen kohteena tässä projektissa olivat ammatillisen koulutuksen ammattiosaamisen näytöt, jotka tuolloin olivat vielä kokeiluvaiheessa.

Asiasanat: osaaminen, ammattitaito, osaamisen arviointi, autenttinen arviointi

Johdanto

Ammattitaidon arviointi on tullut ajankohtaiseksi mm. kansainvälistymisen myötä. Ammatillisessa koulutuksessa Kööpenhaminan prosessin ja korkeakoulutuksessa Bologna prosessin myötä on Euroopassa kehitetty strategioita liikkuvuuden lisäämiseksi ja osaamisen tunnustamiseksi. EQF (European Qualification Framework) on yhteinen suositus läpinäkyvyyden edistämiseksi ja osaamisen tunnustamiseksi, ja ECVET (European Credit Transfer System for Vocational Education and Training) pyrkii kehittämään ammatillisen koulutuksen opintosuoritusten siirtojärjestelmän, jonka avulla voidaan edistää toisessa maassa suoritettujen opintojen tai muuten hankitun osaamisen tunnustamista (Commission of the European Communities 2005). Kansainvälisesti läpinäkyvyys helpottaa tutkintojen ja osaamisen vertailua eri maiden välillä. Parempi läpinäkyvyys tarkoittaa myös sitä, että opiskelijoiden aikaisemmat opintosuoritukset tai ammatillisissa oppilaitoksissa suoritettu työssäoppiminen on helpompaa hyväksyä (Rouhiainen ja Valjus 2003).

Yksi keskeinen suuntaus ammatillisessa koulutuksessa on ollut työelämäjaksojen yhdistäminen osaksi opetussuunnitelmaa. Myös Suomessa on toteutettu toisen asteen ammatillisen koulutuksen uudistus vuodesta 2001 lähtien, jolloin puolen vuoden työssäoppimisen jakso sisällytettiin opetussuunnitelmaan (Pohjonen 2005; Tynjälä, Virtanen & Valkonen 2005; Uusitalo 2001). Koulutuksen ja työelämän yhteistyö tiivistyneenä edelleen toisen asteen ammatillisessa koulutuksessa, kun ammattiosaamisen näytöt on liitetty syksyllä 2006 ammatillisen peruskoulutuksen opetussuunnitelmaperusteiseen koulutukseen. Suomalaiseen ammatilliseen perustutkintoon kuuluvat näytöt edustavat uudenlaista arviointijärjestelmää, jossa yhdistyvät koulu- ja työelämälähtöisen ammatillisen koulutuksen elementit ja jossa näytöt nähdään ammattiin oppimisen prosessina (Nyyssölä 2003; Stenström 2001; Virolainen 2001). Ammattiosaamisen näytöt ovat koulutuksen järjestäjän ja työelämän yhdessä suunnittelema, toteuttama ja arvioima työtilanne tai -prosessi. Ammattiosaamisen näytössä opiskelija osoittaa tekemällä käytännön työtehtäviä mahdollisimman aidoissa työtilanteissa, miten hyvin hän on saavuttanut opetussuunnitelman perusteiden tavoitteissa määritellyn työelämän edellyttämän ammattitaidon (Räkköläinen 2005). Ammattiosaamisen näytöt eroavat (aikuiskoulutuksen) näyttötutkinnoista siten, että viimeksi mainituissa arvioidaan työtehtävissä tarvittavaa ammattitaitoa kiinnittämättä huomiota siihen, miten osaaminen on hankittu. Sen sijaan ammattiosaamisen näytöissä arvioidaan ammattitaidon kehittymistä koko koulutuksen aikana (Räsänen & Räkköläinen 2005).

Tämän artikkelin tarkoituksena on kuvata ammatillisen osaamisen arviointia, joka rinnastetaan käsitteeseen käytännön osaamisen arviointi. Termiä käytettiin vuosina 2003–2006 toteutetussa Leonardo da Vinci -projektissa ”QUAL-PRAXIS”

(Quality Assurance and Practice-Oriented Assessment in Vocational Education). Suomen osalta tutkimuksen kohteena tässä projektissa olivat ammatillisen koulutuksen ammattiosaamisen näytöt, jotka tuolloin olivat vielä kokeiluvaiheessa (Stenström, Laine & Kurvonen 2006). Tässä artikkelissa hyödynnetään edellä mainitussa projektissa saatuja tuloksia (Grollmann & Stenström 2005; Stenström & Laine 2006a & b).

Käytännön osaamisen arvioinnin taustalla olevia suuntauksia

Arviointiajattelun paradigma on muuttunut mittaamisesta arviointiin. Birenbaum (1996) on erottanut kaksi suoritusten mittaamisen kulttuuria: mittauskulttuurin ja arviointikulttuurin. Perinteisessä mittauskulttuurissa paino on tulosten objektiivisuudessa, jolloin opetuksen ja testauksen ajatellaan olevan erillisiä toimintoja. Oppimisyhteiskunnassa arviointikulttuuri nähdään vaihtoehtona testauskulttuurille. Arviointikulttuuri korostaa prosessia ja siihen vaikuttavien subjektiivisten tekijöiden huomioimista. Lisäksi opetus ja arviointi nähdään yhdeksi prosessiksi. Tästä arvioinnista käytetään myös nimityksiä autenttinen arviointi, suoritusten arviointi tai vaihtoehtoinen arviointi. Uusissa konteksteissa tapahtuvalle oppimisen (esim. työssä oppimisen) arvioinnille on tyypillistä, että arviointi voidaan toteuttaa useammin kuin kerran ja sillä voi olla erilaisia funktioita, päämääriä ja menetelmiä, jotka huomioivat myös uudet arviointimuodot (Dochy & McDowell 1997).

Mittauksen vaihtoehtoisia lähestymistapoja on tutkittu 1990-luvulta alkaen (Birenbaum & Dochy 1996; Lesh & Lamon 1992; Nitko 1995; Resnick & Resnick 1992). Nämä tutkimukset kuvaavat kontekstuaalisen ja autenttisen lähestymistavan kehittämistä. Kompetenssiperustaisten oppimistulosten arviointi, joka liittyy autenttiseen arviointiin, edustaa paradigman muutosta mittauskulttuurista arviointikulttuuriin (Birenbaum 1994; 1996; Dierick & Dochy 2001). Testaus yhdistetään usein tietoperustaisiin kirjallisiin kokeisiin vastakohtana kompetenssien arvioinnille. Biemans, van den Elsen, Wesselink ja Mulder (2006) ovat esittäneet kuusi kriteeriä kompetenssiperustaiselle arvioinnille:

- Arvioinnin tulisi tapahtua yhdessä tai useammassa autenttisessa ympäristössä. Arviointitehtävien tulisi vaatia kompetensseja, joita tarvitaan tulevissa työympäristöissä.
- Tieto, taidot ja asenteet pitäisi integroida arviointikäytäntöihin (komponenttien integrointi), ja teoria ja käytäntö tulisi arvioida yhdessä.
- Arviointi tulisi tapahtua ennen oppimisprosessia, sen aikana ja sen jälkeen (integraatio oppimisprosessissa). Arviointi pitäisi integroida oppimisprosessiin.

- Arviointi pitäisi olla yhä enemmän oppijan vastuulla (oppijan vastuu). Oppijalla tulisi olla aktiivinen rooli arvioinnissa.
- Arvioinnin moniulotteisuus on tärkeä lisä reliabiliteettiin ja validiteettiin. Eivät vain perinteisen testauksen kriteerit kuten reliabiliteetti ja validiteetti ole tärkeitä, vaan myös muut kriteerit kuten läpinäkyvyys ja autenttisuus (Dierick & Dochy 2001). Lisäksi arvioinnin laatua voi parantaa moniulotteisen lähestymistavan avulla (erilaiset arviointimenetelmät, useampi arvioitsija ja useat arviointiajankohdat).
- Arvioinnin tarkoituksena pitäisi olla sekä oppimistulosten formaali mittaaminen ja kompetenssien kehittyminen (tulokset ja kehittyminen). Arvioinnilla ei saisi olla vain summatiivinen, vaan myös formatiivinen tehtävä siinä mielessä, että arvioinnin tulisi kannustaa ammattitaidon (kompetenssien) kehittymistä (assessment for learning) (Segers, Dochy & Cascallar 2003).

Käytännön osaamisen arvioinnille tyypillisiä piirteitä

QUAL-PRAXIS-projektissa vaihtoehtoisesta arvioinnista käytettiin nimitystä käytännön osaamisen arviointi (practice-oriented assessment), jolla viitattiin suori- tusperustaiseen, kompetenssiperusteiseen ja autenttiseen arviointiin (Stenström 2005). Käytännön osaamisen arviointi määriteltiin kohteen, kontekstin, arviointimenetelmien ja arvioitsijoiden perusteella (vrt. Pellegrino 2003). Käytännön osaamisen arvioinnin kohde on työelämän edellyttämä ammattitaito. Arviointi tehdään mahdollisimman autenttisisessa kontekstissa, ja menetelmänä käytetään observointia, kirjallisia tehtäviä ja arviointikeskustelua, joka osoittautui yhdeksi keskeiseksi oppimista edistäväksi tekijäksi. Arvioitsijana toimii opettajan lisäksi työpaikkaohjaaja; myös opiskelija osallistuu keskusteluun ja arvioi suoritus- taan.

QUAL-PRAXIS-projektissa voitiin havaita käytännön osaamisen arvioinnissa yhteisiä piirteitä huolimatta siitä, että tutkimukseen osallistuneiden maiden koulutusjärjestelmissä esiintyi eroavuuksia (Stenström & Laine 2006b). Nämä yhteiset piirteet voidaan luokitella seuraavasti:

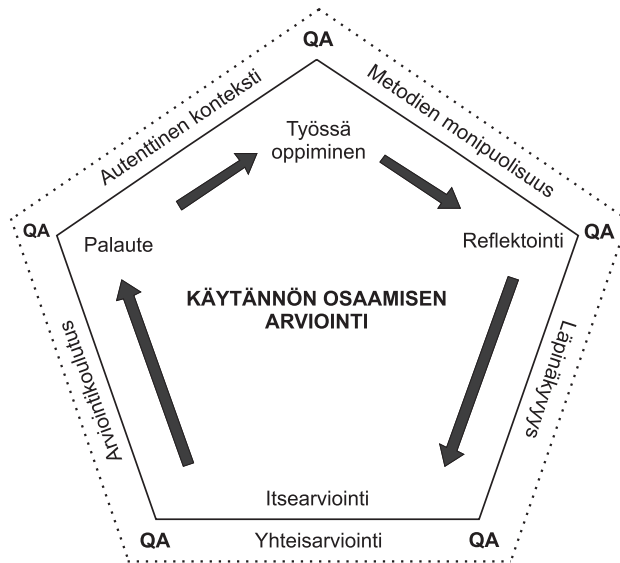
Oppimista edistävät tekijät:

- arviointi osana oppimista
- reflektio
- itsearviointi
- palaute.

Laadunvarmistusta edistävät tekijät:

- autenttinen konteksti
- arvioitsijakoulutus
- läpinäkyvä arviointi
- yhteisarviointi
- arviointimenetelmien moninaisuus.

Yllä olevia tekijöitä voidaan luonnehtia siten, että ne edustavat sekä toimintaympäristöä että oppimisprosessia. Kuviossa 1 sisäkehä kuvaa käytännön osaamista hyödyntävää oppimisprosessia, kun taas ulkokehä kuvaa käytännön osaamisen arviointia. Arviointiprosessiin liittyvät ympäristötekijät toimivat laadunvarmistuksen tekijöinä ja auttavat siten käytännön osaamisen arvioinnin toteuttamista menestyksellisesti. Arvioinnin piirteet kuvaavat myös ammattiosaamisen näyttöjä (ks. Stenström, Laine & Kurvonen 2006).



Kuvio 1. Käytännön osaamisen arvioinnin hyvien käytäntöjen piirteet

Käytännön osaamisen (työssä oppimisen) arvioinnin tyypillisiä piirteitä verrataan seuraavassa perinteisen arviointiajattelun vastaaviin piirteisiin (vrt. Biggs 1994; Black 1999; Eisner 1993; Tynjälä 1999; ks. taulukko 1).

Taulukko 1. *Perinteisen arvioinnin ja käytännön osaamisen arvioinnin keskeiset piirteet (Tynjälän 1999 pohjalta)*

Perinteinen arviointi	Käytännön osaamisen arviointi
Painottaa määrällistä arviointia	Painottaa laadullista arviointia
Toistavaa, painottaa ulkoa oppimista	Painottaa teorian ja käytännön yhteyksiä sekä tietämyksen muuntamista
Keinotekoiset arviointiympäristöt	Mahdollisimman autenttiset arviointiympäristöt
Arviointi erillään oppimisprosessista	Arviointi osana oppimisprosessia
Arvioijana opettaja	Arviointi opettajan, oppilaan ja työpaikkaohjaajan yhteistyönä
Keskittyminen tuloksiin	Keskittyminen oppimisprosessiin, muutoksiin oppilaan tiedoissa ja taidoissa sekä oppimistuloksiin

Perinteisessä arvioinnissa painotetaan määrällistä arviointia ja ulkoa oppimista, kun taas käytännön osaamisen arvioinnissa korostetaan laadullista arviointia, eli huomio kiinnitetään opiskelijan esittämän tiedon laatuun. Keskeinen ero perinteisen ja käytännön osaamisen arvioinnissa on käsitys arviointikontekstista. Perinteinen arviointi tapahtuu kurssin päättyessä luokkahuoneessa, kun sen sijaan käytännön osaamisen arviointi suoritetaan mahdollisimman autenttisessa ympäristössä osana oppimisprosessia. Oppimisprosessin osana toteutettu arviointi on väline, jonka tarkoituksena on eri tavoin kehittää ja edistää opettamista ja oppimista. Lisäksi perinteinen ja käytännön osaamisen arviointiajattelu eroavat siinä, ketkä nähdään arvioijiksi. Perinteisen arviointiajattelun mukaan opiskelija-arviointi on yksin opettajan tehtävä, kun taas käytännön osaamisen arvioinnissa opettaja toimii yhteistyössä muiden arvioijien kanssa (mukana työpaikkaohjaaja, myös opiskelija osallistuu arviointikeskusteluun).

Käytännön osaamisen arvioinnin voidaan katsoa kuuluvan autenttiseen arviointiin. Autenttisen synonyymeja ovat aito, todellinen ja luonnollinen. Autenttiselle arvioinnille ovat tyypillisiä elämänläheisiin tilanteisiin liittyvät laajahkot oppijan omaa tiedon rakentamista painottavat tehtävät, jotka vaativat suunnittelua, ideointia, ongelmien havaitsemista ja niiden ratkaisua sekä ratkaisujen argumentointia ja kommunikointia. Tyypillisiä menetelmiä ovat näytöt, näyte- ja projektityöt, tutkielmat, erilaiset esitykset ja portfoliot. (Linnakylä & Välijärvi 2005, 82.)

Käytännön osaamisen hyvien käytäntöjen elementtejä (Stenström & Laine 2006b) voidaan verrata autenttisen arvioinnin käytäntöihin (Linnakylä & Välijärvi 2005) ja huomata yhtäläisyyksiä, vaikka termit ovat erilaisia (ks. taulukko

Taulukko 2. *Autenttisen arviointikäytännön piirteet ja käytännön osaamisen arvioinnin hyvien käytäntöjen elementit*

Autenttisen arviointikäytännön piirteet	Käytännön osaamisen arvioinnin hyvien käytäntöjen elementit
Tiedon tuottaminen	Käytäntökeskeinen oppiminen: näyttö, tutkinto
Oppijan oma aktiivisuus	Itsearviointi
Tutkiva ja soveltava ote	Reflektointi
Merkityksellisyys oppijan oman elämän kannalta	Arviointi osana oppimisprosessia
Kompleksiset tieto- ja taitorakenteet	Autenttinen konteksti
Yhteisöllisyyden arvostaminen	Yhteisarviointi (opettaja, työpaikkaohjaaja, opiskelija)
Laadullinen, arvioitsijoiden asiantuntevaan harkintaan perustuva arvostelu	Arvioitsijoiden koulutus

2).

Autenttisessa arvioinnissa tiedontuotto merkitsee suuntautumista tuotantoprosessiin ja tuotteeseen. Autenttisessa arvioinnissa keskitytään myös sekä prosessiin että tuloksiin. Molemmat näistä arviointitavoista tähdentävät oppijan omaa toimintaa, mitä voidaan kutsua myös itsearviointiksi. Lisäksi molemmat arviointikäytännöt painottavat tutkivaa ja soveltavaa otetta, toisin sanoen reflektiivisyyttä. Autenttisessa arvioinnissa korostetaan, että arvioinnin tulisi olla merkityksellistä opiskelijalle. Samoin käytännön osaamisen arvioinnissa painotetaan, että arviointi tulisi nähdä holistisesti, osana oppimista. Autenttiseen arviointiin sisältyy tyypillisesti kompleksisia tieto- ja taitorakenteita. Samalla sille on yhteistä käytännön osaamisen arvioinnin kanssa periaate, että arvioinnin tulisi tapahtua autenttisessa kontekstissa. Molemmissa arviointitavoissa on myös tärkeää, että arviointi on yleistettävää ja luotettavaa, mitä voidaan parantaa arviointisajakoulutuksella.

Ammattiosaamisen näytöt ammattitaidon arviointimenetelmänä

Arviointia voidaan tarkastella suhteessa erilaisiin oppimisteoreettisiin suuntauksiin (ks. Poikela 1998). Arviointi ja oppiminen voidaan nähdä joko erillisinä tai toistensa osaprosesseina. Jos oppiminen ja arviointi ovat erillisiä prosesseja, niin arvioinnin tehtävänä on mitata eikä suinkaan tukea oppimista. Tällöin oppimisen oletetaan ohjautuvan opetussuunnitelman ja opetuksen perusteella. Jos

koulutus nähdään oppimisen kautta, niin arviointi voidaan vastaavasti nähdä osana oppimisen prosessia näyttöjen toimiessa osana arviointia. (Poikela 1998; Stenström 2005.)

Suomalaisessa ammatillisessa koulutuksessa arviointi on osa oppimista (Laki ammatillisesta koulutuksesta annetun lain muuttamisesta 15.6.2005/601). Myös ammattiosaamisen näyttöjen arviointiin liittyy aineksia, jotka mahdollistavat oppimisen ja arvioinnin yhdistämisen. Ammattiosaamisen näyttö on uudentyyppistä arviointia, joka toteutetaan yleensä opiskelijan työssäoppimisjakson aikana. Näyttö koostuu tietyistä etukäteen määriteltävistä ammattitehtävistä. Ammattiosaamisen näyttöjen lähtöperiaate on, että ne toteutetaan työtilanteissa, jotka ovat niin autenttisia kuin mahdollista (Suursalmi 2003, 18). Näytöille sopivien tilanteiden löytäminen taas riippuu mm. työpaikan aikataulusta. Arvioinnin painopisteenä on, miten opiskelija käyttää ja soveltaa tietojaan ja taitojaan. Ajatuksena on, että aidot ja realistiset tehtävät auttavat aloittelijoiden kouluttamisessa kohti asiantuntemusta ja ammattimaisuutta. (Räkköläinen 2005, 13.)

Tärkeä osa ammattiosaamisen näyttöä on oppimisprosessia tukeva, kunkin näytön jälkeen toteutettava arviointikeskustelu. Ammatillisessa koulutuksessa arviointikeskustelu tuo yhteen kolme eri näkökulmaa koskien opiskelijan arviointia: opiskelijan, opettajan ja työpaikkaohjaajan (Stenström, Laine & Kurvonen 2006). Tämän keskustelun kuluessa tarkastellaan opiskelijan suoritusta ammattiosaamisen näytössä. Arviointi keskittyy siihen, miten opiskelija hallitsee työmenetelmät, työvälaineet ja -materiaalit, työprosessit, työturvallisuuden, yhteiset painopistealueet (esim. tekniikan, informaatiotekniikan ja yrittäjyyden hyödyntäminen), kaikille aloille yhteisen ydinosaamisen (esim. ongelmanratkaisutaidot sekä vuorovaikutus- ja viestintätaidot) sekä työn perustana olevan tiedon. Nämä ovat myös opetussuunnitelman mukaisia arvioinnin alueita (Kinnunen 2003, 71).

Opiskelijoiden itsearviointi ja reflektointi ovat tekijöitä, jotka yhdistävät ammattiosaamisen näytöt ja työssäoppimisen tiiviisti oppimisprosessiin. Itsearviointi on yksi arviointikeskusteluissa käsiteltävistä aiheista. Opiskelijat arvioivat oppimistaan sekä ammattiosaamisen näyttöön että koko työssäoppimisjaksoon liittyen. Reflektiota tapahtuu tällöin koko työssäoppimisprosessin ajan. Arviointikeskusteluihin liittyvän reflektoinnin lisäksi ammattiosaamisen näytöt ja työssäoppimisjaksot ovat saaneet aikaan reflektointia paitsi niin sanottujen ensisijaisten oppijoiden eli opiskelijoiden niin myös työpaikkaohjaajien ja opettajien keskuudessa. Työpaikkaohjaajat ovat sitä mieltä, että opiskelijoiden ohjaaminen on saanut heidät ajattelemaan uudelleen omia toimintatapojaan, työmenetelmiään sekä löytämään uusia näkökulmia omaan työhönsä. Opettajat puolestaan katsovat, että ammattiosaamisen näyttöjen käyttöönotto on laajentanut heidän näkökulmiaan ja auttanut heitä pysymään ajan tasalla. (Stenström, Laine & Kurvonen 2006.)

Näyttöjen aikana kaikilla näillä osapuolilla on omat roolinsa ja vastuualueensa (Stenström, Laine & Kurvonen 2006). Opiskelijan tärkein tehtävä on näyttää taitonsa. Opiskelija laatii myös kirjallisen näyttösuunnitelman. Lisäksi opiskelija on opettajan ja työpaikkaohjaajan mukana arviointikeskustelussa. Ammattiosaamisen näytössä työpaikkaohjaaja arvioi opiskelijaa asianmukaisten arviointikriteerien perusteella. Opettaja on arvioinnin asiantuntija, joka vastaa ammattiosaamisen näyttöjen arvioinnista. Sen vuoksi hänen täytyy olla perillä käytettävistä arviointikriteereistä. (Ks. Vehviläinen 2004, 65.)

Arvioitsijoiden ammattitaito ja koulutus ovat keskeisiä laadun takaajia (Stenström & Laine 2006). Tämän vuoksi pidetään tärkeänä, että työpaikkaohjaajat saavat koulutuksen ammattiosaamisen näyttöjen toteuttamista ja arviointia varten. Työelämän edustajat, jotka ovat mukana arvioimassa ammattiosaamisen näyttöjä, ovat asiantuntijoita omilla ammattialoillaan. Heidän peruskoulutuksensa kuitenkin vaihtelee, eivätkä he ole välttämättä selvillä arviointikäytännöistä ja näyttöjen hallinnoimisesta. Myös opettajien pitäisi olla vastaavasti riittävästi koulutettuja.

Ammattikoulutuksessa arvioinnin läpinäkyvyys perustuu avoimeen kommunikaatioon ja yhteistyöhön arvioinnin eri osapuolten välillä. Kaikkien opiskelijoiden arviointiin osallistuvien tulee olla tietoisia arvioinnin peruspiirteistä ja käytännön järjestämisestä, kuten arvioinnin tavoitteista ja kriteereistä, arvosanojen määräämisperusteista ja opintoasiakirjoihin ja tutkintotodistuksiin kirjattavista asioista (Ammatillisen peruskoulutuksen opetussuunnitelman ja näyttötutkinnon perusteet 2005, 2).

Arvioinnin laatuksiteerit

Riippumatta siitä, mikä tai mitkä arviointimenetelmät kulloinkin valitaan, tärkeimmiksi arvioinnin laatuksiteereiksi ovat määritelty mm. seuraavat kriteerit:

- validiteetti
- luotettavuus
- objektiivisuus
- läpinäkyvyys
- tasapuolisuus ja
- oikeudenmukaisuus.

Arvioinnin validiteetti tarkoittaa, että arviointi mittaa juuri sitä, mitä sen on tarkoitus mitata. Validiteetti viittaa tarkoituksenmukaisuuteen, merkityksellisyteen ja mittaustuloksista tehtyjen erityisten päätelmien käyttökelpoisuuteen (AERA 1999). Validiteettiin liittyy oleellisesti myös läpinäkyvyyden periaate.

Sen mukaan arvioinnin kriteerien tulee olla kaikkien arviointiin osallistuvien osapuolten tiedossa. Lisäksi läpinäkyvyyteen liittyy oikeudenmukaisuuden periaate, joka arvioinnissa pitää sisällään sen, että kukaan ei saisi joutua muita huonompaan asemaan arviointiprosessin takia. (Williams ja Bateman 2004, 32, 34.) Lisäksi käytännön osaamisen arvioinnille asetetaan erityisvaatimuksia. Käytännön osaamisen arvioinnin tulisi pohjautua työpaikan todellisuuteen, sen erityisiin sääntöihin, normeihin, odotuksiin ja kieltoihin (ulkoinen validiteetti). Arviointiin tulisi sisältyä riittävä määrä relevantteja työtehtäviä, jotka on mahdollista suorittaa aidossa oppimisympäristössä (sisällöllinen validiteetti). Lisäksi käytännön osaamisen arvioinnin validiteettivaatimus koskee oppimistulosten lisäksi oppimisprosessia (Benett 1999, 277–282) ja myös oppijan kykyä kehittyä ammatissaan (ennustevaliditeetti).

Työssäoppimisen kontekstissa luotettavuus (reliabiliteetti) on sidoksissa siihen, missä määrin käytetty arviointimenetelmä on altis vaihtelulle arviointiprosessin kuluessa. Yksi vaihtelutyyppi on tehtävien vaihtelu (toistettavat tehtävät). Ajalliset vaihtelut aiheuttavat myös ongelmia. Maailma ei pysy muuttumattomana, vaikka arviointitehtävät eivät muutukaan. Tämän vuoksi toistettavuus on ongelmallisempaa, jolloin arvioinnin luotettavuus kärsii. Kuitenkin erilaisia taitoja arvioitaessa yhteinen tekijä todennäköisesti vaikuttaa yhtäläistävästi ja tasaavasti ja tekee siten arvioinnista luotettavampaa. (Benett 1999, 284–286.)

Luotettavuus liittyy siihen, että arviointiprosessiin ei saa sisältyä mitään satunnaisia virheitä (Straka 2004). Arviointimenetelmää valittaessa on mietittävä, millä tavoin opiskelija voi parhaiten osoittaa osaamisensa. Samassakin ryhmässä opiskelevia tulee arvioida erilaisin menetelmin, jolloin arvioinnissa tapahtuvia virheitä voidaan välttää. Opiskelijan tulee voida luottaa siihen, että arviointiin eivät vaikuta opettajan tai työpaikkaohjaajan henkilökohtaiset näkemykset tai tunteet opiskelijaa kohtaan. Reliabiliteetin keskeinen osa on objektiivisuus. Arvioinnin objektiivisuus varmistetaan siten, että opiskelijan arvioinnissa on useampi henkilö ja että näiden kokemuksellinen tausta on erilainen. (Helakorpi 2001, 188.)

Vertailtavuus on myös huomionarvoinen seikka, joka vaikuttaa luotettavuuteen arvioitaessa käytännön osaamista (työssä oppimista). Kysymys on siitä, onko opiskelijoiden suoritusaste muuttumaton paikasta toiseen ja vuodesta toiseen ja siten vertailukelpoinen suhteessa aikaisempiin suorituksiin tai muiden suorituksiin (Benett 1999, 286).

Tasa-arvo on yksi oleellisimmista arvioinnin kriteereistä. Tasa-arvoisuus nousee esiin jo suunnitteluvaiheessa, jolloin työssäoppimispaikkojen laadunvarmistuksella pyritään siihen, että kaikki opiskelijat saavat tietyt perusvalmiudet ja oppivat työelämässä vaadittavia metataitoja (Vehviläinen 2004, 87). Arvioinnin tasa-arvoisuus tarkoittaa sitä, että opiskelijaa pyritään arvioimaan mahdollisimman objektiivisesti, henkilökohtaiset ominaisuudet ja persoonallisuuspiirteet

huomioiden (Williams & Bateman 2003, 33). Tasa-arvoisuus tarkoittaa myös sitä, että opiskelijoille pyritään laadunvarmistuksen avulla takaamaan riittävän hyvät olosuhteet työssäoppimis- ja näyttöpaikoissa. (Vehviläinen 2004, 87.)

Lopuksi

Räisänen ja Rökköläinen (2005) nostavat esiin kysymyksen, pitäisikö ammatilliseen osaamiseen liittyvässä oppimistulosten arvioinnissa luopua perinteisestä käsityksestä arviointitiedon luotettavuudesta, totuudellisuudesta ja yleistettävyydestä. Yleistettävyyden sijaan voisi painottaa asioiden suhteellisuutta ja paikallisen tason ymmärryksen lisäämistä, jatkuvaa oppimista ja prosessinomaisuuteen liittyvää näkemystä laadunparantamisen ja oppimisen jatkuvuudesta. Tällöin ulkoinen laadunvarmistus ja laatukontrolli takaavat riittävän luotettavuuden pitkällä aikavälillä. (Räisänen & Rökköläinen 2005.)

Arviointitulosten siirrettävyys toiseen kontekstiin riippuu siitä, miten samankaltaisia arviointiympäristö ja sovellusympäristö ovat (Lincoln & Guba 1985). Käytännön osaamisen arvioinnissa olisi tärkeää, että arviointi tapahtuu mahdollisimman autenttisissa olosuhteissa. Arvioinnin autenttisuuden lisääminen oletetaan parantavan opiskelijan oppimista ja motivaatiota. Tällöin tulevat tärkeäksi muut laadunvarmistuksen tekijät, joita esiteltiin edellä.

Arvioijan tekemät johtopäätökset suodattuvat aina hänen oman viitekehyksensä ja tarkastelunäkökulmansa lävitse. Onkin esitetty, että arvioijien ennakkoon saamalla arviointiohjeilla, -kriteereillä ja -lomakkeilla on ratkaiseva vaikutus siihen, mihin arvioijat keskittävät tarkkaavaisuutensa ja havaintonsa varsinaisessa arviointitilanteessa. Arvioinnin avulla voidaan siis tavoittaa pikemminkin subjektiivisia tai kollektiivisesti jaettuja näkökulmia kuin objektiivinen totuus sinänsä.

Voidaan todeta, että ei ole olemassa kaikkiin tapauksiin ja tilanteisiin soveltuvaa arvioinnin muotoa. Kaikentyyppisillä arvioinneilla on sekä positiivisia että negatiivisia vaikutuksia opetukseen ja oppimiseen (Dochy & McDowell 1997). Vaihtoehtoista (autenttista, suoritusperusteista tai käytännön osaamisen) arviointia kannattaa kuitenkin kehittää ja tutkia yhtenä ammattitaidon arviointimenetelmänä, sillä arviointitutkijat ovat yhtä mieltä siitä, että käytännön osaamisen arviointi on parempi suorittaa autenttisissa kontekstissa kuin luokkahuoneessa kirjallisin kokein (Wolf 1995, 43; Benett 1999, 278). Siksi on suuri haaste kehittää käytännön osaamisen arvioinnin metodologiaa, joka pystyy tavoittamaan kontekstispesifin ja hiljaisen tiedon (Björnväld 2001, 25).

Lähteet

- AERA (American Educational Research Association). 1999. Standards for educational and psychological testing. American Educational Research Association, American Psychological Association and National Council on Measurement in Education. Washington D. C.: American Psychological Association.
- Ammatillisen peruskoulutuksen opetussuunnitelman ja näyttötutkinnon perusteet. 2005. Opiskelijan arviointi, opetussuunnitelma. Määräys 32/011/2005. Helsinki: Opetushallitus.
- Benett, Y. 1999. The validity and reliability of assessment and self-assessments of work-based learning. Teoksessa P. Murphy (toim.) *Learners, learning and assessment*. Trowbridge, UK: Cromwell Press, 277–289.
- Biemans, H., van den Elsen, E., Wesselink, R. & Mulder, M. 2006. Competence-based assessment in Dutch VET: Criteria and experiences. Paper presented at the ECER 2006 Geneva, 13.9.2006.
- Biggs, J. 1994. Student learning research and theory. Where do we currently stand? Teoksessa G. Gibbs (toim.) *Improving student learning. Theory and practice*. Oxford: Oxford Centre for Staff Development, 1–9.
- Birenbaum, M. 1994. Toward adaptive assessment – The student’s angle. *Studies in Evaluation* 20, 239–255.
- Birenbaum, M. 1996. Assessment 2000: Towards a pluralistic approach to assessment. Teoksessa M. Birenbaum & F. Dochy (toim.) *Alternatives in assessment of achievement, learning processes and prior learning*. Boston: Kluwer, 3–30.
- Birenbaum, M. & Dochy, F. 1996. Introduction. Teoksessa M. Birenbaum & F. Dochy (toim.) *Alternatives in assessment of achievements, learning processes and prior learning*. Boston: Kluwer, xiii–xv.
- Bjönåvåld, J. 2001. Making learning visible: Identification, assessment and recognition of nonformal learning. *European Vocational Training Journal* 22, 24–32.
- Black, O. 1999. Assessment, learning theories and testing systems. Teoksessa P. Murphy (toim.) *Learners, learning & assessment*. London: Paul Cahpman in association with the Open University, 118–134.
- Commission of the European Communities. 2005. Commission staff working paper. Towards a European qualifications framework for lifelong learning. Brussels.
- Dierick, S. & Dochy, F. 2001. New lines in edometrics. New forms in assessment lead to new assessment criteria. *Studies in Educational Evaluation* 27, 307–329.
- Dochy, J. & McDowell, L. 1997. Introduction: Assessment as a tool for learning. *Studies in Educational Evaluation* 23 (4), 279–298.
- Eisner, E. 1993. Reshaping assessment in education: Some criteria in search of practice. *Journal of Curriculum Studies* 25, 219–233.
- Grollmann, P. & Stenström, M.-L. (toim.) 2005. Quality assurance and practice-oriented assessment in vocational education and training: Country studies. University of Bremen, Germany. Institute Technology and Education. Working Papers 55.
- Helakorpi, S. 2001. Osaamisen arviointi. Teoksessa M. Rökköläinen & I. Uusitalo (toim.) *Työssäoppiminen ja ohjaus ammatillisissa oppilaitoksissa*. Tampere: Tammi, 173–193.
- Kinnunen, E. 2003. Arvioinnin toteuttaminen. Teoksessa E. Kinnunen & T. Halmevuo (toim.) *Opiskelijan arvioinnin opas ammatilliseen koulutukseen*. Helsinki: Opetushallitus, 16–32.
- Laki ammatillisesta koulutuksesta annetun lain muuttamisesta 15.6.2005/601.
- Lave, J. & Wenger, E. 1991. *Situated learning. Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Lesh, R. & Lamson, S. 1992. *Assessment of authentic performance in school mathematics*. Washington DC: American Association for the Advancement of Science.

- Lincoln, Y. S. & Guba, E. G. 1985. *Naturalistic inquiry*. Beverly Hills, California: Sage.
- Linnakylä, P. & Välijärvi, J. 2005. Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulun kehittämiseksi. Jyväskylä: PS-kustannus.
- Nitko, A. 1995. Curriculum-based continuous assessment: A framework for concepts, procedures and policy. *Assessment in Education* 2, 321–337.
- Nyyssölä, N. 2003. Näytöt ammatillisessa peruskoulutuksessa 2002. Väliraportti näyttöprojektien toiminnasta. Helsinki: Opetushallitus.
- Pelligrino, J. W. 2003. Knowing what students know. Recent advances in the cognitive and measurement sciences should be the foundation for developing a new system of student assessment. *Issues in Science and Technology* 19 (2), 48–52.
- Pohjonen, P. 2005. Työssäoppiminen: ammatillisen osaamisen perusta. Jyväskylä: PS-kustannus.
- Poikela, E. 1998. Oppiminen, arviointi ja osaaminen. Teoksessa A. Räisänen (toim.) *Halitanko ammatti? Pätevyyden määrittelyä arvioinnin perustaksi*. Opetushallitus. Arviointi 2. Helsinki, 35–46.
- Resnick, J. & Resnick, D. 1992. Assessing the thinking curriculum. Teoksessa B. Gifford & M. O'Connor (toim.) *Changing assessment: Alternative views of aptitude, achievement and instruction*. Boston: Kluwer, 37–77.
- Rouhiainen, P. & Valjus, S. 2003. Kohti tiiviimpää koulutuksen eurooppalaista yhteistyötä. Leonardo-ohjelma Kööpenhaminan julistuksen tukena. Helsinki: Opetushallitus, Leonardo-keskus & CIMO.
- Räisänen, A. & Rökköläinen, M. 2005. Oppimistulosten arviointi laadunvarmistuksena. *Ammatillinen koulutus*. Teoksessa H.K. Lyytinen & A. Räisänen (toim.) *Kehittämismuunta arvioinnista*. Koulutuksen arviointineuvoston julkaisuja 6. Jyväskylä, 193–211.
- Rökköläinen, M. 2005. Kansallisen näyttöperusteisen oppimistulosten arviointijärjestelmän kehittäminen ammatillisiin perustutkintoihin. Arviointikokeilusta kohti käytäntöä. Helsinki: Opetushallitus.
- Segers, M., Dochy, F. & Cascallar, E. 2003. *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht: Kluwer Academic Publishers.
- Stenström, M.-L. 2001. Näytöt ammatillisessa peruskoulutuksessa. Kokemuksia ja tutkimustarpeita. Helsinki: Opetushallitus.
- Stenström, M.-L. 2005. Assessment of work-based learning in VET as a subject for research: Quality assurance and practice-oriented assessment in vocational education and training. Special edition of *The Finnish Journal of Vocational and Professional Education*, 99–105.
- Stenström, M.-L. & Laine, K. (toim.) 2006a. Quality and practice in assessment. New approaches in work-related learning. University of Jyväskylä, Institute for Educational Research.
- Stenström, M.-L. & Laine, K. (toim.) 2006b. Towards good practices for practice-oriented assessment in European vocational education. University of Jyväskylä: Institute for Educational Research. Occasional Papers 30.
- Stenström, M.-L., Laine, K. & Kurvonen, L. 2006. Practice-oriented assessment in Finnish VET – Towards quality assurance through vocational skills demonstrations. Teoksessa M.-L. Stenström & K. Laine (toim.) *Quality and practice in assessment: New approaches in work-related learning*. University of Jyväskylä: Institute for Educational Research, 89–120.
- Straka, G. A. 2004. Measurement and evaluation of competence. Teoksessa P. Descy & M. Tessaring (toim.) *The foundations of evaluation and impact research. Third report on vocational training research in Europe: Background report*. Luxembourg: Office for Official Publications of the European Communities. CEDEFOP Reference Series 58, 263–311.
- Suursalmi, P. 2003. Kolmikantaperiaatteen toteutuminen ammatillisen peruskoulutuksen näytöissä. Helsinki: Opetushallitus.

- Tynjälä, P. 1999. Oppiminen tiedon rakentamisena. Konstruktivistisen oppimiskäsityksen perusteita. Helsinki: Kirjayhtymä.
- Tynjälä, P., Virtanen, A. & Valkonen, S. 2005. Työssäoppiminen Keski-Suomessa. Taitava Keski-Suomi -tutkimus. Osa I. Jyväskylän yliopisto. Koulutuksen tutkimuslaitos. Tutkimusselesteita 23.
- Uusitalo, I. 2001. Työssäoppimisen ja kouluoppimisen vuorovaikutus. Teoksessa M. Räcköläinen & I. Uusitalo (toim.) Työssäoppiminen ja ohjaus ammatillisissa oppilaitoksissa. Helsinki: Tammi, 13–27.
- Vehviläinen, J. 2004. Ammattiosaamisen näyttöjen vaikutukset ammatillisen koulutuksen laatuun. Helsinki: Opetushallitus.
- Violainen, M. 2001. Ammatillisen peruskoulutuksen näytöt Englannissa, Saksassa, Hollannissa ja Tanskassa. Helsinki: Opetushallitus.
- Williams, M. & Bateman, A. 2003. Graded assessment in vocational education and training. An analysis of rational practice, drivers and areas for policy development. Leabrook, Australia: National Centre for Vocational Education Research.
- Wolf, A. 1995. Competence-based assessment. Bury St Edmunds, UK: St Edmundsbury Press.

IV

Kokoavaa tarkastelua



Arvotietoisuus, teorialähtöisyys ja vaikuttavuus arviointimenetelmien kehittämisessä

Tämä artikkeli on yhdistävä, tulkinnallinen yleiskatsaus julkaisun artikkeleissa esitetyistä keskeisistä asioista. Katsauksen tueksi on rakennettu malli, jonka puitteissa voidaan jäsentyneesti tarkastella arviointimenetelmien kehittämiseen yhteydessä olevia tekijöitä (kuvio 1, s. 206). Näkökulmallisesti on pidetty tärkeänä sitä, että tämä tieto voisi palvella myös koulutuksen kansallisen ja paikallisten arviointijärjestelmien menetelmien kehittämistä. Katsauksessa ei tarkastella yksittäisiä arviointimenetelmiä, vaan niiden kehittämisen tarpeet perustellaan arvioinnin ja arviointijärjestelmän kehittämisen yleisistä lähtökohdista.

Julkaisun kirjoittajat tarkastelevat artikkeleissaan keinoja, joita apuna käyttäen voidaan todeta, miten asiat ovat, kuinka hyvin asiat ovat ja osin myös sitä, miten asioiden tulisi olla. Tällöin liikutaan osin arvottamisen maailmassa, jonka päättelyn logiikka poikkeaa osin tieteellisen toiminnan tavanomaisista menettelytavoista. Artikkeleita selkeästi yhdistävinä juonteina on tässä nähty kolme tekijää: arvioinnin arvotietoisuus, teorialähtöisyys ja vaikuttavuus. Tarkasteluissa edetään käyttäen näitä avaimia. Katsauksessa ei viitata artikkelien kirjoittajiin.

Arvioinnin tarkoitus ja tavoitteet

Koulutuksen arvioinnin tarkoituksena (esim. perusopetuslaki 32/2003) on 1) turvata voimassa olevien koululakien tarkoituksen toteuttamista, 2) tukea koulutuksen kehittämistä ja 3) parantaa oppimisen edellytyksiä. Lakeja täydentäen asetus koulutuksen arvioinnista (150/2003) määrittelee arvioinnin tavoitteeksi 1) hankkia ja analysoida tietoa valtakunnallisen ja paikallisen päätöksenteon ja kehittämisen pohjaksi ja 2) tukea opiskelijoiden oppimista, opetustoimen henkilöstön työtä ja oppilaitosten kehittämistä.

Edellä todetun perusteella arvioinnin tarkoituksessa ja tavoitteissa korostuvat tiivistetysti seuraavat asiat: 1) tiedon tuottaminen ja 2) koulutuksen, opetuksen ja oppimisen kehittäminen. Viimeksi mainittu alleviivaa koulun ydintehtävää. Nämä tekijät ovat myös selkeästi lähtökohtina silloin, kun on valittava ne strategiat, menetelmät ja toimintatavat, joilla arvioinnin tarkoitus ja tavoitteet toteutetaan. Kyse on perimmältään arviointia ohjaavien arvojen toteutumisesta käytännön arviointitoiminnassa. Kun arvioinnin kohde on tiedossa, tehdään päätökset siitä, miten arvioinnin tarkoitus ja tavoitteet voivat parhaiten toteutua. Kyse on menetelmävalinnoista.

Arviointimenetelmien valinta

Arviointimenetelmästä voidaan puhua suppeamman käsityksen silloin, kun määritetään kohteen arvo tai merkitys eli arvotetaan. Laajemman käsityksen mukaan arviointimenetelmä voi sisältää koko sen prosessiketjun, jonka avulla tieto ja ymmärrys arvioinnin kohteesta otetaan haltuun, arvotetaan ja jonka tiedon perusteella muodostetaan kanta kehittämisen tarpeisiin. Se voi sisältää useita perättäisiä menetelmävalintoja. Nämä valinnat vaikuttavat tavalla tai toisella myös lopputuloksena esitettäviin kannanottoihin. Kehittämistarpeita ei kuitenkaan voida johtaa suoraan arviointitiedosta, vaan arviointitieto tuotetaan tavallisesti juuri päättäjille ja kehittäjille. Heillä on poliittinen vastuu päätöksistä ja niiden toimeenpanosta.

Arvottamisen ohella muitakin arvovalintoja tehdään koko arviointiprosessin ajan. Arviointikriteerit ovat keskeisiä valintoja. Arvioinneissa on olennaista tunnistaa koko arvovalintojen ketju ja tarkastella arvottamisvaiheita kriittisesti myös tästä näkökulmasta.

Jos arvioinnin tavoitteena on aikaansaada välitöntä käytännön kehittämistä, arvot ohjaavat käyttämään kehittävän arvioinnin menetelmiä. Jos taas tavoitteena on tiedon tuottaminen ja ilmiöiden selittäminen, käytetään tiedon tuottamisessa ja analyyseissä tutkimusta ja sen menetelmiä. Jos taas halutaan vain kuvailevaa perustietoa asiasta ilman arvottamista tai kehittämistietoa, riittänee

usein asiaa koskeva selvitystieto. Tämäkin valinta on jo selkeä menetelmävalinta, jonka taustalla on ymmärrys tietotarpeen luonteesta.

Koska arvioinnin ilmiöt voivat vaihdella yksittäisen oppilaan osaamisen arvioinnista (mikrotaso) koko koulutusjärjestelmän tai koulutuspolitiikan arviointiin, on menetelmäkirjokin suuri. Koulutuspoliittinen arviointi on jo makrotason arviointia. Menetelmähaasteita syntyy myös siitä, missä määrin arviointitoiminnan ulkoisuus ja riippumattomuus korostuvat, kehittävyttä painotetaan ja itsearvioinnin aineistoja hyödynnetään sekä siitä, missä määrin arvioitavina olevat kohteet ja niiden parissa työskentelevät ovat jo arviointivaiheessa osallisia prosessissa. Näin ollen metodologiset ja metodiset lähtökohdat voivat vaihdella hankkeittain laajastikin.

Arviointiprosesseja ohjaavia arvovalintoja

Kun organisoimme arviointihankkeet verkostomaisen toimintatavan mukaan, kyse on arvovalinnoista kehittävyden, osallistavuuden, riippumattomuuden ja monitahoisuuden puolesta. Kun valitsemme arviointiryhmään ansioituneita tiedemiehiä, arvostamme luotettavuutta, syvällisyyttä ja totuudellisuutta. Kun valintamme osuu hallinnon parissa tai käytännössä ansioituneisiin asiantuntijoihin, arvostamme tulosten käyttökelpoisuutta päätöksenteossa ja käytännön toimijoiden tarpeita. Käytännössä asiat eivät ole kuitenkaan aivan näin selviä.

Arvioijan arvotietoisuudessa on perimmältään kyse siitä, kuinka hyvin arvioija on tietoinen siitä, millaisen valinnan hän tekee vaihtoehtoisten mahdollisuuksien joukosta. Tällöin kyse myös arvioijan itseymmärryksestä, mitä hän on arvioidessaan tekemässä. Eettiset arvostukset ovat aina jollain tapaa mukana näissä valinnoissa. Eettiset arvot voivat estää arvioijaa käyttämästä arviointia epäoikeudenmukaisesti tai kohdettaan vahingoittavasti.

Arviointikriteerien valinnat ovat arvovalintoja. Jos meillä on esimerkiksi opetuksen ja oppimisen yhteydestä vahvaa tutkimustietoa, käytämme todennäköisesti arviointikriteereinä tekijöitä, jotka johtavat tutkimustiedon mukaan hyvään oppimiseen. Kriteerit saattavat korostaa myös eri tavoin toiminnan tehokkuutta, taloudellisuutta ja erilaisten suoritteiden määrää ja laatua. Kriteereissä voidaan painottaa myös yhdenvertaisuutta, oikeudenmukaisuutta ja laadukkuutta. Arvioijille arvotietoisuus on erityinen pätevyysvaatimus. Se on näkemyksellisyyttä tehdä tarkoituksenmukaisia arvovalintoja erilaisten vaihtoehtojen välillä.

Arvioinnin säädöksissä ei tule selvästi esiin arvottaminen, joka on arvioinnin varsinainen avainprosessi, ei tulosvastuu, ei eettisten tavoitteiden edistäminen eikä valvonnan tai tarkastuksen suorittaminen. Säädösten määrittämä arviointitehtävä viittaa korosteisesti tiedon tuottamisen funktioon. Näin ollen on pääteltävissä, että ulkoisella arvioinnilla hankitun tiedon tai uuden tietämyksen ajatel-

laan ensisijaisesti edistävän vasta tuloksina kehittämistä. Ajatus on kestävätilanteessa, jossa arviointihanke katsotaan käytännössä päättyväksi siihen, kun arviointiraportti on julkistettu. Menetelmällisesti haasteellinen työ liittyy juuri tähän saumaan eli toteutetaanko arviointeja niin, että kehittävä näkökulma tulee mukaan vasta sen jälkeen, kun arvioinnin tulokset ovat valmistuneet vai onko näkökulma mukana koko arviointiprosessin ajan. Näillä ratkaisuilla on merkitystä juuri arviointitoiminnan vaikuttavuuden kannalta.

Arvioinnin teorialähtöisyys

Arviointitoiminta ei ole johonkin tieteenalaan sidoksissa. Se on lähtökohtaisesti monitieteistä. Teorialähtöisyyttä ei kuitenkaan tavallisesti korosteta arvioinnissa. Sen katsotaan kuuluvan tutkimukseen ja arviointitutkimukseen. Arviointi voi olla kuitenkin tutkimuksen kohteena kuten mikä tahansa inhimillinen ilmiö. Arviointitoiminnan laatua parantaa se, jos sitä tehdään tutkimuksen tapaan kurinalaisesti. Tämä tarkoittaa myös sitä, että myös arvioinnin ilmiöitä tulee käsitteellistää. Uuden tiedon perusta on aikaisemmassa tiedossa. Tutkimustieto on kumuloituvaa. Myös arviointitieto voi olla ja sen tulisikin olla kumuloituvaa. Viitekehysten liittäminen arviointeihin lisää mahdollisuuksia kumuloituvan tiedon tuottamiseen.

Jos arvioinnin aineistoa ei käytetä niin, että siitä tehtävien päätelmien käsitteellisyys taso olisi pyrkimys jotenkin kohottaa, ehkä luokittelua lukuun ottamatta, kapenevat arvioinnin mahdollisuudet osua avainasioihin. Käsitteiden huolellinen määrittely on osa arvioinnin laadun perustaa. Nämä valinnat ovat arviointistrategisesti avainvalintoja, ja niillä on seuraamuksia menetelmävalintoihin.

Arvottaminen on tyypillistä kaikessa inhimillisessä toiminnassa, mutta systemaattisena ja professionaalisenä toimintana arvottamiselle on asetettava vaatimuksia, jotka nousevat hallinnon, asioiden käytännöllisen hoitamisen ja teknisen suorittamisen yläpuolelle. Kyse on erityisesti arviointitoiminnan teoreettisuustason tai teorialähtöisyyden vahvistamisesta. Tällä tarkoitetaan muun muassa arviointi-ilmiöiden käsitteellistämistä, arvioinnin käsitteiden tarkastelua toisiinsa sidoksissa olevana käsitteistöinä, arviointi-ilmiöiden oletamusperustan vahvistamista ja arviointia koskevien mallien rakentamista.

Arviointien vaikuttavuus

Arviointien välittömien vaikutusten ja vaikuttavuuden kannalta on tärkeää, miten arvoinnit organisoidaan, miten eri toimijat osallistetaan arviointeihin ja

miten arvioinnin asiakkaat saavat tiedon tuloksista. Kansallisessa ulkoisessa arvioinnissa arvostetaan arviointien verkostomaista organisointia, monitahoista ja kehittävää arviointia. Paikallinen arviointi ja kansallinen ulkoinen arviointi voivat parhaimmillaan tukea toisiaan eri tavoin. Jos itsearviointi on kehittynyt, se luo kysyntää myös ulkoiselle arvioinnille. Ulkoinen arviointi avartaa taas vertailu-ulottuvuuksillaan itsearvioinneilla hankittua tietoa.

Kun arviointiin kuuluu aina arvottaminen ja arvioinnin tarkoituksesta nouseva praktisen kehittämisen vaatimus, avautuu arviointitoiminta menetelmiseen tutkimusta enemmän aikaan, kontekstiin ja asiakkaiden tarpeisiin kytkeytyvänä työnä. Tällöin kehittävyden toteutumista joudutaan pohtimaan kaikkien arvioinnista hyötyä saavien osapuolien näkökulmista. Tilannekohtaiseen tiedonkeruuseen liittyy kuitenkin se pulma, että arviointitieto voi vanheta nopeastikin. Kehittämistarpeet ja yksityiskohtia koskevat painotuserot elävät kaiken aikaa monien tekijöiden seurauksena. Mitä enemmän tulostiedossa jäädytään yksityiskohtien varaan, sitä arvottomampaa tulostieto on myöhempää hyötykäyttöä ajatellen. Tutkivien menetelmin tuotettu arviointiaineisto on kuitenkin soveliasta myös aineiston jatkokäyttöä ajatellen.

Tässä julkaisussa korostuu linjaus, jossa itsearviointi, kehittävä arviointi, verkostomaisuus ja laadunhallinta erilaisine menetelmineen ja mekanismeineen kehittävät yhtäältä itse arviointitoimintaa, mutta antavat myös eväitä kohdetoiminnan kehittämiseen. Arviointimenetelmien tarkoituksenmukaisuutta on myös tarkasteltava siitä näkökulmasta, miten niiden avulla voidaan kehittää kohdetoiminnan tehokkuutta, vaikuttavuutta ja taloudellisuutta. Vaatimus ei ole ongelmaton.

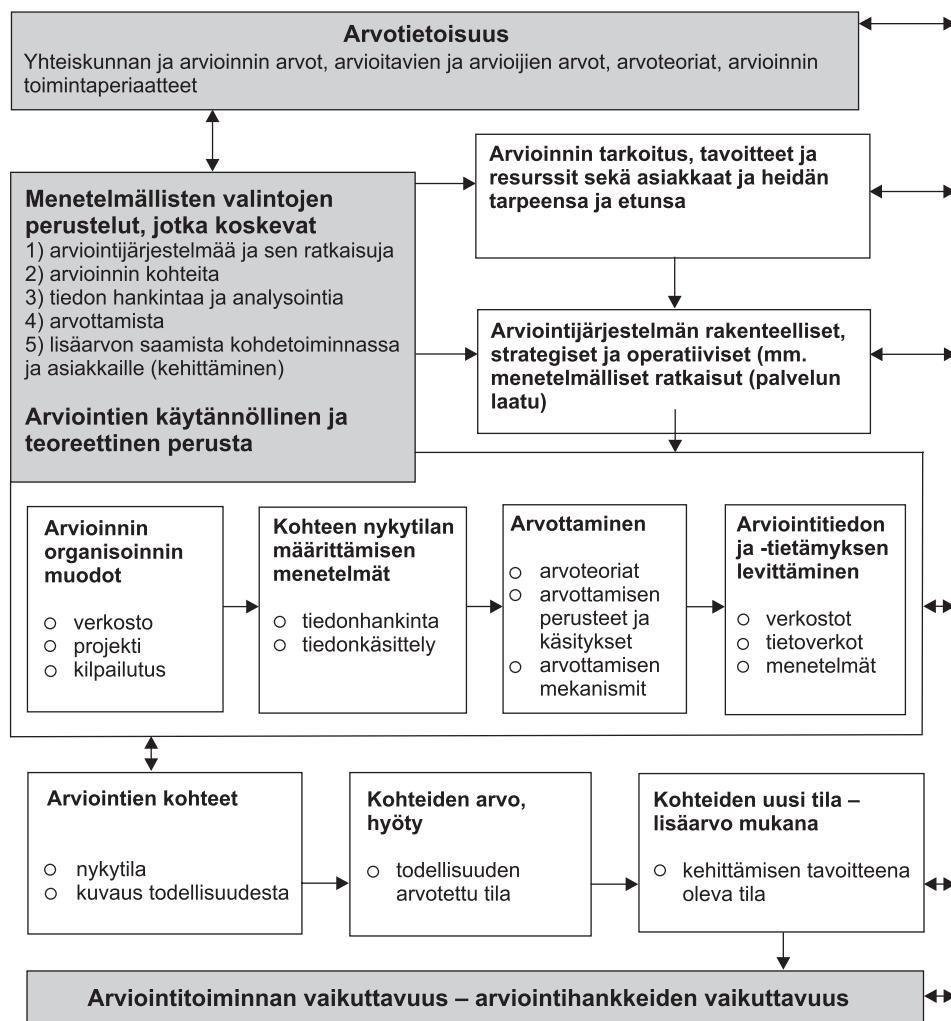
Keskeisiä kansallisen arvioinnin kohteita ovat muun muassa koulutuksen taso, opetuksen laatu ja koulutuksen ohjausjärjestelmän ohjauksen toimivuus. Olennaista on, miten ohjausjärjestelmän osat toimivat yhdessä kokonaisjärjestelmänä (integraatio). Kyse on organisaatioiden toiminnan tehokkuudesta, taloudellisuudesta ja vaikuttavuudesta. Kun arviointikohteiden ala laajenee ja niiden keskinäiset kytkennät lisääntyvät, esimerkkinä toiminnan ja tulosten yhteyksien tarkastelu, on se haaste myös menetelmälliselle kehittämiselle.

Arviointi on osajärjestelmä koulutuksen kehittämisessä ja osajärjestelmä kansallisessa koulutustiedon tuottamisessa. Koulutuksen laadunhallinnan järjestelmät ja tietotuotannon järjestelmät vaikuttavat yhdessä koko koulutusjärjestelmän kehittämiseen.

Arviointitoiminta menetelmäkehittämisen kannalta

Koulutuksen arvioinnin tarkoitus, tavoitteet ja tarpeet ovat arviointitoiminnan lähtökohtia. Ne ovat oleellisia arviointihankkeita organisoitaessa ja arviointitoi-

mintaa kehitettäessä. Niiden taustalla ovat arvot, joiden syvällinen tunteminen on arviointitoiminnan ydintä. Niihin perustuvat arviointitoimintaa koskevat valinnat ja arvottavat päätelmät (ks. kuvio 1). Arvioinnin organisoinnin muodot, arvioinnin kriteerit, arviointimenetelmät ja arviointitiedon levittämisen menetelmät tulisi olla linjassa arviointitoiminnassa noudatettavien julkilausuttujen arvojen kanssa.



Kuvio 1. Arvotietoisuus, teorialähtöisyys ja vaikuttavuus arviointitoiminnan menetelmäkehittämisen lähtökohtina

Arvioinnin teoriaperustan tai teorialähtöisyyden vahvistaminen parantaa edellytyksiä sovittaa arviointityö laajempaan viitekehykseen ja esittää arviointitietoa koskevia väitteitä loogisessa ja systemaattisessa muodossa sekä koetella näiden väitteiden paikkansapitävyyttä. Näitä vaatimuksia voidaan tukea arviointiin kohdistuvalla monitieteisellä tutkimuksella. Tämä tarkoittaa myös arvioinnin, arviointitutkimuksen ja tutkimuksen jatkuvaa monitieteistä vuoropuhelua ja sellaisia menetelmällisiä ratkaisuja, jotka noudattavat tieteellisiä pelisääntöjä.

Yksittäinen arviointimenetelmä voidaan ymmärtää keinoksi tai toimintamuodoksi jonkin arviointihankkeen tarkoituksen ja tavoitteiden toteuttamiseksi. Arvioinnin mahdollisuuksia ja keinoja tulisi tarkastella osana laadunhallinnan keinovalikoimaa. Näin arviointitoiminta voi paremmin täyttää järjestelmäluonteen kriteerit ja edistää koulutuksen vaikuttavuutta systemaattisemmin (ks. kuvio 1).

Seuraavissa luvuissa tarkastellaan kuvioon 1 perustuen arvotietoisuutta, teorialähtöisyyttä ja vaikuttavuutta arviointimenetelmien kannalta. Sen jälkeen tarkastellaan arvioinnin menetelmiä omina kokonaisuuksinaan.

Arvotietoisuus arviointitoiminnan valintojen taustalla

Arvotietoisuus sisältää arvoja koskevaa tietoa, arvioijan oman suhteen arvoihin, tietoisuuden kohdeilmille ominaisista arvoista ja arvioitavaa toimintaa ohjaavista yhteiskunnallisista arvoista. Nämä arvot ovat taustalla silloin, kun arvioija priorisoi valintojaan. Arvottamisen pätevyys nousee arvotietoisuudesta.

Arvotietoisuus ei ole vain tietoa arvoista. Se on myös ymmärrystä siitä, miksi teemme sitä, mitä teemme tai miksi valitsemme sen, minkä valitsemme. Valintojen perusteet ovat avainasioita niin arvioinnissa kuin tutkimuksessakin. Tyyppillisimmillään arvovalinnat liittyvät arvottamisessa käytettäviin kriteereihin ja niiden soveltamiseen. Kohteen arvo tai hyöty, tilanteen hyvyys tai puutteet osoitetaan niiden suhteen.

Ovatko arvot asetettavissa hierarkkiseen järjestykseen? Toiminnassa arvojen ”järjestys” näkyy tavallisesti siinä, mihin panostetaan. Tämä järjestys muuttuu usein kontekstin vaihtuessa. Olennaista on se, että kulloinkin tunnistetaan ja tunnetaan ne arvot, joiden välillä valinta tehdään tai paremmuus todetaan. Tästä prosessista esimerkkinä mainittakoon arviointiryhmissä käytävä arvokeskustelu ja arvoarviointi. Tällaisen prosessin tuloksena voi syventyneen arvotietoisuuden lisäksi olla konkreettisenä tuotoksena läpinäkyvä kuvaus arvioinnin arvovalinnoista. Tämä prosessin vaihe on avain arviointiprosessien ja arvottavien päätelmien ymmärtämiseen.

Intressit ja haluaminen kuvaavat arvojen luonnetta. Toimintaa ohjataan niiden suuntaan – tai toiminta ohjautuu vain ”luonnostaan” johonkin suuntaan

silloin, kun valinnat tehdään tiedostamatta. Arvoilla on arvioijille tavallisesti jokin arvolataus. Tilanteiden, olosuhteiden, toiminnan ja tulosten arvottaminen tapahtuu suhteessa näihin arvoihin. Arvojen läpinäkyvyys arviointiprosesseissa parantaa tulosten uskottavuutta ja luotettavuutta.

Arvottaminen on arvioinnin menetelmällinen prosessi. Sen onnistuminen on arvioinnin laadun avainkysymys. Arvioijien arvottamislogiikan syventäminen, arvottamista koskevan itseymmärryksen parantaminen ja arvottamisen mekaniismien kriittinen tarkastelu kehittävät arviointien laatua ja vaikuttavuutta. Kriittikin kohteena voi olla esimerkiksi tulkinnallisen päättelyn pätevyys. Tällöin voidaan kysyä sitä, miksi eri arviointiryhmien päätelmät poikkeavat toisistaan samassa kohteessa.

Teorialähtöisyyden vahvistaminen arvioinnin kehittämässä

Arviointien pätevyuden parantamiseksi on tärkeää rakentaa arvioinnille teoreettisesti vahva arviointikehys, joka kertoo asianosaisille, missä käsitemaailmassa liikutaan. Arviointikehys ilmaisee selkeästi esimerkiksi sen, mitä matematiikan osaamisella kyseisessä arvioinnissa tarkoitetaan ja miten arviointikohde jäsennetään niin, että sitä voidaan systemaattisesti, pätevästi ja luotettavasti arvioida.

Arviointeja varten tieto on hankittava tieteessä noudatettavin toimintaperiaattein ja tutkivin menetelmin. Vaatimus on sitä tähdellisempää, mitä suurempia odotuksia asetetaan arviointien laadulle. Kyse on toimintaprosessien laadullisesta kurinalaisuudesta. Sillä edistetään arviointitiedon laajempaa ja pitkäkestoisempaa käyttöä.

Arviointimetodologialla voidaan tarkoittaa arvioinnin menetelmien tutkimusta ja arviointia ohjaavien yleisten periaatteiden tutkimusta. Metodologinen työ tukee näin ollen arvioinnin teoriaperustaisuutta ja laadun kehittämistä. Arvottaminen voi olla usein haavoittuvain arvioinnin elementti siksi, että esimerkiksi jatkuvilla tulkinta- ja neuvotteluprosesseilla voidaan saavuttaa kannanottoja, jotka hyväksytään lähinnä vain asianomaisten arvioijien piirissä. Arvovalintojen paremmuudelle ei ole kuitenkaan objektiivista perustaa.

Metodologisesti olisi tarpeellista vertailla myös erilaisten menetelmällisten lähestymistapojen käyttöä arvioinneissa. Mitä lisäarvoa itsearviointiaineistojen käyttö antaa kansallisessa ulkoisessa arvioinnissa, riippuu siitä, miten nämä aineistot menetelmällisesti kerätään ja miten varmistutaan niiden uskottavuudesta. Milloin taas pitkittäisarviointien käyttö on perusteltua? Mitä vaatimuksia niiden aineistojen keruulle on asetettava? Miten voimme parantaa mahdollisuuksia hyödyntää eri aikoina kerättyjä arviointiaineistoja meta-analyysejä varten? Milloin arviointiaineistot tarjoavat edellytyksiä syy- ja seuraussuhteiden tarkasteluille?

Tiedon saamisen mahdollisuudet kohteesta, sen hankkimisen kustannukset ja käytettävissä oleva aika sanelevat myös arviointikysymysten muotoilun ja näin ollen määräävät usein myös menetelmällisen ratkaisun. Strategisesti tärkeä seikka onkin tietää, millaista tiedon järjestelmää arviointitiedon tuottamisella rakennetaan ja miten se on yhteydessä muuhun koulutustiedon tuottamiseen.

On hyödyllistä tarkastella monien eri tieteenalojen näkökulmaa arviointitiedon tuottamisessa. Miten sosiologinen ja psykologinen näkökulma eroavat kouluhyvinvoinnin tai koulutuksen tuottaman inhimillisen pääoman arvioinnissa? Arvioinnin metodologiset intressit voivat näin vaihdella hyvin laajoista tieteidenvälisistä näkökulmaeroista spesifeihin metodisiin kysymyksiin. Tarvitaan jäsentävää perustyötä sen suhteen, minkä näkökulman, lähestymistavan tai konseptin mukaan arvioinneissa kulloinkin edetään.

Arviointitoiminnan teoriaperustaa vahvistettaessa tarkasteltavat tekijät voidaan ryhmitellä a) järjestelmätekijöihin, b) kohdetekijöihin, c) tiedon hankkimista koskeviin, d) tulosten arvottamista koskeviin ja e) tiedon ja tietämyksen levittämistä koskeviin tekijöihin. Kaikkiin tekijäryhmiin liittyy metodologisia ja metodisia erityispiirteitä ja arvovalintoja.

Arviointia koskevat tekijäryhmät muodostavat kokonaisuuden, jota voidaan tarkastella toimintaympäristönä, jota tulee tutkia, arvioida ja jota tulee kehittää. Arvioinnin näkökulmasta tarkastellen kyse on pyrkimyksestä edistää ajattelultaan esimerkiksi kriittisemmän ja systemaattisemmän toimintakulttuurin luomista arvioinnin piirissä.

Koska arviointitoiminnan kohteiden kirjo kattaa lähes kaiken inhimillisen toiminnan, ei aina ole tiedossa tieteellisiä keinoja tai välineitä, joilla ilmiötä voitaisiin mielekkäästi empiirisesti lähestyä. Arvioinnin metodologinen ja metodinen kehittäminen kaipaavat myös näihin tarpeisiin osuvia innovaatioita.

Arviointeihin voidaan kytkeä myös teoreettista taustatietoa ja aikaisempaa tutkimustietoa viitekehyyksi. Tämä lähtökohta edellyttää tavanomaista syvällisempää ymmärrystä kohdeilmiöstä, sen vaatimuksista, siihen liittyvästä aikaisemmasta tiedosta ja lisäksi usein kohdeilmiön käsitteellistämistä. Kun arviointitietoa hankitaan luottamalla liikaa esimerkiksi itsearviointeina tuotettuihin tietoihin, kehityskeskustelutietoihin, epäyhtenäisin formaatein tuotettuihin portfolioihin ja laatujärjestelmien tuottamaan kuvaustietoon vailla laatutakeita, voi tiedon luotettavuus olla jo lähtökohtaisesti kyseenalainen. Kyseisten tietojen alkuperäisen tuottamisen funktio on ehkä ollut toisaalla – ei kansallisen ulkoisen arvioinnin käyttöä ajatellen. Tietojen tuottamisen tavat, lähtökohdat ja menetelmät voivat olla itsearvioinneissa alttiina monille virhelähteille, jos tiedon hankintaa ei ole rakennettu yhtenäisen viitekehyyksen varaan eikä tietoa ole tuotettu tutkivin menetelmin. Tässä suhteessa tutkijoilla ja arviointitutkimuksella olisi annettavaa arviointitoiminnan kehittämiseksi.

Arviointien vaikuttavuus metodisten valintojen taustalla

Koulutuksen lainsäädäntöuudistuksessa painottuvat kansallisen arviointijärjestelmän kehittämisen lähtökohtina 1) tehokkuus, 2) taloudellisuus ja 3) vaikuttavuus. Näitä dimensioita yhdessä voidaan kuvata tuloksellisuuden käsitteellä. Tässä julkaisussa ei oteta kantaa arvioinnin käsitteiden hierarkiaan, mutta otetaan kantaa yksittäisiin keskeisiin arvioinnin käsitteisiin, joita käsitellään luvussa 5. Käsitteiden valintaa, sisältöä ja alaa kirjoittajat ovat tarkastelleet myös yhdessä, vaikka eri kirjoittajat ovatkin yksittäisten käsitelmäärittelyjen takana.

Arviointi-ilmiöiden käsitteellistämällä, käsitteiden muodostamisella ja käsitteiden määrittelyllä luodaan perustaa arvioinnin menetelmien kehittämiseksi, arvioinnin teoriaperustaisuudelle ja vaikuttavuudelle. Arvioinnin käsitteiden avaamisella, luokittelulla ja käsittehierarkioiden luomisella rakennetaan arviointitoimintaa, arviointitutkimusta tai käyttäytymistieteiden tutkimusta ylittävää yhteistä teoriaa. Toistaiseksi arviointia koskevien käsitteiden käyttö on ollut horjuvaa.

Metodologisesti ja metodisesti on haasteellinen kysymys, miten arviointi parantaa koulutuksen laatua (arviointitoiminnan/arvioinnin vaikuttavuus) ja mikä rooli arvioinnilla on koulutuksen yhteiskunnallisen vaikuttavuuden parantamisessa. Tämä vaikuttavuus olisi tyypillisesti monitieteistä tutkimusta edellyttävä interventio asian selvittämiseksi.

Vaikuttavuuden käsitettä voidaan tarkastella esimerkiksi organisaation sisäisen tuotoksen (esim. oppimistulosten) suhdetta sen ulkopuoliseen kriteeriin (esim. työelämän tarpeisiin tai jatko-opintokelpoisuuteen). Kysymys on siirtovaikutuksesta, ja organisaation tuotoksen käyttö- ja vaihtoarvosta markkinoilla. Vaikuttavuuden käsite antaa myös keinon mitata palvelun tuottajan ja asiakkaan kohtaamisen laatua. Toisin kuin muut tuloksellisuuskäsitteet (tehokkuus, taloudellisuus, tuottavuus) vaikuttavuus siirtää tarkastelun organisaation sisäisestä palvelukyvyistä ulkoiseen palvelukykyyn ja palvelun itsensä laatuun. Ei riitä, että tekee asiat oikein; on tehtävä myös oikeita asioita.

Vaikuttavuusarvioinnin kriteerinä on joko yksilöllisen tai yhteisön (asiakasvaikuttavuus) tai kollektiivisen (yhteiskunnallinen vaikuttavuus) hyvinvoinnin lisääntyminen. Tällöin arvioinnin lähtökohtana on hyödyllistä pitää esimerkiksi koko ulkoisen arviointitoiminnan vaikuttavuutta pitkällä aikavälillä esimerkiksi joidenkin avainindikaattorien suhteen. Arviointiprosessien, -tulosten ja -päätelmien korkealaatuisuus eivät itsessään vielä takaa arviointitoiminnan vaikuttavuutta. Tällöin on siis pyrittävä arvioimaan/tutkimaan kaikkien arviointihankkeiden kokonaisuuden ja arviointitoiminnan tukipalvelujen yhteisvaikutuksia.

Yksittäisissä arviointihankkeissa merkitykselliseksi voidaan katsoa muun muassa seuraavat asiat 1) miten hyvin arviointi täyttää arvioinnin tarkoituksen ja tavoitteet sekä tavoittaa kohteeseen liittyvät tarpeet 3) miten hyvin arviointi-

suunnitelma on perusteltu, 3) miten tärkeäksi arvioinnin tarkoitus on osoitettu, 4) miten luotettavaan aineistoon arvottaminen perustuu, 5) ovatko tehdyt analyysit perusteltavissa, 6) miten riippumattomia ja uskottavia tulokset ja arvottavat päätelmät ovat ja 7) miten monipuolisesti ja tehokkaasti tulokset saatetaan asiakkaiden käyttöön ja tietoon? Kyse on tällöin muun muassa arviointitiedon levittämisen menetelmistä. Niihin kuuluu myös palautetiedon antaminen koulutuksen järjestäjille ja oppilaitoksille, joilta tietoa on kerätty. Nämä kysymykset ovat avainasioita silloin, kun tarkastellaan hanketoiminnan vaikuttavuutta.

Koulutuksen ”tuloksia” saatetaan usein tarkastella vain sellaisena palautetietona, joka koskee esimerkiksi asiakastytyväisyyttä. Oppimistuloksia mitataan koulusaavutusmittarein, testein, opinnäyttein tai hyvälaatuisin kokein, joihin kuuluvat myös näyttökokeet. Oppimistuloksia voidaan tarkastella muun muassa yksilöiden, yhteisöjen ja organisaatioiden sisällä ja välillä sekä tulosten (hyötyjen) alueellisia vaikutuksia elinkeinoelämään ja hyvinvointiin. Tällöin vaikutukset voivat olla monenlaisia (esim. odotetut ja odottamattomat, lyhyt- ja pitkäaikaiset). Yksilövaikutukset voivat koskea tiedon karttumisesta oman toiminnan muutoksiin, organisaatiossa vaikuttamisen kehittymiseen ja elämäntehtävän muotoutumiseen saakka. Edellä todetut vaikutukset ja vaikuttavuus kietoutuvat erittäin komplisoidulla tavalla toisiinsa.

Arvioinnin vaikuttavuus syntyy järjestelmällisen ja pitkän työn, arvojen kannalta johdonmukaisen toiminnan ja sen toiminnan hyvän laadun seurauksena. Ne luovat asiakkaiden ja sidosryhmien luottamusta arviointitoimintaa kohtaan, mikä on edellytys myös sille, että arviointitoiminnalle syntyy uusia tarpeita, arviointia arvostetaan ja että sillä on sekä asiakas- että yhteiskunnallista vaikuttavuutta.

Arviointijärjestelmän kehittäminen

Arviointimenetelmien kehittämisen tavoitteena on myös tukea tietoperustaisen ja rationaalisen päätöksenteon mallia, jossa arviointitiedon merkitys (tieto, oppiminen) saa keskeisemmän roolin koulutuspoliittisessa ohjauksessa, päätöksenteossa ja paikallisen koulutuksen kehittämisessä. Paikallinen kehittäminen kattaa myös oppilaitosten työn kehittämisen, opetuksen kehittämisen ja oppimisen jatkuvan parantamisen.

Arviointimenetelmien kehittäminen on osa koko arviointijärjestelmän kehittämistä. Tällöin menetelmien kehittäminen on yhteydessä siihen, millaiset arvot, toimintaperiaatteet ja arviointijärjestelmää koskevat päätökset ovat lähtökohtina. Kyse on ensisijassa lainsäädännöstä, joka säätelee arviointia. Arviointijärjestelmän metodisessa kehittämisessä on tärkeää analysoida myös seuraavia pääkysymyksiä:

- 1) Mitä poliittinen päätöksentekojärjestelmä odottaa arvioinnilta?
- 2) Millaisiin teoreettisiin lähtökohtiin arviointitoiminta ja arviointihankkeet nojaavat?
- 3) Mitä, miksi, miten ja milloin arvioimme? Millaista tiedon järjestelmää arvioinnit palvelevat?
- 4) Mitkä erot, riippuvuudet, trendit, syy- ja seuraussuhteet ovat merkittäviä selvitetäviksi, analysoitaviksi, arvoitettaviksi ja seurattaviksi?
- 5) Miten varmistumme arviointitiedon uskottavuudesta, totuudellisuudesta ja vaikuttavuudesta?
- 6) Miten kehitämme arviointeja osana a) arviointijärjestelmän, b) koulutuksen tietotuotannon ja b) koko koulutusjärjestelmän kehittämistä?
- 7) Milloin uusi tieto tai tietämys on riittävää kertomaan, mitä, miten ja milloin pitäisi kehittää kohdetoiminnassa, koulutusjärjestelmässä ja yksittäisten toimijoiden tasoilla?

Suomen arviointijärjestelmä on nykyisellään kevyt. Arvioinnin lähestymistapa on ollut hallinnollinen ja pragmatistinen. Ajatus on ollut usein se, että arviointitieto hankitaan päätöksenteon ja kehittämisen akuutteja tarpeita varten ja arviointi kehittämistoiminteen pyritään rajaamaan ja kohdentamaan sinne, missä tarve kulloinkin on huutavin. Arviointihankkeita meillä organisoidaan muun muassa seuraavilla tavoilla:

- 1) verkostomainen toimintatapa
- 2) projektiorganisaatio (hankekohtaisesti esim. laitoksen sisäisenä projekti-ryhmänä)
- 3) hanke annetaan kilpailutuksen perusteella parhaan tarjouksen tehneelle hakijalle
- 4) arvioinnit toteutetaan itsearviointeina.

Olenainen kysymys on se, miten jo arvioinnin organisoinnilla voidaan tukea kehittämistä. Arviointitoiminnan organisointi vaikuttaa arviointien suunnitteluun, toimeenpanoon, raportointiin, hanketulosten kytkemiseen muuhun arviointitietoon ja tulosten levittämiseen. Pääpainon ollessa arviointien prosessimaisessa toteutuksessa ja asianosaisten osallistamisessa, saattaa arviointitiedon jatkokäytön mahdollisuudet olla vähäisiä – niillä voi joskus olla jopa kertakäyttöisyyden luonne. Kerätty tieto ei kumuloidu eikä tarjoa mahdollisuutta meta-analyyseihin tai meta-arviointeihin. Tilanne edellyttää arviointijärjestelmän toimivuuden kokonaisvaltaista kehittämistä ja sen tuottaman hyödyn toistuvia arviointeja.

Arviointimenetelmien kehittäminen

Arviointimenetelmien kehittämiseen liittyy olennaisesti se taustatyö, jolla arvioinnin ilmiötä tarkastellaan teoreettisen viitekehyksen puitteissa, analysoidaan ilmiötä, käsitteellistetään ja luodaan malleja arviointitoiminnan jäsentämiseksi ja kehittämiseksi. Arvioinnin tutkimus on osa tätä työtä. Arviointitoiminnan menetelmällisen kehittämisen tavoitteena on myös

- 1) arviointikäsitteistön selkiyttäminen ja vakiinnuttaminen
- 2) arviointiprosessien ja -tiedon hyödyntämisen parantaminen
- 3) suurten aineistojen hallinnan ja käsittelyn (mm. meta-analyysit ja -arvioinnit, pitkittäisanalyysit) toimintamallien ja niitä palvelevien käyttöteknologioiden (esim. arviointiportaali) kehittäminen
- 4) ennakoinnin menetelmällinen kehittäminen
- 5) arviointityökalujen parantaminen vertailukehittämisen tarpeisiin
- 6) arviointien vaikutusten ja vaikuttavuuden tutkimuksen kehittäminen
- 7) monimenetelmäisten ja -tieteisten näkökulmien lisääminen arvioinnissa sekä sektorirajat ylittävien arviointistrategioiden luominen.

Menetelmällisen kehittämisen ydinaluetta on arvottamista koskevan osaamisen parantaminen. Kysymys on lähinnä kolmesta asiasta 1) kohteen tilaa koskevan luotettavan ja monipuolisen tiedon hankkimisesta, 2) kohteen arvon tai hyödyn määrittämisen läpinäkyvyyden ja pätevyyden lisäämisestä ja 3) arvotetun evidenssiaineiston käytön lisäämisestä kehittämisen ja päätöksenteon ohjaamisessa. Kehittämissuosituksen esittäminen on jo harppaus arvioinnin sallimasta evidenssistä vapaamman tulkinnan suuntaan. Niiden esittämistä voidaan kuitenkin perustella siksi, että asiakkaat arvostavat niitä ja arvioija tuntee yleensä parhaiten sen taustan, josta suositukset nousevat. Päätöksenteko ja toimeenpano jäävät kuitenkin kehittäjille itselleen.

Arviointi kehittämisen välineenä

Perustan arvioinneille antaa tieto, joka kohteesta on tuotettu. Tällä tarkoitetaan muun muassa tietoa määristä, eroista, yhteyksistä ja trendeistä. Kyse voi olla esimerkiksi tilasto-, arviointi- ja tutkimustiedosta. Jos ilmiöstä on saatu syy- ja seuraussuhteita koskevaa tietoa, se on kehittämisen kannalta hyödyllistä tietoa. Niin sanotuissa hyvissä käytännöissä ei ilman tutkimusta välttämättä pystytä identifioimaan niitä tekijöitä, jotka ovat hyvyyden takana. Tulostietoja koskevat päätelmät liittyvät normaalisti tilastolliseen tai loogiseen päättelyyn.

Arvottamisen kannalta on tärkeää, että tulostietoa ja näyttöjä on riittävästi päätelmien tekoa varten. Jos arvottamisessa ei löydy perusteita niin sanotuille tasopäätelmille, saatetaan tyytyä toteamaan ne asiat, jotka arvioijien mielestä ovat kunnossa ja vastaavasti ne asiat, jotka kaipaavat kehittämistä. Yksilöidymppi kehittämistarpeiden osoittaminen, puhumattakaan kehittämissuosituksista, edellyttää arvioijilta kohteen monipuolista käytännöllistä tuntemusta. Vasta parantamistoimien seurauksena arviointitoiminnan hyödyt muuttuvat näkyviksi.

Prosessina arviointeihin voi kuulua a) vertailuja soveltuviissa kohteissa yksiköiden, organisaatioiden ja järjestelmien välillä, b) sertifiointeja, joiden nojalla voidaan antaa todistus toiminnan laadusta, c) akkreditoiteja, joihin voi sisältyä instituution aseman oikeutuksen arviointi tai d) auditointeja, jotka sisältävät muun muassa tutkintaa siitä, miten prosessit ja tulokset vastaavat suunnitelmia (esim. laatujärjestelmän toimivuuden systemaattinen arviointi). Arviointien toiminnalliset kontekstit voivat vaihdella laajasti, mutta niiden prosessien erot syntyvät lähinnä arvioinnin funktioista. Edellä todetuissa arvioinnin menetelmällisissä ratkaisuisa kohdetoiminnan kehittäminen on ensisijaista.

Arvioinnin menetelmiä on hyvä tarkastella myös yksittäisiä arviointimenetelmiä laajemmasta näkökulmasta. Tällöin menetelmillä voidaan tarkoittaa:

- 1) tiedonhankinnan ja -analysoinnin menetelmiä
- 2) tulosten arvottamisen menetelmiä
- 3) arvioinnin kehittämisen menetelmiä
- 4) arviointitiedon levittämisen menetelmiä
- 5) kohdetoiminnan laadunhallinnan menetelmiä.

Edellä kuvatun kokonaisuuden puitteissa arviointitoiminnan menetelmällinen kehittäminen voi tapahtua rinnan asianomaisten tieteenalojen menetelmäkehittelyn kanssa ja arvioinnin piirissä tapahtuvana omana metodologisena työnä sekä omana innovatiivisena kokeilu- ja kehittämistyönä.

Kehittävän arvioinnin menetelmille on yhteistä tietoinen pyrkimys kohdetoiminnan kehittämiseen jo arvioinnin prosesseissa (ks. edellä kohta 5). Se voi tarkoittaa esimerkiksi prosesseihin osallistumalla oppimista ja toisilta oppimista. Kehittävän arvioinnin prosesseihin liittyy myös tiedon hankintaa, analysointia ja arvottamista joko itsearviointin ja / tai ulkoisin arvioinnin keinoin. Prosessien ajatellaan aiheuttavan oppimista, joka on kehittämisdynamiikan käyttövoima.

Laatujärjestelmät ovat laadunhallinnan välineitä, joilla pyritään kokonaisuutena arvioimaan esimerkiksi toimintayksikön panoksia, prosesseja ja tuloksia ja tekemään niistä tiedoista johtopäätöksiä kehittämisen kannalta. Vertaisarviointeissa on mukana esimerkiksi samanlaisten yksiköiden vastavuoroisuus arviointaessa ja yhteistoimintaa työn kehittämisessä (mm. parivertailut). Benchmarking liittyy vertailukehittämiseen tai -oppimiseen, jolloin taustalla ovat hyvät käytän-

nöt ja esikuvat, joihin omaa tilannetta verrataan. Mukana voi olla arviointia ja palautteen antoa vertailukohteelle.

Monitahoarvioinnissa on mukana eri tahojen näkemyksellisyys prosessin eri vaiheissa, mutta prosessi voi sisältää myös useiden arviointimenetelmien samanaikaista soveltamista.

Portfolio soveltuu hyvin itsearviointiin, mutta myös ulkoiseen arviointiin. Portfolio tekee näkyväksi myös oppimisprosessia ja kehityksen suuntaa. Kehittävän arvioinnin menetelmät muodostavat kokonaisuuden, jossa on tärkeää paikantaa erityisesti ne elementit, mitä arvioidaan, miten ja millä seuraamuksella.

Tavoite tuottaa jopa yksikkötasolle räätälöityä tietoa ei vähennä yleistettävän tiedon ja monipuolisen tiedon tuottamisen tarvetta. Se on kehittämisen kannalta merkityksellistä. Kehittävässä arvioinnissa tuotettu tieto pyrkii jo lähtökohteisesti edistämään muun muassa yksilöllistä, yhteisöllistä ja organisaatioiden oppimista. Tällöin rajan vetäminen kehittämisen ja arvioinnin välille ei ole aina itsestään selvä.

Arvioinneissa tyydytään tavallisesti toteamaan, ovatko toiminta ja tulokset asetettujen tavoitteiden mukaisia tai kuinka hyvin asetetut tavoitteet on saavutettu. Tavoitteiden hyvyyden tai tärkeyden arviointi (tehdäänkö oikeita asioita) on taas oma metodinen pulmansa esimerkiksi siksi, että organisaatioiden asettamat tavoitteet ovat toiminnan virallisia lähtökohtia eikä prosessien onnistumista voida arvioida joistakin muista lähtökohdista. Arvioinnin intentio ei kuitenkaan voi olla arvokeskustelusta pidättäytyminen. Arvottaessaan arvioijat osallistuvat aina arvoja ja niiden merkittävyyttä koskevaan keskusteluun.

Hankittavan arviointitiedon tarpeet voivat ovat hetkellisiä ja koskea hyvin rajattua kohdetta tai ne voivat olla myös laajoja tila-arviointeja (esim. koulutuspoliittisen tahdon toteutuminen). Tietotarpeet määräävät myös sen, tavoitellaanko palautetietoa, seurantatietoa, selvitystietoa, tilastotietoa, arviointi- vai tutkimustietoa. Tiedon keruu on aina sidoksissa myös siihen, mitä tietoa on jo olemassa ja mitä tietoa kerätään samanaikaisesti muiden tahojen toimesta. Tämä tarkoittaa koko arviointijärjestelmän kehittämistä siltä kannalta, miten kansallisen tietotuotannon eri osajärjestelmät täydentävät toisiaan.

V

Koulutuksen arvioinnin peruskäsitteitä



Koulutuksen arvioinnin peruskäsitteitä

Seuraavaksi lyhyesti määriteltävät arvioinnin peruskäsitteet on tarkoitettu apuvälineeksi niille, jotka työssään tai opinnoissaan tarvitsevat tietoa arvioinnin terminologiasta. Monet tämän kirjan artikkelien kirjoittajista ovat osallistuneet myös käsitteistöä avaavan osuuden laadintaan. Kirjoittajat ovat yhdessä tuumin ottaneet kantaa terminologian linjauksiin sekä käsitteiden kuvauksiin.

Arvioinnin metodologia (methodology of evaluation)

Arvioinnin metodologialla tarkoitetaan suppeassa mielessä arvioinnin menetelmien, arvottamisen menettelytapojen ja toimintaperiaatteiden kuvausta ja tutkimusta. Metodologiset kysymykset koskevat myös arvioinnin arvoja ja olettamusperustaa. Metodologinen työ voi sisältää myös edellä mainittujen kohteiden ja arviointitiedon tarkoituksen ja tavoitteiden kriittistä erittelyä ja tarkastelua.

Arviointi (evaluation)

Suppeassa merkityksessä arviointi voi tarkoittaa suorituksen tason mittaamista, tavoitteiden ja saavutettujen tulosten välisen suhteen vertaamista, esimerkiksi oppimistulosten arviointia (assessment, grading). *Laajemmassa merkityksessä* arviointi ulotetaan koskemaan koulutuksen tai minkä tahansa inhimillisen toiminnan panoksia, prosesseja ja tuotoksia sekä arvioinnin tavoitteista riippuen myös panoksien, prosessien ja tuotosten välisiä suhteita (evaluation).

Arvioinniksi kutsutaan toimenpiteitä, joiden avulla kohdetta koskevan tiedon (panokset, prosessit ja/tai tulokset) tai kohteen tilan arvo, merkitys tai hyöty määritetään asetettuun mittatikkueen eli suhteuttamisperustaan nähden. Arvon määrittäminen voi tapahtua muun muassa asetettujen tavoitteiden, kriteerien sekä konteksti- ja trenditiedon suhteen. Arvottamisprosessiin voi liittyä sekä arvottamisen että arviointia ohjaavien arvojen, tarkoituksen ja toimintaperiaatteiden pohdintaa.

Arvioinnilla voidaan tuottaa tietoa jo *tapahtuneesta* (ex post facto), *meneillään olevista prosesseista* (ex nuc) ja *suunnitellun toiminnan tulevista vaikutuksista* (ex ante). Evidenssi voi koskea muun muassa määriä, eroja, selityksiä ja syitä. *Arvioinnin tehtävä* voi olla toteavaa (esim. päätösten seuraamuksiin liittyvä hallinnollinen funktio), formatiivista (kehittävää, oppimista tukevaa), summaatiivista tai prognostista. *Luonteeltaan* arviointi on konteksti- ja aikasidonnaista käytännön kehittämistä ja oppimista auttavaa ja tukevaa.

Arviointikriteeri (evaluation criterion, assessment criterion)

Kriteeri on vertailuperuste, mittapuu arvioitavan ilmiön tai ominaisuuden määrittämiseksi ja erottelunsa. Kriteeri on siis peruste, johon tukeutuen kohteen arvon määrittävä johtopäätös tehdään. Kriteeri on tavoitteen saavuttamisen taso tai toiminnan tila siten, että sen perusteella voidaan tulkita, miten hyvin tavoite on saavutettu.

Arviointimalli (evaluation model)

Arviointimalli on yksinkertaistettu kuvaus koko arvioinnista lähtien arviointikohteesta, tiedonhankinnasta ja menetelmävalinnoista päättyen raportointitapaan. Arviointimalli kuvaa arviointiprosessin etenemisen vaiheineen.

Arviointimenetelmä (evaluation method)

Arvioinnin menetelmillä voidaan tarkoitaa arvioinneissa käytettävän tiedon 1) hankinnan menetelmiä, 2) tiedon analysoinnin menetelmiä ja 3) tulostiedon arvottamisen menetelmiä. Arviointimenetelmät voidaan ymmärtää suppeammasa mielessä keinoiksi tuottaa systemaattisesti ja luotettavasti kohdetta koskevaa arviointitietoa.

Laajemmassa merkityksessä arviointimenetelmistä puhutaan myös silloin, kun tarkoitetaan kaikkia niitä keinoja ja ratkaisuja, joilla arvioinnin tavoitteet voidaan toteuttaa. Tällöin on kyse esimerkiksi interventioista, joilla arvioinnin kehittävä funktio toteutuu osin jo arvioinnin prosessien avulla tai arviointitoi-

minnan organisointi verkostomaisesti antaa arviointitiedon levittämiseksi lisää mahdollisuuksia

Arviointiparadigma (evaluation paradigm)

Arviointiparadigma tarkoittaa arvioinnissa noudatettavia ja sitä määrittäviä sääntöjä ja periaatteita, jotka perustuvat ontologisiin, epistemologisiin ja metodologisiin käsityksiin. Paradigma sisältää todellisuuden ja tiedon sekä tietämisen luonnetta ja ihmistä ja ihmisen toimintaa koskevia käsityksiä. Paradigma perustelee, mikä on tärkeää, järkevää ja legitiimiä. Käsitys oikeanlaisesta ja järkevästä arvioinnista muuttuu, kun vallitseva arviointiparadigma vaihtuu toiseen paradigmaan. Arvioinnin eri tavoin näkeville koulukunnilla on erilaiset arviointiparadigmansa. Kun useilla paradigmoilla on omat kannattajakuntansa, ne kilpailevat keskenään järkevimmän ja oikeutetuimman paradigman asemasta.

Arviointitoiminta

Arviointitoiminta voidaan ymmärtää kokonaisuudeksi, joka sisältää kaiken toiminnan arviointijärjestelmän sisällä, mutta myös järjestelmän luomisen ja kehittämisen. Arviointitoimintaan kuuluu myös yhteistyö arviointeihin osallistuvien asiantuntijoiden ja osapuolien kanssa, arviointihankkeiden organisointi, arvioinnin menetelmien, prosessien ja mittareiden kehittäminen sekä arviointitiedon tuottamisesta vastaaminen kokonaisuutena. Tällöin arviointitoimintaan liittyy myös hallintoa ja arviointitiedon ja -tulosten levittämistä.

Arviointitutkimus (evaluation research)

Arviointitutkimuksessa kohteena olevasta ilmiöstä, joka on usein tilaajan käytännöllisiä tarpeita ja päätöksentekoa palveleva, hankitaan tietoa, jota analysoidaan tieteellisin menetelmin, määritetään tarkasteltavan kohteen tai ilmiön arvo tai hyöty ja pohditaan erilaisten ratkaisumallien etuja ja haittoja erilaisissa tilanteissa.

Arviointitutkimuksen tulokset palvelevat tavallisesti päätöksentekoa ja käytäntöä. Sen prosesseille on kuitenkin ominaista tutkimuksellisen tiedonhankinnan kurinalaisuus ja analyyseissä on mukana myös yhteyksien etsiminen ja mallintaminen. Arviointitutkimusta voidaan ajatella arvioinnin ja tutkimuksen rajapintana, jossa ongelmiin tai arviointitehtäviin vastaaminen tuottaa arviointia enemmän yleistettävää, yhteyksiä tarkastelevaa, hypoteeseja rakentavaa sekä tuloksia ja ratkaisumalleja kriittisesti tarkastelevaa tietoa.

Arvo (value)

Arvoiksi sanotaan asioita tai asiantiloja, joita pidetään arvokkaina, joita toisin sanoen arvostetaan, arvotetaan positiivisesti. Puhuttaessa arvoista käytetään laatusanaa ”hyvä”, ja se vertautuu ”huonoon” (esimerkiksi moraalista puhuttaessa ”pahaan”). Arvottava suhtautuminen maailmaan on ihmisen perustaipumus. Arvottamisen näkökulmasta voidaan maailmaa luokitella esimerkiksi taloudellisiin, moraalisiin ja esteettisiin arvoihin. Pedagogisia arvoja ovat kasvatuksessa kulloinkin hyvänä pidetyt tavoitteet. Kirjallisuuden lukuisissa määritelmissä arvon käsite liitetään yksilön tai ryhmien intresseihin, haluihin, tarpeisiin, mielityksiin, toivomuksiin, odotuksiin, arvotunteisiin, sisäistettyihin preferensseihin, toiminnan päämääriin, valintataipumuksiin tai toiminnan tulosten arviointikriteereihin. Arvot eivät ole yksilöiden mielivaltaisen arvottamisen tuloksia, vaan niiden perusta on yhteisöjen, yhteiskuntien ja kulttuurien perinteissä ja käytännöissä. Arvoja omaksutaan yhteisöjen jäsenyyden kautta. Arvot ohjaavat ihmisten valintoja, toimintaa ja suhtautumista ympäröivään maailmaan.

Arvoluokituksia on mahdollista tehdä monilla erilaisilla tavoilla. Yksi mahdollinen jako on jo Aristoteleelta tuttu jako itseisarvoihin ja välinearvoihin. Itseisarvoiksi sanotaan asioita, joita halutaan niiden itsensä vuoksi, kun taas välinearvot ovat askelmia johonkin muuhun tavoitteeseen. Esimerkiksi useimmille raha on välinearvo. Arvohierarkialla tarkoitetaan arvojen järjestystä. Esimerkiksi itseisarvot sijoitetaan arvojärjestyksessä korkeammalle kuin välinearvot.

Arviointi ei voi koskaan olla irrallaan arvoista. Arviointiin liittyvien valintojen taustalla on aina arvoja, joita arvioitsija on oppinut arvostamaan. Arviointikohteen tilaa verrataan arvottamisessa johonkin toivottavaan tai tavoitteeksi asetettuun asiantilaan, esimerkiksi tasa-arvoon tai oikeudenmukaisuuteen. Arvottaminen merkitsee vertailussa hierarkkisen aineksen mukaantuloa, jolla ilmaistaan suosimista, arvonantoa, hyväksyvää myöntämistä tai puoltamista tai vastaavasti edellä mainittujen vastakohtia. Tämä johtaa erilaisista arvoista johdettujen kriteerien ja standardien perusteella arvottaviin päätelmiin arviointikohteen hyödyistä, ansioista ja merkittävydestä.

Arvottaminen (evaluation)

Arvottamisessa arviointikohteen tilaa verrataan toivottavaan tai tavoitteeksi asetettuun asiantilaan, esimerkiksi tasa-arvoon tai oikeudenmukaisuuteen. Arvottaminen merkitsee vertailussa hierarkkisen aineksen mukaantuloa, jolla ilmaistaan suosimista, arvonantoa, hyväksyvää myöntämistä tai puoltamista tai vastaavasti edellä mainittujen vastakohtia. Arvottaminen johtaa erilaisista arvoista johdettujen kriteerien ja standardien perusteella arvottaviin päätelmiin arviointikohteen

hyödystä, ansioista ja merkittävydestä. Arvostaminen on taas suhtautumistapa, jossa joku arvosubjekti X pitää jotain kohdetta eli arvo-objektia A arvokkaana.

Pohjimmiltaan arvottamisessa on kysymys valintojen tekemisestä tiettyjen arvojen perusteella punnittaessa esimerkiksi eri toimintavaihtoehtojen perusteltavuutta tai hyvyyttä. Arvoilla on selkeä status arvottamisessa, kun sitä ei yleensä objektiivisuuteen pyrkivässä tieteessä ole.

Periaatteessa arvottaminen voi olla absoluuttista tai suhteellista. Absoluuttisessa arvottamisessa arvotettavaan kohteeseen liitetään tai sille johdetaan jokin arvon mitta. Esimerkiksi voimavarojen arvottamisessa mittana käytetään rahaa. Tällöin voidaan määrittää, kuinka paljon parempi tai huonompi asiantila on. Koulutuksen arvioinnissa absoluuttinen arviointi ei useinkaan ole mahdollista, koska arviointikohteen mittauksen taso ei ole riittävä.

Suhteellisessa arvottamisessa määritetään, onko arviointikohde A parempi tai huonompi kuin B. Absoluuttisessa arvottamisessa taas määritetään, paljonko A:n tuottama asiantila on B:n tuottamaa parempi tai huonompi. Arvottaminen voidaan tehdä kummallakin tavalla riippuen siitä, millä mittauksen tasolla arvottamisen kohde on mitattavissa.

Indikaattori (indicator / educational indicator)

Indikaattorit (mittaimet, osoittimet, ilmaisimet) ovat tavallisesti tilastollisia lukuja tai näiden johdannaisia (indeksejä), joita käytetään suunnittelun, seurannan, ohjauksen ja kehittämisen tarkoituksiin. Varsinaisina koulutusindikaattoreina on totuttu pitämään mittaimia, jotka edustavat kokonaista kohdejoukkoa, toimintayksikköä tai aluetta (kuten kuntaa, lääniä tai maata). Kuitenkin näillä tarkoitetaan yleensä sellaista informaatiota, joka kertoo (indikoi, ilmaisee, osoittaa) jotakin olennaista järjestelmän tai sen tietyn osan tilasta, toiminnasta tai tuloksista. Tiedon on siis oltava jotenkin hyödyksi (tässä) koulutuksen ilmiöiden hahmottamisessa, kuvailussa ja niiden ongelmia koskevassa pohdinnassa – joskus myös päätöksenteossa. Itse asiassa mittaimien tavoite ja ongelmanäkökulma, niiden mielekkyys ja merkittävyys seurannan ja arvioinnin tarpeisiin tekevät niistä indikaattoreita. Käyttötarkoitukset myös sanelevat sen, millaista tietoa alun perin hankitaan ja prosessoidaan mittaimiksi.

Itsearviointi (self-evaluation, self-assessment)

Itsearviointi tarkoittaa arvioitavan yksilön tai arvioitavan kohteen omaan toimintaan, sen edellytyksiin ja tuloksiin kohdistamaa kriittistä ja refleksiivistä tarkastelua. Itsearvioinnilla tarkoitetaan myös sellaisen tiedon hankkimista, jota organisaatio kerää itsestään oman toimintansa kuvaamiseksi, kehittämiseksi ja oppimisen edellytysten parantamiseksi.

Itsearvioinnista voidaan käyttää myös käsitteitä sisäinen arviointi ja oman toiminnan arviointi. Sisäinen arviointi on vakiintunut käsite arvioinnille, jota tehdään säännöllisesti oman toiminnan kehittämiseksi tai jos toiminnasta kerätään arviointitietoja. Tällöin itsearviointi ei ole osa ulkoista arviointia, vaan se toimii jatkuvan arvioinnin periaatteen mukaisesti. Itsearvioinnin ja sisäisen arvioinnin käsitteet eivät ole vakiintuneita, vaan ne tulee sitoa käyttöyhteyteensä. Itsearvioinnin määrittelyyn liittyy, millaisessa toimintaympäristössä sitä toteutetaan, millaisia arvoja sillä halutaan tukea ja millainen rooli reflektiolle annetaan.

Kehittävä arviointi

Kehittävää arviointia voidaan tarkastella eräänä toimintamuotona tyypiteltäessä arviointitoimintaa tarkoituksen ja/tai toimintaperiaatteiden mukaan. Kehittävää arviointia voidaan luonnehtia käyttäjälähtöiseksi prosessiksi, jossa arvioinnin menetelmät räätälöidään arvioinnin tavoitteiden, arviointikohteen ja osallistujien tarpeiden mukaan. Kehittävyuden tarkoituksen toteutumista tavoitellaan arvioinnissa vuorovaikutteisuuksella, asianosaisten tarpeiden huomioon ottamisella sekä positiivisella ja kriittisellä suhtautumisella omaan ja muiden prosessissa olevien toimintaan. Kehittävyys periaatteena hallitsee arviointia koko prosessin ajan ohjaten sitä alusta saakka. Näin ollen prosessiin osallistuneille muodostuu tavanomaista syvällisempi tiedollinen ja kokemuksellinen valmius kehittää toimintaansa. Koska arvioija ja arvioitava ovat tässä toimintatavassa subjekteja, on prosessi monin osin rinnasteinen itsearvioinnin kanssa.

Kehittävän arvioinnin toimintamallissa prosessin osapuolet osallistuvat yhteisymmärryksessä arviointimenetelmien valintaan, tulkinnallisen analyysin tekoon sekä arviointituloksien ja kehittämissuosituksen määrittelyyn. Tässä toimintatavassa ei tieteellistä kurinalaisuutta voi täysin noudattaa. Arviointiin osallistuneiden itsetunnon vahvistuminen on eräs kehittävän arvioinnin tehtävä, mutta tarkoitus voi ulottua myös yhteisön, organisaation ja koko järjestelmän kehittämiseen. Kehittävyuden periaate toteutuu hyvin silloin, kun arviointi on mukana toimintaprosesseissa suunnittelusta toiminnan kehittämiseen saakka.

Laadunhallinta (quality management)

Laadunhallinta voidaan määritellä toimintatavoiksi, järjestelmiksi ja prosesseiksi, joilla pyritään kontrolloimaan, ylläpitämään ja parantamaan toiminnan laatua. Laadunhallintaan ja laadunvarmistukseen sisältyy piirteitä, jotka painottuvat usein toiminnan ulkoiseen ohjaukseen ja kehittämiseen. Laadunhallinta on myös olennainen osa toimijoiden omaehtoista autonomista työtä.

Laadunvarmennus (quality assurance)

Laatua voidaan varmistaa, hallita, arvioida, kehittää, vahvistaa, valvoa ja kontrolloida. Varmistus on mainituista sanoista ehkä neutraalein. Laadunvarmistus viittaa menettelytapoihin, järjestelmiin ja prosesseihin sekä ylläpitää ja parantaa laatua. Laadunvarmistusta koskeva kirjallisuus voidaan jakaa kahteen ryhmään: 1) laadunvarmistukseen liittyvät teoreettiset keskustelut, jotka sisältävät arvioinnin, akkreditoinnin, auditoinnin, suoritusten arvioinnin jne. sekä 2) kansallisten järjestelmien, organisaatioiden ja instituutioiden käytäntöjen kuvaukset.

Laadunvarmistuksen tavoitteet ovat yleistyneet koulutuksessa Kööpenhaminan ja Bolognan prosessien myötä. Laadunvarmistus sisältää korkeakoulutuksessa sekä laadun ylläpitämisen että kehittämisen. Ammatillisessa koulutuksessa haasteet kohdistuvat Kööpenhaminan prosessin tavoitteisiin, joista keskeinen tavoite on koulutuksen laadunvarmistus. Tätä varten on luotu yleinen laadunvarmistuksen viitekehys. Ammatillisen koulutuksen laadunvarmistuksen mekanismeja voidaan erottaa kahdenlaisia: 1) normatiiviset mekanismit kuten lisenssit, opetussuunnitelma ja kvalifikaatiovaatimukset, näytöt, rahoitus ja arviointi sekä 2) vapaaehtoisia mekanismeja kuten laadunhallinta, erilaiset suositukset ja laa-
tupalkinnot.

Laatu (quality)

Laatu on kulloinkin tietyin arvoperustein valittu joukko arvioitavan kohteen, tuotteen, toimintakokonaisuuden tai palvelun (esimerkiksi opetus) ominaisuuksia tai piirteitä (liittyvät mm. resursseihin, prosesseihin ja tuloksiin), jotka ovat vuorovaikutuksessa toisiinsa ja joilla on kontekstissaan erilaisia merkityksiä ja painoarvoja. Laadun paraneminen merkitsee yleensä lisääntyvää (yksilön, yksikön, organisaation tms.) kykyä täyttää sille asetetut vaatimukset ja tavoitteet. Laadun parantaminen edellyttää myös tietoa laatua kuvaavien tekijöiden vuorovaikutusmekanismien keskinäisestä tarkoituksenmukaisuudesta. Laadun määrittelyssä kohteittain ovat mukana myös toimintaan käytetyt resurssit (siihen sijoitetut panokset) ja muu tätä toimintaa tukeva tai rajoittava ympäristö kokonaisuutena. Prosessien laatu määrittyy ajan, toimintakontekstin, tavoitteiden sekä arvioijien arvojen ja käsitysten vuorovaikutuksessa.

Monitahoarviointi

Arvioitavaan kohteeseen liittyvien intressitahojen, toimijoiden ja stakeholdereiden yhdessä tekemä tulkinnallinen analyysi. Usein eri tahot arvottavat kohteen eri puolia eri tavoin. Monitahoarvioinnilla saadaan arvioitava kohde kuvattua useasta näkökulmasta. Monitahoisuuteen kuuluu usein myös useiden menetel-

mien käyttö yhdessä arvioinnissa, kun kohteen lähestyminen eri näkökulmista edellyttää myös erilaisia tiedonhankintamenetelmiä.

Selvitys (account)

Selvitys on luonteeltaan tavallisesti yleinen ja kertaluonteinen. Se ei täytä tutkimuksen ja arvioinnin kriteerejä eikä sen funktio ole tutkimusta edeltävä kartoitus tai sen eksploraatiivinen vaihe. Selvitys on kuitenkin enemmän kuin asiaa koskeva lausunto tai selonteko. Selvitys valaisee kohdetta ajantasaisella tiedolla ja vastaa toimeksiantajansa tarpeisiin. Selvitys kuvaa aihettaan monipuolisesti ja johdonmukaisesti. Selvitykseen ei liity välttämättä etukäteen asetettuja kysymyksiä. Selvityspyyntö voi olla kuitenkin hyvinkin yksilöity. Selvitystä varten ei kerätä tavallisesti uusia aineistoja, mutta sitä varten voidaan kuitenkin kuulla kohteen kannalta merkittäviä asiantuntijoita tai tahoja sekä käyttää jo olemassa olevaa tietoa. Selvityksen lähtökohtana on usein käytännössä havaittu ongelmallinen tilanne, jonka ratkaisua varten halutaan nopeasti pohjatietoja, asianosais- tahojen käsityksiä ja perusteltuja ratkaisuvaihtoehtoja.

Seuranta (follow up)

Seurannalla tarkoitetaan systemaattista, toistuvaa, etukäteen suunnitellulla ja sovitellulla tavalla tehtävää empiirisen tiedon keräämistä, dokumentointia ja hyödyntämistä. Seurannan kohteena on jonkin toiminnan kehittyminen, ja seuranta voi kohdistua joko tuloksiin tai laatuun. Seuranta tapahtuu sovittujen indikaattoreiden avulla: joko tulosta kuvaavien tekijöiden tai laadukkaana toimintana pidettyjen avainilmiöiden avulla. Kertaluonteinen tilan arviointi ei muutu säännöllisesti samalla tavoin toistettuna seurannaksi.

Standardi, laatustandardi (quality standard)

Standardi ilmaisee, miten hyvä arviointikohteen tulee olla tai pitäisi olla. Se voi ilmaista myös, miten hyvin jokin asia tulee tehdä. Standardin hyvyystason voi määrittää viranomainen, asiakas, arvioija tai arviointiryhmä. Standardi voi olla myös yleinen käsitys, malli, esimerkki tai hyvä käytäntö. Joskus arvioija määrittää arvionvaraisen standardin etukäteen ja katsoo, miten se toimii arvioinnissa. Usein on hyödyllistä määrittää, mikä on hyvää vasta sen jälkeen, kun arvioitava kohde on tutkittu lähemmin. Asetettaessa standardia aineiston perusteella, joudutaan kysymään, onko tämä ja tämä pistemäärä riittävän hyvä standardiksi. Tulonarvioinnissa tällaisen standardin määrittäminen voi olla suhteellisen yksinkertaista, mutta ohjelma-arvioinnissa vastaus kysymykseen ”Toimiiko tämä ohjelma hyvin?” ei ole ollenkaan yksinkertainen asia. Yksi mahdollisuus on ot-

taa tällöin standardiksi jonkinlainen ideaali. Se on kuitenkin usein kuvitteellinen, jota ei koskaan ole olemassa.

Termejä standardi ja kriteeri käytetään kirjavasti merkitsemässä toinen toisiansa. Kriteeri ymmärretään kuitenkin yleensä tietyksi ominaisuudeksi, kuten esimerkiksi fyysinen kunto tai koulukypsyys. Standardi ilmaisee tällöin tasoa, joka tulee saavuttaa ollakseen esimerkiksi hyvä tai erinomainen. Standardi ilmaisee siis kriteerinä olevan ominaisuuden määrää, joka tulee olla tietyn tason (standarditason) saavuttamiseksi.

Käytettäessä arviointiryhmiä standardit ovat usein arvioijien ”päässä”. Ne ovat ns. implisiittisiä standardeja, joiden pohjalta arviointitulokset määräytyy.

Tehokkuus (efficiency)

Tehokkuus määritellään yksikön tuottavuuden suhteena parhaaseen mahdolliseen (teknologisen tason määrittämään) tuottavuuteen. Tehokkuus on siten aina suhteellinen käsite, ja sen avulla verrataan yksiköiden tuottavuutta keskenään. Tehokas yksikkö tuottaa maksimaalisen tuotoksen käytettävissä olevilla panoksilla tai annetun tuotoksen pienemmillä panoksilla kuin tehoton yksikkö. Vastavasti jos yksikkö on tehoton, se olisi voinut tuottaa saman tuotoksen pienemmällä panoksella tai suuremman tuotoksen annetuilla panoksilla.

Tehokkuutta voidaan kuvata eri näkökulmista. Teknisellä tehokkuudella tarkoitetaan yleensä joko yksikön kykyä tuottaa annetuista panoksista mahdollisimman suuri tuotos tai tietyn tuotoksen tuottamista mahdollisimman pienin panoksilla. *Kustannustehokkuudessa* otetaan huomioon myös panosten hinnat, ja kyse on yksikön kyvystä tuottaa tuotokset mahdollisimman alhaisin kustannuksin. Kustannustehokkuus voidaan jakaa tekniseen tehokkuuteen ja panoskäytön *allokatiiviseen tehokkuuteen*. Allokatiivinen tehokkuus ilmaisee sen, kuinka lähellä tekninen tehokkuus on optimaalista ottaen huomioon myös panosten hinnat. *Skaalatehokkuudella* viitataan yksikön optimikokoon. Kun yksikkö on skaalatehokas, sen koko ei poikkea optimikoosta.

Tuottavuus ja taloudellisuus (productivity, economic efficiency)

Tuottavuudella tarkoitetaan tuotoksen määrän suhdetta panoksen määrään. Tuottavuus kuvaa siten sitä, missä suhteessa panoksia muunnetaan tuotoksiksi tuotantoprosessissa. Tuottavuuden mittaaminen on mielekäästä silloin, kun sille on olemassa jokin vertailukohta. Tavallisimmin tuottavuutta verrataan eri ajankohtina tai eri yksiköiden kuten koulujen välillä. Tuottavuuden mittaamiseksi tarvitaan tietoa arvioitavan prosessin panoksista ja tuotoksista. Koulutuksessa tavallisimpia tuotoksia ovat esimerkiksi oppimistulokset ja sijoittuminen jatko-opintoihin tai työmarkkinoille. Panoksia ovat esimerkiksi opetukseen käytetyt

resurssit kuten opetuksen määrä, opetusta antavien opettajien ominaisuudet sekä oppilaitoksen tilat ja laitteet. Kun tuotoksia mitataan oppimistuloksilla, on tärkeää ottaa huomioon myös opiskelijoiden aikaisempi osaaminen (ns. value-added-malli) sekä muut osaamiseen vaikuttavat tekijät kuten sosioekonominen tausta. Taloudellisuudella tarkoitetaan yleensä tuottavuutta silloin, kun tuotantoprosessin panoksia mitataan kappalemäärien sijaan kustannusten avulla.

Ulkoinen arviointi (external evaluation)

Ulkoisella (tai ulkopuolisella) arvioinnilla tarkoitetaan arvioitavan kohteen kanalta ulkopuolisen tahon tekemää arviointia. Ulkoista arviointia edustaa puh-taimmillaan toiminta, jossa arvioijat eivät osallistu tai eivät ole osallistuneet arvioitavan toiminnan ohjaukseen, sen kulkuun tai resursointiin.

Ulkopuolisuuden vaatimus nousee erityisesti arvioinnin tavoitteista, joissa keskeistä on tiedon tuottaminen kansallisen ja paikallisen päätöksenteon ja kehittämisen tarpeita varten. Ulkopuolinen arvioija voi kohdentaa arviointinsa tärkeinä pitämiinsä asioihin ja voi tuottaa tietoa, joka on luotettavaa ja riippumatonta.

Vaikuttavuus (effectiveness)

Vaikuttavuuden käsitettä tarvitaan tarkasteltaessa organisaation sisäisen tuotoksen (esim. oppimistulosten) suhdetta sen ulkopuoliseen kriteeriin (esim. työelämän tarpeisiin tai jatko-opintokelpoisuuteen). Kysymys on siirtovaikutuksesta, impaktista, ja organisaation tuotoksen käyttö- ja vaihtoarvosta markkinoilla. Käsite antaa keinon mitata palvelun tuottajan ja asiakkaan kohtaamisen laatua. Toisin kuin muut tuloksellisuuskäsitteet (tehokkuus, taloudellisuus, tuottavuus) se siirtää tarkastelun organisaation sisäisestä palvelukyvyistä ulkoiseen palvelukykyyn ja palvelun itsensä laatuun. Arvioitavana ovat silloin mm. palvelujen relevanttius (ei riitä, että tekee asiat oikein; on tehtävä myös oikeita asioita), riittävyys, kohdentuminen, tavoitettavuus, ajantasaisuus ja kustannus-hyöty-suhde.

Vaikuttavuusarvion kriteerinä on joko yksilöllisen (asiakasvaikuttavuus) tai kollektiivisen (yhteiskunnallinen vaikuttavuus) hyvinvoinnin lisääntyminen eli asiakkaan tarpeiden tyydyttyminen ja hänen odotuksiinsa vastaaminen tai yhteiskunnan tavoitteiden saavuttaminen tai niitä kohti eteneminen. Käsitteen avulla voidaan siis tarkastella sekä yhteiskunnallisten ja poliittisten järjestelmien toimivuutta ja tuloksellisuutta (instituutiotarpeiden tyydyttymistä ja funktioiden toteuttamista) että asiakastyytyväisyyttä (yksilön hyvinvointitarpeiden tyydyttymistä).

Kirjoittajat

Päivi Atjonen
Professori
Kasvatustieteen laitos
Joensuun yliopisto
paivi.atjonen@joensuu.fi

Mira Huusko
Suunnittelija, KM
Käyttätymistieteellinen tiedekunta
Helsingin yliopisto
mira.huusko@helsinki.fi

Antti Kauppi
Johtaja
Helsingin yliopisto, Koulutus- ja
kehittämiskeskus Palmenia
antti.kauppi@helsinki.fi

Tanja Kirjavainen
Economisti, KTL
Valtion taloudellinen tutkimuskeskus
tanja.kirjavainen@vatt.fi

Gunnel Knubb-Manninen
Pääsuunnittelija, PsL
Koulutuksen arviointisihteeristö
Jyväskylän yliopisto
gunnel.knubb-manninen@eval.jyu.fi

Esko Korkeakoski
Pääsuunnittelija, KT, dosentti
Koulutuksen arviointisihteeristö
Jyväskylän yliopisto
esko.korkeakoski@eval.jyu.fi

Timo Laine
Tutkija, FT
Kasvatustieteen laitos
Jyväskylän yliopisto
timo.laine@edu.jyu.fi

Pirjo Linnakylä
Professori
Koulutuksen tutkimuslaitos
Jyväskylän yliopisto
pirjo.linnakyla@ktl.jyu.fi

Heikki K. Lyytinen
Pääsihteeri, Koulutuksen arviointi-
neuvosto
Johtaja, Koulutuksen arviointi-
sihteeristö
Jyväskylän yliopisto
hekalyyt@eval.jyu.fi

Anita Malinen
Lehtori, KT
Kasvatustieteen laitos
Jyväskylän yliopisto
anita.malinen@edu.jyu.fi

Jyri Manninen
Tutkimusjohtaja, KT, dosentti
Helsingin yliopisto, Koulutus- ja
kehittämiskeskus Palmenia
jyri.manninen@helsinki.fi

Pentti Nikkanen
Erikoistutkija, KT, dosentti
Koulutuksen tutkimuslaitos
Jyväskylän yliopisto
pentti.nikkanen@ktl.jyu.fi

Reijo Raivola
Professori (emeritus)
Tampereen yliopisto
reijo.raivola@uta.fi

Heikki Silvennoinen
Pääsuunnittelija, VTT, dosentti
Koulutuksen arviointisihteeristö
Jyväskylän yliopisto
heikki.silvennoinen@eval.jyu.fi

Marja-Leena Stenström
Professori, dosentti
Koulutuksen tutkimuslaitos
Jyväskylän yliopisto
marja-leena.stenstrom@ktl.jyu.fi

KOULUTUKSEN ARVIOINTINEUVOSTON JULKAISUJA

- 1:2004 **Koulutuksen arvioinnin uusi suunta.** Arviointiohjelma 2004–2007. 55 s. 10 e. Myös verkkoversio.
- 2:2004 **Utbildningsutvärderingens nya inriktning.** Utvärderingsprogram för perioden 2004–2007. 57 s. 10 e. Myös verkkoversio.
- 3:2005 **New Directions in Educational Evaluation.** Evaluation Programme 2004–2007. 64 s. 10 e. Myös verkkoversio.
- 4:2005 Ståhle, B.: **Toisen asteen koulu Pohjoismaissa.** Toisen asteen koulujen pohjoismainen vertailu ”Pohjoismainen ISUSS-raportti”. 143 s. 20 e. Myös verkkoversio.
- 5:2005 Räsänen, A.: **EFQM-arviointimalli ammatillisen koulutuksen järjestäjien arvioinnin tukena.** 72 s. 20 e.
- 6:2005 Lyytinen, H.K. & Räsänen, A. (toim.): **Kehittämissuuntaa arvioinnista.** 246 s. 25 e.
- 7:2005 Rönholm, H. & Räsänen, A. (toim.): **Arviointi tukee kehittymistä – miten arvioinnin kehittymistä tuetaan?** Koulutuksen järjestäjien tukeminen arviointiin liittyvissä asioissa. 37 s. 15 e. Myös verkkoversio.
- 8:2005 Korkeakoski, E. (toim.): **Koulutuksen perusturva ja oppimisen tuki perusopetuksessa.** Osaraportti 1: Arviointiraportti. 61 s. 12 e. Myös verkkoversio.
- 9:2005 Korkeakoski, E.: **Koulutuksen perusturva ja oppimisen tuki perusopetuksessa.** Osaraportti 2: Tausta ja tulokset. 156 s. 21 e. Myös verkkoversio.
- 10:2005 Korkeakoski, E. (toim.): **Koulutuksen perusturva ja oppimisen tuki perusopetuksessa.** Osaraportti 3: Syventävät artikkelit. 131 s. 20 e. Myös verkkoversio.
- 11:2005 Mehtäläinen, J.: **Eriyisopetuksen tarve lukiokoulutuksessa.** 100 s. 17 e. Myös verkkoversio.
- 12:2005 Knubb-Manninen, G. (red.): **Grundtryggheten och behovet av stöd i skolan.** 60 s. 12 e. Myös verkkoversio.
- 13:2005 Rönholm, H. & Räsänen, A. (red.): **Utvärdering stödjer utvecklingen – hur kan utvärderingens utveckling stödjas?** Stöd till utbildningsanordnarna i frågor som gäller utvärdering. 42 s. 15 e. Myös verkkoversio.
- 14:2006 Räsänen, A. & Rönholm, H.: **Itsearviointi koulu yhteisöä kehittäväksi.** EFQM-arviointimalli yleissivistävässä koulutuksessa. 79 s. 20 e. Myös verkkoversio.
- 15:2006 Raivola, R., Heikkinen, A., Kauppi, A., Nuotio, P., Oulasvirta, L., Rinne, R., Kamppi, P. & Silvennoinen, H.: **Aikuisten opiskelumahdollisuudet ja järjestäjaverkko toisen asteen ammatillisessa koulutuksessa.** 219 s. 24 e. Myös verkkoversio.
- 16:2006 Vaherva, T., Malinen, A., Moisio, A., Raivola, R., Salo, P., Kantasalmi, K., Kamppi, P. & Silvennoinen, H.: **Vapaan sivistystyön oppilaitosrakenne ja palvelukyky.** 207 s. 24 e. Myös verkkoversio.
- 17:2006 Knubb-Manninen, G. (red.): **Vuxenutbildningsfältet.** Utvärdering av fritt bildningsarbete och yrkesinriktad vuxenutbildning. 79 s. 16 e. Myös verkkoversio.
- 18:2006 Silvennoinen, H. (toim.): **Koulutuksen arviointi verkostoituu.** 86 s. 16 e. Myös verkkoversio.
- 19:2006 Räsänen, A. & Silvennoinen, H.: **Virtaa vertaisten verkosta.** Ammatillisten erikoisoppilaitosten laadunhallinta. 59 s. 12 e.
- 20:2006 Tynjälä, P., Räsänen, A., Määttä, V., Pesonen, K., Kauppi, A., Lempinen, P., Ede, R., Altonen, M. & Hietala, R.: **Työpaikalla tapahtuva oppiminen ammatillisessa peruskoulutuksessa.** Arviointiraportti. 254 s. 30 e. Myös verkkoversio.
- 21:2006 Räsänen, A. (toim.): **Työpaikalla tapahtuva oppiminen ammatillisessa peruskoulutuksessa.** Tiivistelmä. 39 s. 15 e. Myös verkkoversio.
- 22:2006 Räsänen, A. (red.): **Inlärnning på arbetsplatsen inom grundläggande yrkesutbildning.** Sammandrag. 37 s. 15 e. Myös verkkoversio.
- 23:2007 Mäensivu, K., Mäenpää, H., Määttä, M., Volanen, M.V., Knubb-Manninen, G. & Mehtäläinen, J. & Räsänen, A.: **Lukiokoulutuksen ja ammatillisen koulutuksen yhteistyö opetuksen järjestämisessä.** 156 s. 25 e. Myös verkkoversio.
- 24:2007 Räsänen, A. & Hietala, R. (toim.): **ESR-rahoitus ja työssäoppimisen järjestäminen.** Arviointiraportti. 109 s. 20 e. Myös verkkoversio.
- 25:2007 Vaherva, T., Malinen, A., Moisio, A., Raivola, R., Salo, P., Kuusipalo, P., Silvennoinen, H. & Vaahtera, K.: **Vapaan sivistystyön vaikuttavuus ja suuntaviivaopinnot.** 189 s. 23 e. Myös verkkoversio.

- 26:2007 Raivola, R., Heikkinen, A., Kauppi, A., Nuotio, P., Oulasvirta, L., Rinne, R., Knubb-Manninen, G., Silvennoinen, H. & Vaahtera, K.: **Aikuisten näyttötutkintojärjestelmän toimivuus.** 323 s. 30 e. Myös verkkoversio.
- 27:2007 Räisänen, A. & Hietala, R. (toim.): **Yhteisiin pöytiin.** Ammatillisen koulutuksen aluekehitysvaiikutukset. Arviointiraportti. 319 s. 30 e. Myös verkkoversio.
- 28:2007 Räisänen, A. (toim.): **Yhteisiin pöytiin.** Ammatillisen koulutuksen aluekehitysvaiikutukset. Tiivistelmä. 36 s. 15 e. Myös verkkoversio.
- 29:2007 Räisänen, A. (red.): **Till gemensamma bord.** Yrkesutbildningens regioneffekter. Sammandrag. 36 s. 15 e. Myös verkkoversio.
- 30:2008 Atjonen, P., Halinen, I., Hämäläinen, S., Korkeakoski, E., Knubb-Manninen, G., Kupari, P., Mehtäläinen, J., Risku A-M., Salonen, M. & Wikman, T.: **Tavoitteista vuorovaikutukseen.** Perusopetuksen pedagogiikan arviointi. 288s. 30 e. Myös verkkoversio.
- 31:2008 Korkeakoski, E. & Silvennoinen, H. (toim.): **Avaimia koulutuksen arvioinnin kehittämiseen.** 229 s. 28 e.

Tilaukset ja tiedustelut:

Koulutuksen arviointisihteeristö
PL 35, 40014 Jyväskylän yliopisto
puh. (014) 260 3220
faksi (014) 260 3241
ktl-asiakaspalvelu@ktl.jyu.fi
www.edev.fi

Osa julkaisuista on saatavissa myös verkosta:

www.edev.fi/portal/julkaisu



Koulutuksen
arviointineuvosto

Kansi: Martti Minkkinen

Avaimia koulutuksen arvioinnin kehittämiseen käsittelee arvioinnin menetelmäkysymyksiä.

Arvioinnin arvot ja toimintaperiaatteet, arvioinnin tarkoitus ja tavoitteet ovat arviointimenetelmien kehittämisen peruslähtökohtia. Arviointikohteiden laaja kirjo sekä vaatimus tukea koulutuksen kehittämistä ja parantaa oppimisen edellytyksiä ovat haasteita myös menetelmälliselle kehittämiselle. Julkaisun artikkeleissa tarkastellaan mainittuja asioita eri näkökulmista. Julkaisussa käsitellään myös arvioinnin perusterminologiaa tavanomaista monipuolisemmin.

Kirja on tarkoitettu arvioinnin parissa työskenteleville asiantuntijoille, arviointitietoa työssään hyödyntäville päättäjille ja kehittäjille tai muuten arvioinnin metodologisista ja metodisista asioista kiinnostuneille.

Koulutuksen
arviointineuvoston julkaisu 31
ISBN 978-951-39-5347-8
ISSN 1795-0163
2008