

# Kansainväliset vertailevat oppimistulosarvioinnit perus- ja korkea-asteella

Jani Ursin & Jouni Välijärvi

## ABSTRACT

International comparative assessments of learning outcomes in basic and tertiary education

International assessments of learning outcomes have rapidly increased over the last two decades, especially as a consequence of PISA studies. There are now international comparative studies of learning outcomes ranging from basic to tertiary and adult education. The purpose of the present article is to describe the history and background of international comparative assessments with particular respect to two OECD assessments: PISA and AHELO. We argue that educational policy, pedagogical and scientific interests exert influence on international comparative assessments. These different interests promote both complementary and conflicting expectations toward international assessments. However, international assessments also have the potential to increase understanding of learning outcomes on the part of those who represent educational policy, pedagogical and scientific interests, thereby increasing the likelihood that educational systems and learning environments can be developed in an ethically equitable way.

## JOHDANTO

Kansallisen ja kansainvälisen arvioinnin merkitys yhteiskuntapoliittisen ohjauksen välineenä on kasvanut voimakkaasti 1990-luvulta lähtien. Näin on myös koulutuksessa aina esiopetukselta korkea- ja aikuiskoulutukseen. Lisäksi koulutuksen järjestäjien ja oppilaitosten itsearvioinnit ovat nopeasti yleistyneet ja lainsäädäntö on tässä suhteessa varsin velvoittava erityisesti perusasteella, mutta myös korkea-asteella on arviointivaade (ks. esim. Huusko 2009; Perusopetuslaki 1998). Niinpä aikaamme on luonnehdittu sellaisilla käsitteillä kuten arvioiva valtio (Neave 1998) ja audit-yhteiskunta (Power 1999).

Jo 1980-luvulla monet pohjoismaisten hyvinvointiyhteiskuntien itsestään selvyyksinä pidetyt tavat huolehtia yhteisvastuullisesti kansalaisten sivistyksellisistä, sosiaalisista ja terveydellisistä palveluista asetettiin kyseenalaisiksi. Verovaroin rahoitettuja ja lähes yksinomaan julkisen vallan ylläpitämiä koulutusjärjestelmiä alettiin haastaa kilpailuun, valinnanvapauksien lisäämiseen, rahoituspohjan monipuolistamiseen, läpinäkyvyyteen ja tulosvastuuseen. Paineet verotuksen keventämiseen, jota Suomen liittyminen Euroopan Unionin jäseneksi voimisti, kasvoivat. Desentralisaation aalto poisti peruskoulutuksen tarkastusjärjestelmän, menoperusteisen rahoituksen, oppikirjojen kontrollin, koulupiiri- ja vähensi kansallisen opetussuunnitelman ohjausvaltaa. Talouden lama kiihdytti tätä muutosta ja korosti myös julkisessa koulutuksessa

taloudellisuuden ja tehokkuuden vaatimuksia. Tarve tietää, mitä kouluissa tapahtui – ja myös halu ohjailta kouluja – kuitenkin säilyi. Resurssi- ja prosessiohjauksen sijaan alettiin painottaa tulosohtajausta kaikilla koulutusasteilla. Näin arvioinnille eri muodoissa syntyi nopeasti kasvava kysyntä.

Tässä artikkelissa tarkastelemme koulutuksen kansainvälisten arviointien taustoja, kehityshistoriaa ja näiden liittymistä yleisempään yhteiskunnalliseen muutokseen. Kansainvälisillä arvioinneilla viittaamme tässä laajoihin eri maiden koulujärjestelmiä kvantitatiivisten aineistojen pohjalta vertaileviin hankkeisiin, joissa mielenkiinnon kohteena ovat pääasiassa oppimistulokset. Pyrimme myös avaamaan arviointien sisältöä, toteutustapaa ja vaikuttavuutta tarkastelemalla hieman yksityiskohtaisemmin OECD:n organisoimia PISA-ohjelmaa (Programme for International Student Assessment) ja AHELO-hankkeen esiselvitystä (Assessment of Higher Education Learning Outcomes – Feasibility Study).

### **ARVIOINTI LAAJENEET KOKO KOULUTUSJÄRJESTELMÄÄ KOSKEVAKSI**

Kansainvälisten arviointien perinne nojaa vahvimmin perusasteen päättövaiheen oppimistulosten vertailuun. PISA edustaa tätä perinnettä ja on tällä hetkellä laajin ja vaikutuksiltaan merkittävin koulutuksen alueen kansainvälinen arviointiohjelma. PISA:ssa arvioidaan kolmen vuoden välein 15-vuotiaiden nuorten lukutaitoa sekä matematiikan ja luonnontieteiden osaamista. Kukin yksittäinen tutkimus kestää neljä vuotta. Kullakin mittauskerralla yksi yllä mainituista kolmesta alueesta on ”pääroolissa” ja muilla alueilla arvioidaan lähinnä osaamisen kehittymistä. Ensimmäinen PISA-tutkimus toteutettiin vuonna 2000, jolloin lukutaito oli pääalueena, ja joulukuussa 2010 julkistetaan neljännen PISA-tutkimuksen (PISA 2009) tulokset, jolloin lukutaito on jälleen pääalueena. PISA:n ytimen muodostavat 33 OECD-maata. Näiden lisäksi ohjelma on vetänyt mukaansa kasvavan määrän muita maita: vuoden 2009 PISA:an osallistui 64 maata.

AHELO-hanke on ensimmäinen laajamittainen yritys ulottaa kansainvälisen vertailun näkökulma korkeakoulujen tuottamaan osaa-

miseen. AHELO:n synty ajoittuu vuoteen 2008, jolloin OECD:n jäsenvaltioiden opetusministerit kiinnittivät huomiota korkeakoulutuksen laadun arviointiin – tai paremminkin sen puuttumiseen kansainvälisellä tasolla. Tämän jälkeen asiantuntijatapaamisissa pohdittiin, onko mahdollista käynnistää kansainvälinen vertaileva tutkimus- ja arviointihanke korkeakoulutuksen oppimistulosten tarkastelemiseksi. Näiden tapaamisten seurauksena syntyi AHELO.

AHELO:n esiselvityksen ensisijaisena tavoitteena on tuottaa tietoa siitä, onko mahdollista kansainvälisesti mitata opiskelijoiden oppimistuloksia. Lisäksi hankkeen tarkoituksena on selvittää, mitä geneerisiä taitoja, taloustieteellistä osaamista ja teknisen alan valmiuksia korkeakouluissa opitaan. Tällä hetkellä AHELO:ssa on mukana 15 maata, joista valtaosa on OECD:n jäsenvaltioita. Esiselvityksen tulokset julkaistaan 2012 lopulla, jonka jälkeen OECD:n koulutuspolitiikkakomitea tekee päätöksen laajamittaisen AHELO:n käynnistämisestä tai käynnistämättä jättämisestä. (OECD 2010; ks. myös Ursin & Hyytinen 2010.)

### **KOULUTUKSEN VERTAILEVALLA ARVIOINNILLA ON JO HISTORIANSA**

Koulutuksen kansainvälinen vertaileva arviointi ei ole 1990-luvun tuotetta. Suomi osallistui 1960-luvulta vuosituhatteen alkuun ulottuvalla ajanjaksolla noin 20 koulutuksen kansainväliseen arviointitutkimukseen. Jo tätä ennen vertailevalla tutkimusotteella oli koulutusjärjestelmien tarkastelussa melko vakiintunut perinne, mutta vertailukelpoisen kvantitatiivisen tiedon tuottamiseksi tutkimusalueen metodologia oli kehittymätöntä. Kehittämistyötä varten perustettiin pääosin eri maiden tutkimuslaitosten yhteistyönä IEA-järjestö (International Association for the Evaluation of Educational Achievement), jonka keskeisin puuhemies oli Tukholman yliopisto kansainvälisen koulutuksen professori Torsten Husén. Suomalaiset koulutustutkijat olivat alusta saakka aktiivisesti mukana järjestön toiminnassa (Leimu 2004).

IEA-tutkimuksissa on selvitetty lähinnä lukutaidon, matematiikan ja luonnontieteiden sekä yhteiskunnallisen osaamisen tasoa kehittyneissä teollisuusmaissa. Myös tietoteknologian hyödyntäminen on ollut useiden vertailevien

tutkimusten kohteena. Tutkimusten pääpaino on ollut kysymyksessä, kuinka hyvin eri maiden koulujen opetussuunnitelmissa määritellyt tavoitteet on kouluissa saavutettu ja mitkä tekijät tuloksia selittävät. Hankkeissa on kerätty myös laajat aineistot koulun oppimisympäristöistä, opetussuunnitelmista, opettajista ja opetuksesta eri maissa. Pääpaino on ollut perusasteen päätösvaiheen oppimistuloksissa, joskin myös alakoulujen saavuttamista tuloksista lukutaidossa ja matematiikassa on koottu arviointitietoa. Muutamia hankkeita on toteutettu myös toisen asteen koulutuksessa mutta Suomi on osallistunut niihin vain satunnaisesti. (Leimu ym. 2001.)

Erityisesti kuluvan vuosikymmenen loppupuoliskolla koulutuksen kansainvälisten arviointien määrä ja yhteiskunnallinen vaikuttavuus on Suomessa nopeasti kasvanut. Ratkaiseva vaihe oli OECD:n PISA-ohjelman ensimmäinen mittaus vuonna 2000. Tuossa vaiheessa Suomen opetushallinnon kiinnostus kansainvälisiin arviointeihin oli melko laimeaa ja PISA oli ainoa kansainvälinen arviointihanke, johon Suomi tuolloin osallistui. PISA:n siivittämänä kiinnostus kansainvälisiin vertailuihin kuitenkin alkoi nopeasti kasvaa. Tällä hetkellä opetus- ja kulttuuriministeriön sekä vähäisemmässä määrin työ- ja elinkeinoministeriön tuella Suomessa toteutuu, laskutavasta riippuen, 7–8 erilaisista koulutussektorin kansainvälistä arviointia. Kansainvälisen arvioinnin lisäksi myös kansallisten arviointien määrä on kasvanut.

Painopiste on edelleen perusopetuksessa ja erityisesti peruskoulun päättövaiheen oppimistuloksissa, mutta kiinnostus ulottaa arviointi koko koulutusjärjestelmän kattavaksi on lisääntynyt. Koulutuksen varhaisempien vaiheiden tulokset kiinnostavat siksi, että monet keskeiset oppimisen prosessit näyttävät muotoutuvan ja eriytyvän jo koulunkäynnin varhaisissa vaiheissa. Esimerkiksi asenteet ja uskomukset eri oppiaineisiin vakiintuvat jo varhain ja heijastuvat myöhemmin voimakkaina myöhempiin oppimistuloksiin.

Korkea-asteen oppimisesta ja tuloksista halutaan puolestaan saada täsmällisempää tietoa vastapainoksi melko karkeille ja ulkokohtaisia tutkimustoiminnan tuotoskriteereitä painottaville ranking-listoille, kuten Shanghai Jiao Tong yliopiston tuottama Academic Ranking of World

Universities (ARWU 2010). Pyrkinessään määrittämään ja vertailemaan oppimistulosten tasoa korkea-asteella, AHELO-hanke kuvastaa selkeästi tavoitteita irtaantua näistä tutkimustoimintaan keskittyvistä ja rankeerauslistoja laativista arvioinneista (OECD 2010; Palonen 2010).

Tähän päivään tultaessa arviointien fokus on laajentunut kouluoppimisesta osaamisen analyysiin: koulussa omaksuttujen tietojen ja taitojen tarkastelusta opitun hyödyntämiseen vaihtelevissa käyttötilanteissa ja ympäristöissä. Myös yksilöön, kouluun ja yhteisöön kytkeytyvien kontekstitekijöiden merkitys osaamiselle on noussut yhä laajemman kiinnostuksen kohteeksi, kun etsitään keinoja koulutusjärjestelmien toimivuuden parantamiseen ja osaamistason nostamiseen. (Juva ym. 2009; OECD 2009.)

Koulutuksen arviointien ulottaminen kehittyneiden maiden koulutusjärjestelmien ulkopuolisiin maihin on alusta lähtien ollut yhdenmukainen pyrkimys IEA:n ja OECD:n strategioissa. Niinpä nykyiset hankkeet tyypillisesti kattavat 60–80 maata kaikilta eri mantereilta. Esimerkiksi useilla Etelä- ja Keski-Amerikan mailla on jo pitkät perinteet kansainvälisistä arvioinneista ja viime vuosina monet Aasian maat ovat aktivoituneet koulutusjärjestelmiensä kehittämiseen kansainvälisiä vertailuja hyödyntäen. Kansainväliset yhteishankkeet ovat olleet myös osa strategiaa integroida entiset sosialistiset maat läntisiin demokratioihin. Keskeinen selitys kasvavalle aktiivisuudelle on ollut kansainvälisten rahoittajien, kuten Maailmanpankin ja Aasian kehityspankin, taloudellinen tuki. Sen sijaan Afrikan maiden osallistuminen näihin hankkeisiin on säilynyt melko vaatimattomana paria poikkeusta lukuun ottamatta.

## KOULUTUSPOLITIikka JA ARVIOINTI

Koulutuksen arvioinnin motiivit ovat sekä teollisia että poliittisia. Painopiste on muuttunut viimeisen 10–20 vuoden aikana poliittisten intressien suuntaan. Monissa maissa koulutus on tänä päivänä voimakkaiden poliittisten intotohimojen kohde, koska sillä nähdään olevan vahvoja yhteyksiä yhteiskuntien talouteen, hyvinvointiin ja kansainväliseen kilpailukykyyn (esim. Whitty 2010). Kehitys Suomessa on ollut samansuuntainen, joskin vaikutuksiltaan paljon lievempi kuin esimerkiksi Englannissa tai

Ruotsissa. Tämä näkyy muun muassa siinä, että Suomi ei ole ottanut käyttöön koko ikäluokan kattavia kansallisia testejä, joilla useimmissa Euroopan maissa kouluja asetetaan paremmuusjärjestyksiin eri luokkatasoilla. Suomi on myös ainoita maita, joissa ei ole enää koulujen tarkastusjärjestelmää.

Koulutuksen politisoituessa ja arvioinnin yhteiskunnallisen merkityksen kasvaessa, arviointitiedon tuottaminen nousee keskeiseksi kysymykseksi (esim. Harisalo ym. 2007; Simola & Rinne 2004; Rinne & Simola 2005). Tarve tehdä selkeä ero arviointitutkimuksen ja sen pohjalta tehtävien koulutuspoliittisten päätelmien – arvioinnin – välillä korostuu.

Ollakseen käyttökelpoista, puolueetonta ja uskottavaa, arviointia varten tapahtuvan tiedontuotannon tulee noudattaa tiukasti tieteellisen tutkimuksen standardeja ja sen tulee tapahtua viranomaisista ja poliittisista päättäjistä riippumattomana prosessina. Arviointitutkimuksen on täytettävä vaativimmat tieteellisen tutkimuksen standardit – sen on myös pystyttävä kyseenalaistamaan arviointihankkeiden lähtökohtia ja peruskäsitteitä. (Simola & Rinne 2004.) Arviointi puolestaan on tämän tiedon arvottamista koulutuspoliittisin tai muuten arvolutautunein kriteerein. Tässä mielessä koulutuksen arviointien tulokset ja niiden painotukset sekä arviointien pohjalta tehtävät johtopäätökset voivat vaihdella, vaikka arvioinnin perustana oleva tieto on sama. Julkisuudessa tämä näkyy muun muassa kannanotoissa lahjakkaimpien asemasta ja osaamisen tasosta suomalaisessa peruskoulussa tai korkeakoulujen paremmuuskeskusteluissa.

Suomessa rajankäynti tutkimusperustaisen tiedontuotannon ja siihen perustuvan arvioinnin välillä kansainvälisissä arvioinneissa selkiytettiin jo 1960-luvulla, kun päädyttiin arvioinnin edellyttämien tutkimushankkeiden sijoittamisesta yliopistoihin hallintokoneiston sijasta. Useissa maissa kansainvälisten arviointitutkimusten kansallisista velvoitteista vastaavat ministeriöiden yhteydessä toimivat tutkimusyksiköt (Linnakylä & Välijärvi 2005b). Näin erityisesti niissä maissa, joissa koulutuksen arviointiin on kohdistunut vahvoja poliittisia intressejä. Suomalainen käytäntö selittyy sillä, että vielä 1990-luvulle tultaessa koulutuksen vertailevat arvioinnit olivat lähinnä tutkijoiden mielenkiinnon kohde.

1990-luvulta lähtien tietoyhteiskunta-diskurssi, työelämän osaamistarpeiden voimakas muutos ja globalisaation myötä nopeutunut työtehtävien siirtyminen maiden rajojen yli nostivat nopeasti koulutusjärjestelmien tuottavuuden ja tehokkuuden yhteiskunnallisen keskustelun ja poliittisten intressien ytimeen (esim. Ozga 2010). Tämä näkyy selvästi muun muassa OECD:n ja Euroopan Unionin linjauksissa, joista hyvin monet koskettavat koulutuksen sisältöä, tavoitteita ja toteutusmuotoja. Tämä siitä huolimatta, että Unionin perustamisasiakirjoissa koulutus todetaan yksiselitteisesti kansalliseen määräysvaltaan kuuluvaksi asiaksi. OECD:n toteuttamat vertaisarviointit ja indikaattorihankkeet yhdessä Euroopan unionin toimintaohjeiden kanssa ovatkin muodostaneet 2000-luvulla kansallista päätöksentekoa ohjaavia ”pehmeitä lakeja”. Nämä pehmeät lait vaikuttavat vahvasti kansallisiin lakiuudistuksiin. Esimerkiksi Kallo (2009) osoitti, kuinka Suomessa OECD:n vuoden 2006 teematutkiminta on korkea-asteella käytetty useissa korkeakoulupolitiikan uudistuksissa ja kuinka suositusten sisältöjä on muokattu eri poliittisiin asiansyhteyksiin sopiviksi. Myös oppimistulosarviointit ovat saaneet tästä näkökulmasta osakseen kritiikkiä (ks. esim. Uljens 2007).

Eri kansainvälisten organisaatioiden pyrkimykset näyttävät tässä suhteessa olevan pitkälle yhteneväisiä ja myös yksittäisten koulutuksen arviointihankkeiden tasolla niiden yhteistyö on tiivistä. Tässä katsannossa PISA:n ja AHELO:n kaltaiset arviointihankkeet voidaan nähdä pyrkimyksinä edistää koulutusjärjestelmien harmonisoitumista ja tuottavuutta ilman, että julkilausutusta subsidiariteetti-periaatteesta koulutussektorilla luovutaan. Osallistumalla näihin hankkeisiin tietämys toimivimmista rakenteista ja koulutuspolitiikan keinoista välittyy laajaan tietoisuuteen ja vaikuttaa näin yhdenmukaisesti itsenäisesti päätöksensä tekevien maiden ratkaisuihin. Samalla korostuu entistä voimakkaammin riippumattoman ja metodisesti korkeatasoisen tiedontuotannon merkitys ja sen selkeä erottaminen tätä tietoa hyödyntävästä mutta poliittisista intresseistä ja tavoitteista nousevasta koulutuksen rakenteiden, toimintatapojen ja tulosten arvottamisesta. Monissa Euroopan maissa vilkkaana käyvä evidence-based policy -keskustelu kiteytyy pitkälti näiden pyrkimysten äärelle (ks. esim. Head 2008).

PISA-ohjelman rakentuminen indikoi selkeästi sitä pitkää kehitysprosessia, jonka tuloksena koulutusjärjestelmän toimivuus ja tehokkuus tärkeimpänä kansallisen osaamispääoman aktivoijana ja kehittäjänä on noussut myös talous- ja työvoimapolitiittisen keskustelun yhdeksi ydinteemaksi. PISA-ohjelman juuret ulottuvat 1980-luvulle, jolloin OECD käynnisti IEI-projektin (International Education Indicators). Nämä toimet osoittavat poliitikkojen kasvavaa kiinnostusta koulutuksen tuottavuuteen ja kansakuntien taloudelliseen kehitykseen. Taustalla vaikutti vahvistuva usko koulutuksen hyvinvointia edistävään vaikutukseen. Voidaan myös sanoa, että OECD:n toimet viestivät kriittisyydestä perinteiseen koulutukseen sekä tapoihin määritellä ja mitata sen tuloksia: muiden pääosin julkisesti rahoitettujen instituutioiden tavoin koulutus on tulosvastuullinen ja tilivelvollinen sitä ylläpitäville kansalaisille. Koulutusta ei tule tarkastella vain instituution omista lähtökohdista vaan suhteessa ympäristöönsä ja tulevaisuuteensa. Samalla myös koulutuksen hyötytavoite korostuu (Laukkanen 1995; Leimu 1995; OECD 2007).

Sivistyspoliittisesti tarkastellen voidaan kriittisesti todeta, ettei koulutusta ja oppimista enää haluta nähdä arvona sinänsä. Ei ainakaan niin yksituumaisesti kuin suomalaisessa koulutusajattelussa on vuosikymmeniä totuttu ajattelemaan. Sen sijaan koulutuksen arvon nähdään määrittävän entistä enemmän suhteessa yhteiskunnan hyvinvointipäämääriin ja niihin läheisesti kytkeytyviin taloudellisiin lainalaisuuksiin. Tämä on muuttamassa merkittävästi erityisesti yleissivistävän (perus)koulutuksen yhteiskunnallista asemaintia, sisältöä ja sen arvolähtökohtia. Ajattelutavan muutos ilmenee muun muassa avainkompetensseja koskevassa keskustelussa. Kompetensseilla viitataan nimenomaan sellaisiin taitoihin ja valmiuksiin, joilla on välitöntä käyttöarvoa nykyaikaisessa yhteiskunnassa (Rychen & Salganik 2003). Tämä painotus on ohjannut myös PISA-ohjelman rakentamisessa. Tehtävien valinnassa painottuu niiden käytännön relevanssi ja vastaavuus ”todellisen elämän” käyttötilanteisiin.

Samankaltainen kehitys on havaittavissa myös korkeakoulutuksessa. Huomio on siirtynyt entistä enemmän osaamisperustaisiin ja oppimiskeskeisiin opetussuunnitelmiin erityisesti Bolognan

prosessin seurauksena. Muuttuneet painotukset heijastelevat yhteiskunnassa ja työelämässä tapahtuneita muutoksia. Tutkimustulokset osoittavat selkeän ristiriidan työelämän odotusten ja koulutuksen tuottamien kompetenssien välillä. Korkeakoulututkinnot eivät välttämättä vastaa nopeasti muuttuvan työelämän tarpeisiin (Tynjälä ym. 2006; Murtonen ym. 2008). Niinpä viimeaikainen kiinnostus on kohdistunut siihen, millaisia valmiuksia ja osaamista korkeakoulutus tuottaa suhteessa työelämän tarpeisiin. Juuri tähän ongelmaan AHELO:nkin toivotaan tuottavan lisätietoa.

## KANSAINVÄLISET OPPIMISTULOSARVIOINNIT: ESIMERKKEINÄ PISA JA AHELO

### Tavoitteet ja tarkoitus

Kansainvälisissä oppimistulosarvioinneissa arviointi ymmärretään usein koulutusjärjestelmän, koulutuspolitiikan, opetus- ja koulutus-kulttuurien sekä opetussuunnitelmien toteutumisen palaute- ja kehittämisjärjestelmäksi. Siten kansainvälisillä koulutusarvioinneilla on usein monia, toisiaan täydentäviä ja paikoin jopa ristiriitaisia, tavoitteita. Tyypillisesti kansainvälisillä arvioinneilla pyritään tarkastelemaan oppilaiden/opiskelijoiden tai oppilaitosten oppimistulosten tasoa ja laatua, opetusmenetelmien tai koulutusjärjestelmien yhteyttä oppimistuloksiin sekä työelämän osaamisvaatimusten toteutumista koulutusinstituutioissa. Niinpä kansainvälisillä arvioinneilla pyritään yhtäältä saamaan tietoa koulutuksen ja opiskelun tuottamista tuloksista yksilöiden, koulujen ja yhteiskunnan näkökulmasta, toisaalta suuntaamaan huomio arvioinnin kehittävään ulottuvuuteen: miten esimerkiksi opetusmenetelmiä, opiskelijajärjestelyjä ja koulutusjärjestelmiä voitaisiin kehittää niin, että ne entistä paremmin palvelisivat laadukasta opetusta ja oppimista. (Linnakylä & Välijärvi 2005, 13–15, 22.)

Chelimskyn (1997) mukaan arvioinnilla onkin siten tarkoituksena lisätä ymmärrystä koulutuksen tilasta (tiedontuotanto), tuloksellisuudesta (tulosvastuu), ja pyrkiä tämän tiedon varassa parantamaan ja uudistamaan koulutusta ja opettamista (kehittämistarkoitus). Jälkimmäinen tavoite on toteutunut erityisesti

korkeakoulutuksessa, jossa keskeisin kansallinen arviointielin, Korkeakoulujen arviointineuvosto (KKA) on perustanut arviointinsa niin sanottuun kehittävän arvioinnin periaatteeseen. Käytännössä tämä tarkoittaa sitä, että KKA:n toteuttamat arvoinnit tukevat korkeakoulujen oman toiminnan kehittämistä, eivätkä pyri niinkään korostamaan arvioinnin tulosvastuullisuutta esimerkiksi keskeistä rahoittajaa eli valtiota kohtaan. (KKA 2010, 6.)

Linnakylän ja Välijärven (2005) mukaan kansainväliset arvoinnit ovat luonteeltaan ensisijaisesti summatiivisia ja evaluoivia. Summatiivisella arvioinnilla tarkoitetaan tulosarviointia, jossa huomio kiinnitetään ennalta asetettujen tavoitteiden ja standardien saavuttamiseen. Evaluoivassa arvioinnissa kerättyä arviointitietoa suhteutetaan koulutuspoliittisiin päämääriin ja periaatteisiin. Patton (1997) painottaakin juuri tätä ulottuvuutta: arvioinneilla on oltava käyttöä päätöksenteossa ja kehittämistyössä. Käytettävyyteen keskittyvä arviointi nostaa samalla keskiöön kysymyksen siitä, kenen arvot näkyvät arvioinneissa, ketkä käyttävät tuloksia ja miten. Arviointien tuloksia voidaankin tarkastella eri näkökulmista: koulutusjärjestelmän, koulutuksen järjestäjän, oppimisympäristön, opetuksen, oppilaan/opiskelijan tai tiedeyhteisön. Siten kansainvälisiin arviointeihin kohdistuvat odotukset ja intressit voidaan tiivistää koulutuspoliittisiksi, pedagogisiksi, tieteellisiksi ja käytännöllisiksi. (Linnakylä & Välijärvi 2005, 30–34.)

*Koulutuspoliittiset intressit* viittaavat päätöksentekijöiden koulutukselle asettamiin päämääriin ja pyrkimyksiin. Näitä intressejä kuvasimmekin jo edellisessä luvussa, mutta tiivistäen Suomessa koulutuspoliittiset linjaukset ovat viime vuosina korostaneet laadukasta ja korkeatasoista opetusta sekä tehokasta ja taloudellista koulutusjärjestelmää. Kansainväliset arvoinnit antavat siten päättäjille tietoa siitä, miten suomalainen koulutus sijoittuu suhteessa muihin koulutusjärjestelmiin. Kansainvälisten koulutusarviointien odotetaan myös lisäävän tietoa kulttuurisista ja historiallisista tekijöistä, jotka muokkaavat kansallista koulutusjärjestelmää.

*Pedagogiset intressit* kohdistuvat opetuksen ja koulutuksen kehittämiseen opetussuunnitelmien, opetusmenetelmien tai oppimistavoitteiden näkökulmista. Kansainvälisten arviointien

avulla pyritäänkin usein siirtämään hyviä käytäntöjä maasta toiseen – tästä yhtenä esimerkkinä on Suomessa kukoistava PISA-turismi. (Linnakylä & Välijärvi 2005, 31–32.)

Esimerkiksi PISA:ssa pedagogiset tavoitteet ja intressit ovat tärkeässä roolissa, sillä mielenkiinnon kohteena ovat korostetusti nuorten valmiudet käyttää koulussa ja muualla oppimaansa arkielämässä, erilaisissa työtehtävissä ja jatko-opinnoissa. Suhteessa muihin koulutuksen arviointeihin PISA:a voidaan luonnehtia siten, että kyse on osaamisperusteisesta arvioinnista. PISA ei ole niinkään kiinnostunut siitä, missä osaaminen on hankittu, joskin virallinen koulutusjärjestelmä edelleen nähdään osaamisen keskeisimpänä kehittäjänä. Osaamisen ”tuottajaneutraalius” avaa kiinnostavia tutkimuskysymyksiä myös siitä, mikä on virallisen järjestelmän ulkopuolisten vaihtoehtojen, kuten esimerkiksi verkkoympäristöjen, harrastusten, järjestöjen, merkitys nuorten osaamisen kehitykselle ja miten nämä voidaan ottaa huomioon opetusmenetelmissä ja -järjestelyissä. Tämä nostaa keskeiseksi myös kysymyksen ”virallisen” koulutuksen ulkopuolella opitun tunnistamisesta ja tunnustamisesta osaksi tutkintoja.

Osaamisen kriteeri PISA:ssa on siten oppilaan valmius ratkaista autenttisia ongelmatehtäviä erilaisissa oppimis- ja muissa ympäristöissä omaksumiaan tietoja ja taitoja soveltaen. Autenttisuus viittaa tässä sellaisten ongelmallanteiden ratkaisemiseen, joita nuoret suurella todennäköisyydellä kohtaavat jatko-opinnoissaan, erilaisissa työtehtävissä ja kansalaisina nykyaikaisessa yhteiskunnassa.

PISA-tehtävät edellyttävät usein sanomalehdistä, tieteellisistä artikkeleista, Internetistä tai kirjallisuudesta valittujen materiaalien analysointia. Niissä menestyminen edellyttää taustana olevan tiedonalan (mm. matematiikka, luonnontieteet) hyvää osaamista, mutta myös valmiutta tunnistaa olennaiset tiedot ja muokata ne ratkaistavaan muotoon moninaisissa ja jopa ristiriitaisissa tekstiympäristöissä. Missä määrin koulupedagogiikka kannustaa tällaiseen ajatteluun?

AHELO:ssa puolestaan pedagogiset intressit ovat hankkeen ytimessä: miten AHELO:n avulla saatua tietoa voidaan hyödyntää niin, että korkeakoulutuksen laatu paranee. Hankkeessa huomio kiinnitetään neljään osa-alueeseen: genee-

riisiin taitoihin, taloustieteelliseen osaamiseen ja tekniikan hallitsemiseen. Lisäksi huomion kohteena on vielä niin sanottu kontekstuaalisten tekijöiden osa-alue, jossa pyritään kehittämään keskeisiä survey-pohjaisia mittareita selvittämään oppimistulosten ja taustatekijöiden (yksilö- ja korkeakoulutasoisia) välistä suhdetta. (OECD 2010.) Hanke tuottaa siten tietoa korkeakouluille koulutuksen ja opetuksen tuottamista valmiuksista ja osaamisesta. Tämä tiedon pohjalta voidaan kehittää koulutusohjelmia ja pedagogisia toimintatapoja.

PISA:n tavoin AHELO:ssa, ja varsinkin geneeristen taitojen osa-alueessa, korostuu autenttisissa tilanteissa toimiminen. Geneerisiä taitoja arvioidaan CLA-testillä (Collegiate Learning Assessment). Testissä on kaksi suoriutumista mittaavaa osaamistehtävää (engl. performance task), jotka perustuvat tosielämän ongelmanratkaisutilanteisiin. Tietokoneavusteisesti toteuttavassa testitilanteessa opiskelijan on 90 minuutissa läpikäytävä sarja erilaisia kysymyksiä. Vastauksien tulee perustua omien kokemusten sijaan käytössä oleviin materiaaleihin, jotka tyypillisesti ovat tekstipätkiä (esim. tieteellisistä julkaisuista, sanomalehtiartikkeleista tai sähköpostikeskusteluista) tai erilaisia graafeja (kuvia, kuvioita ja taulukoita). Tehtävissä opiskelijan on osattava erottaa epäolennainen tieto olennaisesta, epäluotettava luotettavasta ja näin kirjoittaa vastauksensa esitettyihin kysymyksiin. CLA-testi mittaa opiskelijan analyttistä päättelyä ja arviointia, ongelmanratkaisua, tekstin tuottamista sekä kirjoitustaitoa. (OECD 2010.)

*Tieteelliset intresseistä* käsin tarkasteltuna arviointihankkeet ovat aina myös tutkimushankkeita. Tällöin mielenkiinto kohdistuu siihen, miten arviointikohde teoretisoidaan, käsitteellistetään, operationalisoidaan, mallinnetaan sekä millaisia metodologisia lähestymistapoja käytetään ja miten luotettavina saatuja tuloksia voidaan pitää. (Linnakylä & Välijärvi 2005, 32.) Ymmärtämällä arviointihakkeet myös tieteellisinä tutkimushankkeina pyritään varmistamaan, että tulokset ovat luotettavia ja että mittavälineitä ja tutkimusmenetelmiä jatkuvasti kriittisesti tarkastellaan ja tarvittaessa edelleen kehitetään. Siten kansainväliset arviointihankkeet ovat tärkeitä myös tiedeyhteisön näkökulmasta: niissä usein kehitetään uudenlaisia tutkimusasetelmia ja analyysimenetelmiä. Arviointihakkeiden ai-

neiston ovat laajoja ja tutkijoiden vapaasti käytettävissä arvioinnin päättymisen jälkeen, jolloin voidaan mm. tehdä riippumattomia tarkentavia lisäanalyysieja ja -tutkimuksia.

Tieteellisiin intresseihin liittyvät läheisesti myös erilaiset *käytännölliset*, arvioinnin toteuttamisen, *kysymykset*. Missä vaiheessa lukuvuotta arviointi on järkevintä toteuttaa, miten mittavälineiden käännös toteutetaan, miten oppilaitokset ja oppilaat/opiskelijat saadaan mukaan arviointiin tai miten koetettävien pisteitys järjestetään? (Linnakylä & Välijärvi 2005, 32–33.) Näiden käytännöllisten kysymysten ratkaisussa tarvitaan usein oppilaitosten vahvaa sitoutumista arviointiin, mikä osaltaan auttaa myöhemmässä vaiheessa jalkauttamaan arviointitietoa oppilaitoksiin niiden toiminnan kehittämisen perustaksi.

Seuraava kuvaus PISA:n toteuttamisesta antaa hyvän kuvan siitä, mitä tieteellisiä vaatimuksia ja käytännöllisiä odotuksia jo pelkään otantaan liittyä.

PISA:n kohdejoukon muodostavat pääsääntöisesti mittausvuonna 15 vuotta täyttävät nuoret. Suomessa kohdejoukoksi määritellään kyseisen vuoden helmikuun ja seuraavan vuoden tammikuun välisenä aikana syntyneet. Näin varmistetaan se, että lähes kaikki PISA:an valikoituvat oppilaat ovat peruskoululaisia. PISA-tutkimukseen oppilaat valitaan kaksivaiheisella satunnaisotannalla. Ensimmäisessä vaiheessa valitaan koulut kaikista niistä oppilaitoksista, joissa opiskelee PISA-ikäkriteerin täyttäviä oppilaita. Oppilaitosten otantapainot määräytyvät niiden oppilasmäärien mukaan, joten koulun todennäköisyys valikoitua PISA-otokseen on suoraan verrannollinen sen 15-vuotiaiden oppilaiden määrään. Suomessa kouluotanta on lisäksi ositettu siten, että maan eri osat muodostavat oman ositteensa. Lisäksi kullakin alueella taajama- ja haja-asutusalueiden koulut muodostavat oman ositteensa. Ruotsinkielisistä kouluista muodostetaan oma osite. Jokaisesta ositteesta valitaan kouluja vähintään tietty minimimäärä, jotta alueiden väliseen vertailuun saadaan maan kaikista osista riittävä määrä dataa. Näin ollen Itä- ja Pohjois-Suomesta kouluotokseen valikoituu suhteessa enemmän kouluja kuin Etelä-Suomesta. Koko maan tuloksia analysoitaessa alueiden väliset erot otantasuhteissa korjataan painokertoimilla. Kouluotoksen minimikoko

kussakin maassa on 150. Vuoden 2009 PISA:an Suomessa osallistui 203 koulua. Koulujen määrää kasvatti se, että tässä tutkimuksessa ruotsinkielisistä kouluista otettiin yliotus luotettavien kieliryhmien välisen vertailun mahdollistamiseksi.

Otostetuissa kouluissa oppilaat valitaan määrävälilotannalla siten, että oppilasotoksen kooksi saadaan vähintään 35 oppilasta. Otantaa varten koulut toimittavat listan kaikista PISA:n ikäkriteerin täyttävistä oppilaistaan, mukaan lukien myös erityisoppilaat ja maahanmuuttajaoppilaat. Mikäli koulun oppilaiden määrä koulussa on pienempi kuin 35, valitaan tutkimukseen kaikki oppilaat. Jokainen oppilas vastaa tehtävälomakkeeseen, mihin varataan kaksi tuntia (120 min). Rotatoimalla tehtäviä oppilaille tuotetaan tosistaan osin poikkeavat tehtävälomakkeet, yhteensä 13 erilaista variaatiota. Näin käytettyjen tehtävien kokonaismäärä saadaan laajennettua seitsemään tuntiin ja vastaavasti mitattavat tietoluheet saadaan katettua laajemmin.

Otoksen minimikoko yksittäisessä maassa on 4500 oppilasta. Suomessa PISA:an on valittu eri mittauskerroilla satunnaisesti 4500–6000 oppilasta. PISA-mittauksen lähtökohtana on, että sen tulee mahdollisimman hyvin kattaa kaikissa maissa koko ikäluokka. Kansainvälisenä standardina on, että otantaa varten tavoitetaan vähintään 95% PISA-kriteerin täyttävästä ja koulujärjestelmän piirissä olevasta ikäluokasta. Suomessa osuus on ollut eri mittauksissa 97–98% ikäluokasta. Otannan ulkopuolelle on jätetty pääsääntöisesti erityiskoulut, joissa opiskelee hieman runsas 2% vaikeimmin vammaisista nuorista. Kouluissa oppilaita ei voi jättää pois mittauksesta kuin erityisen perustelluista syistä, joiden kriteerit on täsmällisesti määritelty kaikille maille ja kouluille samat.

Yllä oleva PISA-esimerkki kuvastaa hyvin sitä, mitä tieteellisiä vaatimuksia laajoihin kansainvälisiin vertailuhankkeisiin liittyy. Täsmälliset ja tiukasti kontrolloidut tieteelliset standardit ovat välttämätön ehto sille, että eri maissa tuotettavien aineistojen vertailtavuus voidaan varmistaa. Hankkeissa pitää siten olla osaava henkilöstö joka hallitsee tieteelliset (mittareiden käänös, otannat, aineiston keruu ja analyysi), viestinnälliset (tulosten julkaisu, vuorovaikutus monikulttuurisessa tutkimusympäristössä jne.), johtamiseen liittyvät (henkilöstön palkkaus, ta-

loushallinto, kilpailutus jne.) ja käytännölliset (käännösten organisointi, hankkeen eri vaiheiden ajoitus, koetehtävien pisteytys jne.) vaatimukset. (Kupari & Välijärvi 2005.)

### **Kansainvälisten arviointien edut ja haasteet**

Kansainvälisiin arviointeihin liittyy erilaisia ja erityisiä etuja ja haasteita. Parhaimmillaan kansainväliset vertailut täydentävät kansallisia arviointeja osoittamalla koulutusjärjestelmän heikkoudet ja vahvuudet. Kansainväliset arvioinnit mahdollistavat siten kansallisen koulutusjärjestelmän tarkastelemisen laajemmalla kehyksellä. Täten kansainvälisissä vertailuhankkeissa on aina sisäänrakennettuna benchmarking-ulottuvuus. Esimerkiksi koulutusjärjestelmiltään ja opetuskulttuureiltaan hyvinkin erilaiset maat voivat saavuttaa yhtäläiset oppimistulokset. Tällöin mielenkiintoista onkin tarkastella, mistä tämä johtuu: koulutusjärjestelmästä, opetusjärjestelystä, opettajien pedagogisesta osaamisesta vai jostain muusta. (Leimu ym. 2001). Näiden kysymysten ratkaisemiseen laajojen kansainvälisten hankkeiden aineistot antavat usein paremmat edellytykset kuin kansalliset arvioinnit.

Kansainväliset arvioinnit tuovat myös lisää vaihtoehtoja kansallisten koulutusjärjestelmien kehittämiseen, sillä tällöin voidaan tutustua erilaisiin, samaan ongelmaan liittyviin ratkaisumalleihin ja arvioida erilaisten ratkaisujen toimivuutta ja tuloksellisuutta. Näin voidaan välittää ja hyödyntää eri konteksteissa kehitettyjä hyviä käytäntöjä ja katsoa, toimivatko ne erilaisissa koulutusjärjestelmissä ja -kulttuureissa. (Raivola 1995.) Suomen PISA-menestys onkin saanut osakseen maailmanlaajuisia huomiota, mikä on merkittäväällä tavalla lisännyt suomalaisen koulujärjestelmän ja opettajankoulutuksen tunnettuutta.

Kansainvälisten arviointien laajat aineistot mahdollistavat myös erilaisten selitys- ja ennustemallien tekemisen. Näin voidaan kehittää uusia käsitteellisiä kehyksiä ja entistä luotettavampia mittavälineitä oppimistuloksien kansainväliseen tutkimiseen. Tästä esimerkkinä on, että PISA:ssa ollaan aikaisempien vuosien kokemusten perusteella siirtymässä entistä enemmän monivalintatehtävien sijaan autenttisiin



tehtävyytyyppeihin. Tämä heijastaa PISA:n taustalla olevan käsitteellisen kehysten painopisteen siirtymisen kouluoppimisesta osaamisen (koulussa opitun soveltamisen) tutkimiseen. Tutkimusperustaisuus onkin keskiössä laajoissa vertailevissa oppimistuloksia mittaavissa hankkeissa, mikä takaa sen, että myös teoreettiset ja metodologiset kysymykset nähdään olennaisena osana arviointitutkimusta. Arviointiohjelmaa edeltääkin usein vuosien esiselvitysvaihe, jossa rakennetaan tieteellisesti luotettava arviointikehys. (Döbert ym. 2004; OECD 2007.) Esimerkiksi AHELO:n esiselvitysvaihe, jossa määritetään käsitteellinen kehys, laaditaan ja käännetään mittavälineet, kerätään 2000 opiskelijan aineisto, analysoidaan ja raportoidaan tulokset, kestää 2,5 vuotta.

Kansainvälisistä oppimistulosarvioineista hyötyvät useat eri tahot. Esimerkiksi OECD:n ja IEA:n arviointien tuloksista tiedotetaan oppilaille, vanhemmille, kouluille, hallinnolle, poliitikoille, medialle ja tutkijoille. Erityisesti koulut ja opettajat saavat näin tärkeää tietoa oman toimintansa kehittämiseksi. Päätöksentekijöille arvioinnit tuottavat usein tiettyihin indikaattoreihin perustuvaa vertailutietoa. Vaikka indikaattorit antavatkin pelkistetyn kuvan oppimistulosten laadusta ja muutoksesta, paljastavat ne kuitenkin keskeisimmät vahvuudet ja pahimmat puutteet. Tämä tieto auttaa tarkentamaan kansallisia kehittämiskohteita ja koulutuspoliittisia linjauksia. (Laukkanen 1995; Leimu 1995.)

Tämä tulosten käytettävyys näkyy myös AHELO:ssa, jossa tarkoitus on antaa korkeakoulukohtainen palaute. Näin kukin korkeakoulu voi hyödyntää tietoja oman toimintansa kehittämisessä ja koulutuksellisten heikkouksiensa ja vahvuksiensa tunnistamisessa. Opiskelijat saavat myös aikaisempaa tarkempaa tietoa siitä, mikä korkeakoulu kannattaa valita opinahjoksi ja pidemmällä aikavälillä parempaa koulutusta ja opetusta. Työnantajille AHELO:n toivotaan tuottavan tietoa siitä, miten hyvin korkeakoulutus valmistaa siirtymään työelämään ja erityisesti, miten onnistuneesti korkeakoulutuksessa opitaan työelämässä tarvittavia tietoja ja taitoja. Kansallisille päätöksentekijöille AHELO antaa tietoa siitä, käytetäänkö korkeakoulutuksen suunnattuja resursseja tarkoituksenmukaisesti ja tehokkaasti oppimisen näkökulmasta. (OECD 2010.)

Kansainvälisiin arviointihankkeisiin liittyy myös monenlaisia haasteita. Yksi keskeisimmistä haasteista on vertailtavuuden saavuttaminen, sillä kukin koulutusjärjestelmä ja -kulttuuri on omanlaisensa historiallisen kehityksen tulos. Vertailtavuuden haasteet liittyvät muun muassa yhteiskunnallisiin eroihin (tasa-arvo, oikeudenmukaisuus jne.), koulutusjärjestelmien erilaisuuteen (rinnakkaiskoulu/yhtenäiskoulu jne.), opetussuunnitelmallisiin tekijöihin (oppimiskäsitykset, oppiaineiden jäsentäminen jne.), koulutuksen resursointiin ja arviointitotumuksiin. (Linnakylä & Välijärvi 2005, 53–54.) AHELO:ssa osallistuviin 15 maahan mahtuu hyvin erilaisia yhteiskuntia (pohjoismainen hyvinvointi-ideologia, islamilaisen kulttuurin yhteiskunnat jne.), korkeakoulujärjestelmiä (julkinen vs. yksityinen, vanha vs. nuori järjestelmä jne.) ja koulutukseen suunnattuja resursseja (yhdyksvaltalaiset huippuyliopistot ns. Ivy League vs. kolmansien maiden korkeakoulut). Niinpä voidaan perustellusti kysyä, missä määrin on perusteltua vertailla hyvinkin erilaisten maiden oppimistuloksia keskenään.

Vertailtavuuden saavuttamiseen liittyy läheisesti ekvivalenssin vaatimus (Arffman 2007). Ekvivalenssilla tarkoitetaan riittävää samankaltaisuutta arvioinnin mahdollistamiseksi. Usein ekvivalenssia tarkastellaan eri näkökulmista: kulttuurisesta, kehyksellisestä, rakenteellisesta, toiminnallisesta, korrelatiivisesta ja geneettisestä. Kulttuurisella ekvivalenssilla viitataan koulutuksen ja opetuksen kulttuuristen piirteiden samankaltaisuuteen. Esimerkiksi suomalais-ta koulujärjestelmää voidaan helposti verrata pohjoismaisiin kouluihin samankaltaisten kulttuuristen tekijöiden vuoksi. Tilanne on toinen kun suomalaista koulujärjestelmää verrataan esimerkiksi Aasian maihin. Kansainvälisissä hankkeissa kulttuurinen ekvivalenssin haasteellisuus korostuu mittavälineiden käännoksissä. Mittavälineiden tulisi olla keskenään vertailukelpoisia, eikä käännoksen seurauksena eri maiden vastaajat saa joutua eriarvoiseen asemaan. Tätä käännosten ekvivalenssia pohdittiin esimerkiksi AHELO:n mittavälineiden laadinnassa: missä määrin tehtävissä voi esiintyä sellaisia elementtejä, jotka eivät välttämättä kaikissa kulttuureissa ole itsestään selviä tai edes tuttuja. Ymmärretäänkö esimerkiksi järvi samalla tavoin Suomessa ja Kuwaitissa?

Kehysekvivalenssilla tarkoitetaan taloudellisten, sosiaalisten ja fyysisten tekijöiden samankaltaisuutta. Esimerkiksi kehitysmaissa koulutukseen suunnatut taloudelliset resurssit ovat huomattavasti vähäisemmät kuin teollistuneissa länsimaissa. Tällöin voidaan kysyä, onko perusteltua verrata resurssiltaan erilaisia maita keskenään. Rakenteellinen ekvivalenssi puolestaan kiinnittää huomiota koulutusjärjestelmien ja -rakenteiden samankaltaisuuteen. Perusasteen arvioinneissa on havaittu, että yhtenäiskoulutusjärjestelmissä koulujen väliset erot ovat pienemmät kuin rinnakkaiskoulutusjärjestelmissä, joissa oppilaiden väliset erot koulujen sisällä ovat pienemmät. (OECD 2007) Esimerkiksi yksi keskeisin AHELO:n haasteista on, miten ottaa huomioon erilaiset korkeakoulut ja korkeakoulutusjärjestelmät. Korkeakoulututkimuksessa ei ole luokittelujärjestelmää, jonka avulla voitaisiin luotettavasti erotella erilaiset korkeakoulutyypit. Siten erilaisten korkeakoulukulttuurien ja -järjestelmien vaikutusta opiskelijoiden oppimistuloksiin on haasteellista arvioida.

Toiminnallinen ekvivalenssi liittyy opetus-suunnitelmien erilaisiin tavoitteisiin, sisältöihin ja menetelmiin sekä arviointikäytäntöihin. Esimerkiksi useissa Keski-Euroopan maissa oppimista arvioidaan loppukokeilla, mikä taas ei ole tyypillistä Suomessa. Toiminnalliseen ekvivalenssiin luetaan kuuluvaksi myös etnisten ja kielellisten vähemmistöjen tasa-arvoinen asema. Suomessa kaikilla kouluasteilla etnisten vähemmistöjen osuus on huomattavasti pienempi kuin vaikkapa muissa Pohjoismaissa. Geneettinen ekvivalenssi puolestaan viittaa oppimisen taustalla olevien arvostusten, oppimiskäsitysten, pedagogisten periaatteiden ja oppimisympäristöjen samankaltaisuuteen. Missä määrin esimerkiksi oppimismahdollisuuksien tasa-arvoa pidetään kaiken oppimisen perustana? (Leimu 2004.)

Vertailtavuuden (ekvivalenssin) saavuttamisen lisäksi toinen keskeinen haaste liittyy kansainvälisten ja kansallisten arviointien yhteensovittamiseen. Aina kansalliset ja kansainväliset tarpeet eivät kohtaa: se, mikä on kansainvälisesti ajankohtaista, ei välttämättä juuri sillä hetkellä ole kansallisesti olennaista. Lisäksi kansainväliset vertailuhankkeet saattavat tuottaa tietoa, joka on ristiriidassa kansallisen koulutusjärjestelmän kanssa. Esimerkiksi jos arviointitulokset osoittavat, että tulosstandardeihin perustuvat

opetusjärjestelyt tuottavat parempia tuloksia kuin väljemmin määritellyt opetussuunnitelmat, voivat hajautetun koulutusjärjestelmän maat joutua epäsuotuisaan valoon. Tällöin arvioinnin tuottamia tuloksia on syytä tarkastella kriittisesti, eikä ottaa niitä itsestään selvyyksinä ja kansallisen koulutuspolitiikan uudelleenmäärittämisen lähtökohtina. Näin erityisesti siksi, että kansainvälisissä arviointihankkeissa käytetyt käsitteelliset kehykset ovat neuvoteltuja kompromisseja, eivätkä siten kuvaa kansallisia opetus- ja oppimisympäristöjä. (Linnakylä & Välijärvi 2005, 61–63.)

Kolmas haaste on, miten kansainvälisten vertailuhankkeiden tuottamaa tietoa käytetään. Miten tietoa hyödynnetään esimerkiksi koulutuspolitiikan tekemisessä? Kuten jo koulutuspolitiikkaa ja arviointia käsittelevässä luvussa mainitsimme, kansainvälisten vertailuhankkeiden tuottama tieto toimii usein päätöksentekoa ohjaavana ja sitä muokkaavana (ks. Kallo 2009). Esimerkiksi vuonna 2010 perusopetuksen yleisten valtakunnallisten tavoitteiden ja perusopetuksen tuntijakoa pohtinut työryhmä nosti vahvasti kansainvälisen vertailutiedon, erityisesti PISA:n tulokset, työnsä pohjaksi (ks. OKM 2010). Tiedon käytettävyyttä (vrt. Patton 1997) liittyy kansainvälisissä arvioinneissa myös tulosten julkistamiseen. Eri maissa on erilaisia julkisuusperiaatteita, mikä osaltaan vaikeuttaa yhtenäisten ja oikeudenmukaisten tiedottamis- ja julkaisuohjeistusten tekemistä. Toisaalla (esimerkiksi Englannissa) pidetään tärkeänä, että arviointitieto on mahdollisimman julkista, kun taas toisaalla (kuten Suomessa) arviointitiedon julkisuuteen suhtaudutaan pidättyvämmiin. Suomessa esimerkiksi PISA:ssa yksittäisten koulujen tulosten julkistamiseen suhtaudutaan epäillen, sillä suomalaisten koulujen laatuero eivät ole keskeinen koulutusjärjestelmän ongelma, kun taas Englannissa tämä tieto on olennaista vanhempien valitessa jälkikasvunsa opinahjoja. (Simola 2005; van Petegern ym. 2005.)

### Arvioinnin etiikka

Neljäs, ja kenties keskeisin, haaste liittyy arvioinnin eettisiin kysymyksiin, jotka nousevat kansainvälisissä arvioinneissa esiin monessa eri muodossa. Keskeisimpiä kysymyksiä on, miten arvioida oikein, hyvin ja arvokkaasti, kun pyri-

tään hyödyllisyyteen, yhteiskunnalliseen vakuutavuuteen ja monien eri intressitahojen palvelemiseen (Patton 1997). Koulutuksen arviointi on vahvaa vallankäyttöä järjestelmiä, instituutioita ja niissä toimivia ihmisiä kohtaan. Siksi eettinen tietoisuus hankkeiden yleisistä ja arvioitsijan henkilökohtaisista motiiveista ja päämääristä on olennainen osa arvioinnin asiantuntijuutta.

Arvioinnin kulttuurinen, sosiaalinen ja poliittinen herkkyyksy liittyy olennaisena osana kansainvälisiin vertailuihin. Vaikka laadukas arviointi on tutkimuksen luonteista, sen ensisijaisena tarkoituksena on toiminnan tai tuloksen arvon määrittäminen (Scriven 1994). Arvioinnin tavoitteet, kriteerit ja arvioitsijan intressit väistämättä suuntaavat sitä, mitä, miten ja millä seurauksilla arvioidaan. Koulutuksen arvioinnilla tulisi kuitenkin aina olla oppimista, opetusta ja koulutusta, jopa yhteiskuntaa ja sen kulttuuria kehittäviä päämääriä (House 1993).

Kansainvälisten arviointien julkilausuttuna arvona painotetaan usein demokraattisia pyrkimyksiä tasa-arvoistaa koulutusjärjestelmiä esimerkiksi sukupuolen, iän, asuinpaikan, kielen, etnisen taustan, taloudellisen aseman tai oppimisedellytysten erilaisuuden suhteen (OECD 2007; Leimu 2004). Käytännössä tämä ilmenee muun muassa muuttujavalinnoissa tai tavoissa raportoida tuloksista. Maat voivat oppia toisiltaan, miten uudistaa koulujärjestelmään näitä päämääriä paremmin palvelevaksi.

Koulutuksen kansainvälisissä arvioinneissa herää jatkuvasti myös kulttuurisia, kielellisiä ja taloudellisia syrjäntäepäilyjä. Onko oikeudenmukaista vertailla tai arvioida täysin erilaisia ja eri tavoin resursoituja järjestelmiä ja niiden oppimistuloksia? Esimerkiksi ranskalaiset, italialaiset ja espanjalaiset usein syyttävät englantia puhuvia maita kielellisestä ylivallassa. Pohjoismaalaisia puolestaan huolestuttaa jatkuvasti, ettei hyvinvointiyhteiskuntamme tasa-arvoajattelua arvosteta tarpeeksi. Erilaisten kulttuuristen ja niihin liittyvien eettisten ristiriitojen pohtiminen ja tasapainottaminen käytännön arviointitilanteissa onkin kansainvälisen arvioinnin keskeisimpiä haasteita (House 1997).

Valta-asetelmat, arvostuksiin liittyvät intressit ja eettiset ristiriidat näkyvät kansainvälisissä arvioinneissa usein hyvin selvinä. Esimerkiksi PISA:ssa eri maiden asiantuntijoiden näkemykset arviointikohteesta poikkesivat merkittävästi

sekä metodologisesti että erityisesti eettisesti. Kenen ehdoilla arvioidaan – koulutuksen vai työelämän? Kansainvälisissä neuvotteluissa kompromisseihin onkin usein taivuttava, vaikka omat eettiset näkemykset poikkeaisivat enemmistön kannanotoista. (Linnakylä & Välijärvi 2005.)

Vaikka kansainvälisten arviointien huomio on koulujärjestelmien vertailussa, arviointitiedon tuottajina toimivat lähes aina oppilaat ja opettajat. Arvioitavien oikeuksiin kuuluu se, että arvioinnin tarkoitus, keskeiset kriteerit ja seuraukset selvitetään avoimesti ennakkoon. Arvioitavien oikeuksiin sisältyy myös mahdollisimman monipuolinen palautetieto, jota voi hyödyntää oppimisen tai kehittämisen välineenä. Hyödyllisimmillään arviointi tuottaa tuloksia, jotka ovat luotettavia ja edistävät oppimista ja koulutuksen kehittämistä sekä lisäävät ymmärrystä tuloksien taustalla vaikuttavista tekijöistä.

Arvioinneissa on myös raportoitava tulokset tieteellisen raportoinnin sääntöjä noudattaen. Tuloksia ei tule esimerkiksi yleistää, jos otanta ei anna siihen riittäviä perusteita. Niinpä arvioinneissa on syytä ilmaista selvästi tulosten tulkinnan rajoitukset. Kansainvälisiin vertaileviin arviointeihin kohdistuva julkisuus kohdistuu lähes yksinomaan maiden ja koulujen järjestykseen panemiseen, eikä pohdi tiedon luotettavuutta. Julkinen keskustelu helposti värittääkin ja jopa vääristää kansainvälisten arviointien tuloksia. (Linnakylä 1998.)

Koulutuksen arviointeja toteutettaessa tuleekin korostaa arvioinnin kehittävää tehtävää ja eettisiä ehtoja, jotka kunnioittavat arvioitavien oikeuksia, painottavat arvioijien pätevyyttä sekä arviointiprosessin tasapuolisuutta. Vaikka koulutuksen tuloksena vaaditaan yhä painokkaammin kilpailukykyä ja huippuosaamista, koulutuksen arvioinnissa tulisi korostaa myös tasa-arvoa ja oikeudenmukaisuutta. (Castells 1998; Linnakylä 2001.)

## POHDINTA

Kansainvälisissä arvioinneissa tärkeää on tiedostaa, että ne altistuvat jatkuvasti eri intressitahojen odotusten kohteiksi. Nämä intressit ovat läsnä arviointihankkeiden eri vaiheissa: niiden tavoitteita ja lähtökohtia määritettäessä,

Taulukko 1. Kansainvälisten vertailevien arviointien tavoitteet eri intressien näkökulmasta tarkasteltuna.

Tavoitteet	Koulutuspoliittiset intressit	Pedagogiset intressit	Tieteelliset ja käytännölliset intressit
Tiedontuotanto	Missä määrin arviointi lisää päättäjien ymmärrystä oppimistuloksista?	Missä määrin arviointi lisää oppilaitosten pedagogista tietämystä?	Miten tuottaa tieteellisesti pätevää arviointitietoa?
Tulosvastuu	Missä määrin arviointi lisää päättäjien tietoa koulutuksen tuloksellisuudesta?	Missä määrin arviointi antaa oppilaitoksille tietoa koulutustoimintansa tuloksellisuudesta?	Miten toteuttaa arviointi siten, että se kertoo myös koulutuksen tuloksellisuudesta ja taloudellisuudesta?
Kehittäminen	Missä määrin arviointi tuottaa tietoa päättäjille koulutusjärjestelmän kehittämiseksi?	Missä määrin arviointi tuottaa tietoa oppilaitoksille toimintansa kehittämiseksi?	Miten esittää tulokset siten, että arviointitieto voi toimia koulutusjärjestelmien ja oppilaitosten kehittämisen perustana?

toteutuksessa sekä tulosten raportoinnissa, julkistamisessa ja jalkauttamisessa. Siten eri intressitahojen edustajat joutuvatkin usein esittämään itselleen ja muille kysymyksiä, mitä heidän mielestään kansainvälisillä arvioinneilla oikein tavoitellaan (taulukko yksi).

Koulutuspoliittiset intressit liittyvät vahvasti kansallisten ja ylikansallisten päätöksentekijöiden linjauksiin koulutusjärjestelmien kehittämisen suuntaviivoista arviointitietoon pohjatuena. Juuri tämä on tieteelliseen näyttöön perustuvan päätöksenteon (evidence-based policy) ydin. Pedagogiset intressit jalkautuvat oppilaitosten – on kyse sitten perus- tai korkeakoulusta – arkeen: miten arviointien tuottamaa tietoa voidaan käyttää oppilaitosten pedagogisten käytänteiden, opetusmenetelmien ja toimintatapojen kehittämisessä ja parantamisessa. Tieteelliset ja käytännölliset intressit koskevat arviointien toteuttajia ja tiedeyhteisöä, jonka tehtävänä on valvoa ja taata arviointien korkea tieteellinen taso.

Usein koulutuspoliittiset, pedagogiset ja tieteelliset intressit voivat olla keskenään ristiriitaisia, mikä tekee arviointihankkeiden to-

teuttamisesta paikoin haasteellista. Erityisesti tutkimusyhteisö on vastuussa siitä, että vertailtavuuteen (ekvivalenssi) liittyvät kysymykset ratkaistaan tieteellisesti ja oikeudenmukaisesti. Koulutuspolitiikan tekijöiden harteilla on kansallisten ja kansainvälisten arviointien tuottaman tiedon yhteensovittaminen siten, että se mahdollisimman hyvin palvelee koulutusjärjestelmän kokonaisvaltaista kehittämistä ottaen samalla huomioon kulttuuriset ja kansalliset piirteet. Tärkeää on myös, että arviointien tuottama tieto saavuttaa oppilaitokset ja oppilaat/opiskelijat siten, että koulutusinstituutioissa pystytään kehittämään opetus- ja opiskeluympäristöjä.

Mikäli kansainvälisten arviointien edut pystytään maksimaalisesti hyödyntämään, voidaan lisätä koulutuspoliittista, pedagogista ja tieteellistä ymmärrystä oppimistuloksista eri koulutusasteilla, soveltaa eri maiden ja kulttuurien hyviä käytäntöjä toisiin koulutuskonteksteihin, kehittää oppimistulosarviointien käsitteitä ja menetelmiä sekä ennen kaikkea parantaa opiskelu- ja oppimisympäristöjä eettisesti oikeudenmukaisella tavalla.

## LÄHTEET

- Arffman, I. (2007). *The problem of equivalence in translating texts in international reading literacy studies. A text analytic study of three English and Finnish texts used in the PISA 2000 reading test.* Tutkimuksia 21. Jyväskylä: Koulutuksen tutkimuslaitos.
- ARWU. (2010). *The Academic ranking of world universities.* Institute of Higher Education Shanghai Jiao Tong University. Saatavana [www-muodossa: http://www.arwu.org/](http://www.arwu.org/) [viitattu 30.9.2010].
- 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 (Eds.) *Evaluation for the 21<sup>st</sup> century. A Handbook.* Thousand Oaks: Sage, 1–26.
- Döbern, H., Klieme, E. & Stroka, W. (E.) (2004). *Conditions of school performance in seven countries.* Münster: Waxman.
- OECD. (2010) *Education at a glance.* Paris.
- Guba, E. G. & Lincoln, Y. S. (1989). *Fourth generation evaluation.* Newbury Park, CA: SAGE Publications.
- Harisalo, R., Aarrevaara, T., Stenvall, J. & Virtanen, P. (2007). *Julkinen toiminta – julkinen politiikka.* Tampere: Tampereen yliopistopaino.
- Head, B. W. (2008). Three lenses of evidence-based policy. *The Australian Journal of Public Administration* 67 (1), 1–11.
- House, E. R. (1993). *Professional evaluation. Social impact and political consequences.* Newbury Park, CA: SAGE Publications.
- House, E. R. (1997). Ethics of evaluation studies. Teoksessa J.P. Keeves (Ed.), *Educational research methodology, and measurement: an international handbook.* 2<sup>nd</sup> edition (s. 257–261). Cambridge: Cambridge University Press. Pergamon.
- Huusko, M. (2009). *Itsearviointi suomalaisissa yliopistoissa: arvoja, kehittämistä ja imagon rakentamista.* Kasvatusalan tutkimuksia 46. Jyväskylä: Suomen kasvatustieteellinen seura.
- Juva, S., Kangasvieri, A. & Välijärvi, J. (2009). *Kuntaperustaisen koulujärjestelmän kehittäminen.* Helsinki: Kuntaliitto.
- Kallo, J. (2009). *OECD education policy. A comparative and historical study focusing on the thematic reviews of tertiary education.* Kasvatusalan tutkimuksia 45. Jyväskylä: Suomen kasvatustieteellinen seura.
- KKA. (2010). *Korkeakoulujen arviointineuvoston toimintasuunnitelma 2010–2013.* Korkeakoulujen arviointineuvoston julkaisuja 8:2010.
- Kupari, P. & Välijärvi, J. (2005). *Osaaminen kestävällä pohjalla.* Jyväskylä: Koulutuksen tutkimuslaitos.
- Laukkanen, R. (1995). Koulutusindikaattorien kehittämisen kansainvälisiä lähtökohtia. Teoksessa Laukkanen, R. (toim.), *Koulutusjärjestelmän arviointia* (s. 55–74). Opetusministeriö, Opetushallitus. Helsinki: Hakapaino.
- Leimu, K. (1995). Koulutusindikaattorien tuottamiseen liittyviä periaatteita. Teoksessa Laukkanen, R. (toim.), *Koulutusjärjestelmän arviointia* (s. 75–97). Opetusministeriö, Opetushallitus. Helsinki: Hakapaino.
- Leimu, K. (2004). *Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa.* Jyväskylän yliopisto. Koulutuksen tutkimuslaitos.
- Leimu, K., Linnakylä, P. & Välijärvi, J. (2001). *Merging national and international interests in educational system evaluation.* University of Jyväskylä. Institute for Educational Research.
- 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.
- Linnakylä, P. & Välijärvi, J. (2005b). Kansainvälinen arviointiyhteistyö hyödyntämään kansallista arviointia. Teoksessa H. Lyytinen & A. Räsänen (toim.), *Kehittämissuuntaa arvioinnista* (s. 13–25). Koulutuksen arviointineuvoston julkaisuja (6). Jyväskylä yliopisto: Koulutuksen arviointisihteeristö.
- Murtonen, M., Olkinuora, E., & Tynjälä, P. & Lehtinen, E. (2008). Do I need research skills in working life? University students' motivation and difficulties in quantitative methods courses. *Higher Education* 56 (5), 599–612.
- Neave, G. (1998). The evaluative state reconsidered. *European Journal of Education* 33 (3), 265–284.
- OECD. (2007). *PISA 2006 Science competencies for tomorrow's world.* Paris:OECD.
- OECD. (2009). *PISA 2009 Assessment framework – Key competencies in reading, mathematics and science.* Paris: OECD.
- OECD. (2010). *Roadmap for AHELO feasibility study.* Third version (22 February 2010). Saatavana [www-muodossa: www.oecd.org/edu/ahelo](http://www.oecd.org/edu/ahelo) [viitattu 28.9.2010].
- OKM. (2010). *Perusopetus 2020 – yleiset valtakunnalliset tavoitteet ja tuntijako.* Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2010:1. Yliopistopaino.

- Ozga, J. (2010). Trans-national technologies and national contexts: a comparative analysis of education policy in Scotland and England. Teoksessa J. Kauko, R. Rinne & H. Kynkäänniemi (Eds.), *Restructuring the truth of schooling – Essays on discursive practices in the sociology and politics of education. A Festschrift for Hannu Simola* (s. 61–84). Jyväskylä: Finnish Educational Research Association.
- Palonen, M. (2010). Mitä uusia eväitä AHELO tuo korkeakoulupolitiikan muotoiluun? Teoksessa Laukkanen, R. (toim.), *PISA, PIAAC, AHELO. Miksi ja miten OECD mittaa osaamista?* (s. 71–73). Opetus- ja kulttuuriministeriön julkaisuja 2010:17.
- Patton, M. Q. (1997). *Utilization-focused evaluation. The new century text*. 3. painos. Thousand Oaks: Sage.
- Perusopetuslaki. (1998). Laki 628/21.8.1998.
- Power, M. (1999). *The audit society: Rituals of verification*. Oxford: Oxford University Press.
- Raivola, R. (1995). Mitä evaluation on ja mihin sitä tarvitaan? Teoksessa A. Kajanto (toim.), *Aikuiskoulutuksen arviointi. Panoraamoja ja lähikuvia* (s. 21–60). Aikuiskoulutuksen 36. vuosikirja. Kansanvalistusseura & Aikuiskasvatuksen tutkimusseura. Helsinki: Kirjastopalvelu.
- Rinne, R. & Simola, H. (2005). Koulutuksen ylikansalliset paineet ja yliopistojen uusi hallinta. *Tiede & Edistys* 30 (1), 6–26.
- Rychen, D. & Salganik, L. (2003). A holistic model of competence. Teoksessa D. Rychen & L. Salganik (Eds.), *Key competencies for successful life and well-functioning society* (s. 63–108). Göttingen: Hogrefe & Huber.
- Scriven, M. (1994). Evaluation as a discipline. *Studies in Educational Evaluation* 20 (1), 147–166.
- Simola, H. (2005). Koulukohtaiset oppimistulokset ja julkisuus. *Yhteiskuntapolitiikka* 70 (2), 179–187.
- Simola, H. & Rinne, R. (2004). Laatuvalankumous, arviointiteollisuus ja Korkeakoulut. *Kasvatus* 35 (3), 330–341.
- Tynjälä, P., Slotte, V., Nieminen, J., Lonka, K. & Olkinuora, E. (2006). From university to working life: Graduates' workplace skills in practice. Teoksessa P. Tynjälä, J. Välimaa & G. Boulton-Lewis (Eds.), *Higher education and working life: Collaborations, confrontations and challenges* (s. 73–88). Amsterdam: Elsevier.
- Uljens, M. (2007). The hidden curriculum of PISA – the promotion of neo-liberal policy by educational assessment. Teoksessa S.T. Hopman, G. Brinek & M. Retzl (Eds.), *PISA zuefolg – PISA according to PISA. Does PISA keep what it promises?* (s. 295–303). Vienna: LIT Verlag.
- Ursin, J. & Hyytinen, H. (2010). Mitä korkeakoulutuksessa opitaan? Esimerkkinä AHELO. Teoksessa Laukkanen, R. (toim.), *PISA, PIAAC, AHELO. Miksi ja miten OECD mittaa osaamista?* (s. 65–70). Opetus- ja kulttuuriministeriön julkaisuja 2010:17.
- van Petegerm, P. & Vanhoof, J. & Daems, F. & Mathieu, P. (2005). *Publishing information on individual schools? Education Research and Evaluation* 11 (1), 45–67.
- Väljjarvi, J., Linnakylä, P., Kupari, P., Reinikainen, P. & Arffman, I. (2007). *The Finnish success in PISA – and some reasons behind it*. Jyväskylä: Institute for Educational Research.
- Whitty, G. (2010). Marketization and post-marketization in contemporary education policy. Teoksessa J. Kauko, R. Rinne & H. Kynkäänniemi (Eds.) *Restructuring the truth of schooling – Essays on discursive practices in the sociology and politics of education. A Festschrift for Hannu Simola*. Jyväskylä: Finnish Educational Research Association, 23–41.