

Mitä näytöt näyttävät?

Luotettavuus ja luot- tamus näyttöperus- teisessa arvioinnissa

Mari Räkköläinen

Opetusneuvos, KT

Opetushallitus

mari.rakkolainen@oph.fi

Johdanto

Ammatillisten perustutkintojen opetussuunnitelmaperusteisen koulutuksen opiskelijan arviointia on uudistettu siten, että ammatillista osaamista arvioidaan myös näyttöperusteisesti. Opetusministeriö teki päätöksen vuonna 1998 siitä, että ammatillisiin perustutkintoihin liitetään näyttöön perustuva osoitus ammatillisten opintojen tavoitteiden saavuttamisesta (opetusministeriön päätös 212/430/98). Näyttöä

kutsutaan ammattiosaamisen näytöksi, joka on koulutuksen järjestäjän ja työelämän yhdessä suunnittelema, toteutettava ja arvioima työtilanne tai työprosessi, joka voidaan toteuttaa työpaikoilla tai mahdollisimman aitoina työelämän toimintakokonaisuuksina oppilaitoksissa. Opiskelija osoittaa näytössä, miten hyvin hän on saavuttanut opetussuunnitelman perusteiden ammatillisten opintojen tavoitteissa määritellyn ja työelämän edellyttämän ammattitaidon. Opiskelijan osaamista arvioivat ja ammattiosaamisen näyttöjen arvioinnista päättävät opettaja ja työelämän edustaja joko yhdessä tai erikseen. Näyttöjä to-

teutetaan koko koulutuksen ajan (L 601/2005, 25 §; A 603/2005; Opetushallituksen määräys M 32/011/2005).

Ammattiosaamisen näytöille asetettiin paljon odotuksia. Näytöillä haluttiin yhdenmukaistaa ja monipuolistaa opiskelijan arviointia sekä tuoda myös työelämän edustajat mukaan arvioimaan ammatillisen osaamisen tasoa. Lisäksi näytöillä haluttiin varmistaa ammatillisen koulutuksen laatua, vahvistaa koulutuksen työelämävastaavuutta ja kehittää oppimistulosten arviointia. Aikaisemmin kansallinen arviointitieto oppimistuloksista tuotettiin erillisillä tehtävillä ja testeillä, ns. kansallisilla kokeilla. Nyt tavoitteeksi asetettiin arviointijärjestelmän uudistaminen siten, että näytöistä voidaan koota suoraan vertailukelpoista tietoa ammatillisen koulutuksen oppimistuloksista ja vaikuttavuudesta, jolloin luovutaan erillisistä kansallisista kokeista. Näytöistä saatavan arviointitiedon haluttiin myös tukevan oppilaitoksia niiden oman toiminnan kehittämisessä ja itsearviointissa (Koulutuksen ja tutkimuksen kehittämissuunnitelma 1999 - 2004; L 601/2005 ja A 603/2005; Opetushallituksen määräys ammatillisten perustutkintojen opetussuunnitelman perusteista M 32/011/2005; opetusministeriön työryhmien muistioita 14:1999).

Koulutuksen järjestäjien on pitänyt ottaa näytöt käyttöön vuonna 2006 ja vuodesta 2007 näytöt ovat olleet myös kansallisen oppimistulosten arvioinnin perusta (Opetusministeriön päätös (3/502/2007) 12.2.2007). Kehittämistyö on toteutettu vuosina 1999-2008 laajoina valtakunnallisina kehittämissuunnitelmina pääasiassa Euroopan sosiaalirahaston (ESR) tuella.

Kansallista näyttöperusteista oppimistulosten arviointijärjestelmää kehitettiin ja testattiin Opetushallituksen johtamassa kansallisessa hankkeessa ja siinä toteutetuissa arviointikokeiluissa 2002 - 2008 (ks. tarkemmin Räcköläinen 2005 a; Räcköläinen & Ecclestone 2005). Tämä artikkeli perustuu tutkimukseeni, jonka aineistona ovat olleet kehittämishankkeen viralliset asiakirjat ja työryhmien muistiot, käytännön kokeiluhankkeiden tulokset, osallisten (opettajien, opiskelijoiden ja työpaikahojaajien) kokemukset sekä kansallisen hankkeen ulkoinen arviointi ja palaute. Hankkeessa oli neljä merkittävää kansallista työryhmää ja kymmenen oppilaitosten koordinoimaa arvioinnin kehittämisprojektia, joissa toteutettiin kaksi arviointikokeilua. Arviointikokeiluissa koottiin ja analysoitiin oppimistulokset oppilaitosten toteuttamista näytöistä (3651 näyttöä 2636 opiskelijalta) ja toteutettiin laaja kysely opettajille, opiskelijoille ja työelämän edustajille (943 opettajan vastausta, 1220 opiskelijan vastausta ja 748 työpaikahojaajan vastausta).

Tutkimus on fenomenologis-hermeneuttinen ja siinä on ymmärtävä ja tuloksinainen tutkimusote. Tutkimusaineisto on jäsenetty tutkimuksen kohteen eli näyttöperusteisen arviointiprosessin mukaan eri konteksteihin. Arviointia voidaan tarkastella järjestelmien, organisaatioiden ja yksilöiden tasolla. Ammatillisten perustutkintojen arvioinnin uudistus, näyttöperusteinen arviointijärjestelmä, koskee kaikkia näitä tasoja. Arviointi käsitteenä liittyy sekä opiskelijan osaamisen arviointiin että koulutuksen hallinnolliseen ohjaukseen ja koulutusjärjestelmätason arviointitoimintaan. Arvioinnin tarkastelu eri konteksteissa, yksilön, yhteisön ja

järjestelmän tai organisaation tasoilla on välttämätöntä, jotta näyttöperusteisessa ammatillisen osaamisen arviointiprosessissa ilmenneitä arvioinnin jännitteitä voi ymmärtää. Analyysia on ohjannut **käsitteellinen analyysimalli arvioinnin ulottuvuuksien jännitteistä** (ks. kuvio 1). **Malli näyttöperusteisesta arviointiprosessista** (ks. kuvio 2) jäsentää arvioinnin tarkastelua eri konteksteissaan.

Käytännön kehittämistyö nosti esiin jo melko varhaisessa vaiheessa arviointitiedon ja koko uuden näyttöperustaisen arviointijärjestelmän luotettavuuden ja arvioinnin laadun varmistamisen (oppimistulosten arvioinnin osuvuus, vertailtavuus ja käyttökelpoisuus) haasteellisuuden. Esiin nousi myös kysymys arvioijan ja arvioinnin kohteen välisestä suhteesta ja arvioinnin kohteen osallisuudesta arvioinnin toteuttamisessa (luottamus, joustavuus ja kontrolli arviointisuhteessa). Eri arviointitraditioissa ja arviointinäkemyksissä (arvioinnin lähestymistavoissa) otetaan miltei aina kantaa arvioinnin luotettavuuskriteereihin ja arviointisuhteeseen (kontrolliin perustuva vs. vuorovaikutukseen perustuva). Muun muassa tunnettu jaottelu arvioinnin sukupolvista (Guba & Lincoln 1989) valottaa hyvin erilaisia arviointinäkemymiä. Samoin uusimpaan arviointikeskusteluun liittyvä osallistumista edistävä ja voimaannuttava arviointi (esim. Fetterman 2001, ks. myös Fetterman 1996, 1998) (ks. myös Patton 2002; Pawson & Tilley 1997; Stern 2005).

Tässä artikkelissa keskityn tarkastelemaan tutkimusteni tulosten pohjalta erityisesti arvioinnin luotettavuutta näyttöperusteisessa arviointiprosessissa sekä sitä, miten arviointitiedon luotettavuus varmistetaan ja edistetään luottamusta

uudessa arvioinnin lähestymistavassa.

Luottamus ja kontrolli arviointinäkemyksissä

Luottamuksen perusta

Arviointinäkemys on lähellä arviointiparadigman käsitettä, joka viittaa siihen, miten arvioija näkee maailman ja mitä arvioija ajattelee arviointitiedon luonteesta ja siitä, miten tietoa voidaan tulkita. Paradigma vaikuttaa ennen kaikkea siihen, miten arviointikysymykset määritetään ja miten arvioijan ja arvioinnin kohteen välinen suhde muotoutuu. Arviointinäkemukseen kuuluvia ja eri arviointinäkemykset toisistaan erottavia asioita ovat (Räkköläinen 2011):

- arvioinnin tarkoitus ja tavoite
- arviointiasetelma ja arviointimenetelmät
- arvioijan ja arvioinnin kohteen välinen suhde
- arvioinnin luotettavuuskriteeristö ja luotettavuuden varmistaminen
- arviointitiedon käyttö.

Aiempien sukupolven mittaamiseen, kuvailuun tai johtopäätöksiin perustuvista arvioinneista on siirrytty kohti neuvotteluun ja vuorovaikutukseen perustuvaa arviointia (ks. Guba & Lincoln 1989; Jakku-Sihvonen 2001). Osallistumista edistävässä ja voimaannuttavissa arvioinneissa keskeistä on arvioinnin kohteen aktiivinen rooli arviointien toteuttamisessa ja voimaantuminen omassa asiassa. Osallistaminen ja osallistuminen ovat yhteydessä luottamukseen arvioijan ja arvioinnin kohteen välillä, koska arviointiin osallistuvien suhde ei perustu ulkopuoliseen kontrolliin, vaan yhteistoiminnallisuuteen, vuorovaikutukseen ja tasavertaisuuteen. Ei kuiten-

kaan ole yksiselitteistä, edistääkö osallistuminen luottamusta vai onko luottamus edellytyksenä osallistumiselle, vaan luottamus ilmenee erityisesti arvioinnin kohteen ja arvioijan välisen suhteen luonteessa. Arvioijan rooli muuttuu ohjaavaksi, eikä täydellinen kontrolli eikä objektiivisuus (arvioijan ulkopuolisuus) ole mahdollista, vaan myös arvioija on osallisena arviointiprosessissa. Osallistava arviointi perustuu yhdessä neuvoteltuun arvoperustaan. Arvioinnin hyödyllisyys ja käyttökelpoisuus edellyttää, että arviointi toteutetaan läheisessä yhteistyössä arvioinnin tuloksia käyttävien kanssa ja arviointi kehittää arvioitavaa toimintaa (Fetterman 2001; Rossey, Lipsey & Freeman 2004; Patton 1997).

Toisaalta arvioijan rooli riippuu siitä, missä tarkoituksessa arviointia tehdään. Arviointi tilivelvollisuuden varmistamiseksi saattaa edellyttää arvioijan ulkopuolisuutta ja mahdollisimman objektiivista arviointiasetelmaa, mutta tämä ei kuitenkaan tarkoita, etteikö tilivelvollisuusevaluaatio voisi olla osallistavaa ja edistää luottamusta (ks. myös Räcköläinen 2005b). Osallistavan arvioinnin puolestapuhujatkin pitävät etenkin tilivelvollisuusarvioinnissa arvioinnin laadun kannalta tärkeänä arvioijan riittävää etäisyyttä ja ulkopuolisuutta, koska tällöin varmistetaan, etteivät arvioijan omat intressit ja motiivit ohjaa arvioitavien valintoja. Ulkopuolisuudella ja etäisyydellä arviointikohteesta varmistetaan myös, että huomioidaan, mitä odottamattomia vaikutuksia politiikalla, ohjelmalla tai uudistuksella voi olla. Arvioinnin läpinäkyvyydellä tarkoitetaan, että arvioinnin kohde on tietoinen arvioinnista ja arviointitiedon käytöstä (esim. O'Neil 2002; Saunders 2006; Stern 2005). Kaiken kattavan läpinäkyvyyden vaatimus on O'Neilin (2002)

mukaan kuitenkin lisännyt kontrollia ja näin heikentänyt luottamusta ja jopa edistänyt epäluuloisuuden ilmapiiriä. Avoimuus ja läpinäkyvyys, joka perustuu teknisesti nykyisin helppoon tiedon tuottamiseen, on samalla uhka luottamukselle.

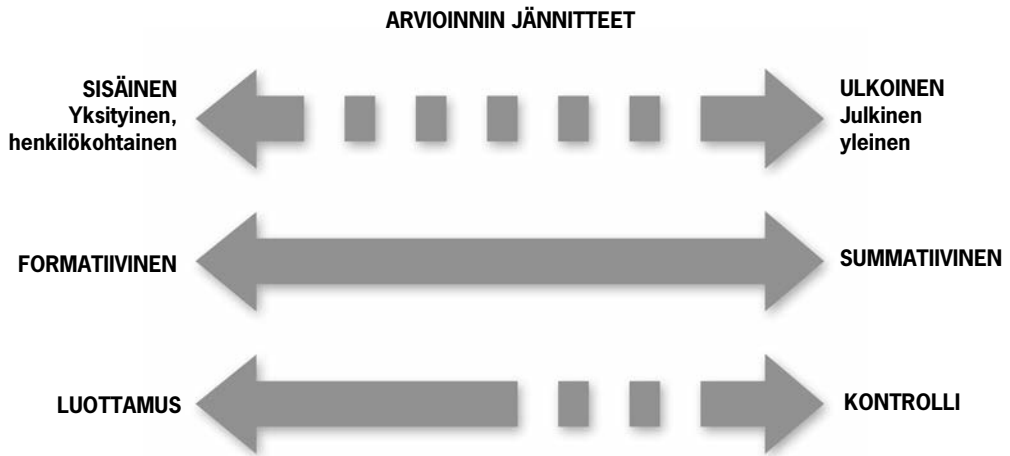
Kehittävässä arvioinnissa arvioijan rooli on aktiivinen; arvioijan on varmistettava sekä arvioinnin kohteen osallisuus että arvioinnin toteuttamisessa ja arviointitulosten oikeudenmukainen käyttö. Tietoisuutta arvioinnin tarkoituksesta, arviointikysymyksistä ja arviointitiedon käytöstä pidetään osallistavan arvioinnin perusehtona. Palaute on olennainen osa kehittävää ja luottamusta edistävää arviointia, ja ihmisillä tulee olla mahdollisuus itse tarkistaa ja arvioida tiedon luotettavuus. Luottamusta edistävässä arviointikulttuurissa arvioinnin kohde, esimerkiksi opettajat ja opiskelijat, ovat oman toimintansa aktiivisia reflektioijia, he ovat osallisina arviointiprosessissa ja kokevat ammatillista vastuuta ja vaikutusvaltaa siinä. Luottamusta edistävässä arviointikulttuurissa painotetaan organisaatioiden, esimerkiksi oppilaitosten, omaa laadukasta hallinnointia eikä niinkään kontrollia ulkoapäin (Crooks 2002; Fetterman 2001; O'Neil 2002; Räisänen 2005; Saunders 2006)

Arvioinnin ulottuvuuksien jännitteet

Kehittämistyön esiin nostamia jännitteitä, arvioinnin luotettavuuden ongelmakohtia, ristiriitoja, on mahdollista ymmärtää Kuviossa 1 esitetyn käsitteellisen mallin avulla. Arvioinnin teoreettisen tarkastelun ja käsitteellisen analyysin perusteella olen nimennyt arvioinnin ulottuvuuksiksi eli dimensioiksi:

formatiivinen – summatiivinen, sisäinen – ulkoinen sekä luottamus – kontrolli (Räkköläinen 2011). Nämä ulottuvuudet ovat toisiinsa nähden jännitteissä suhteessa. Jännitteet näillä ulottu-

vuuksilla nousevat esiin sovittaessa yhteen opiskelijan arvioinnin ja kansallisen arvioinnin tavoitteita näyttöperusteisessa arviointijärjestelmässä.



Kuvio 1. Arvioinnin ulottuvuuksien jännitteet (Räkköläinen 2011).

Arviointi voi olla **summatiivista**, kokoavaa arviointia ja **formatiivista**, kehittävä arviointia (esim. Ecclestone 1996, 2005; Cronbach; 1982; Scriven 1967). Jaottelu on perinteinen, mutta käyttökelpoinen ja sen avulla voi tarkastella opiskelijan arvioinnin ja kansallisen oppimistulosten arvioinnin erilaisia tarkoituksia. Summatiivinen liitetään usein vaikutusten ja tulosten arviointiin ja formatiivinen prosessiarviointiin. Opiskelijan arviointi on usein kehittävä oppimista edistävä ja oppimistulosten arviointi kokoavaa tulosten ja vaikutusten arviointia.

Jaottelu **ulkoiseen** ja **sisäiseen** arviointiin puolestaan tuo tarkasteltavaksi muun muassa sen, miten arviointikysymykset asetetaan; nousevatko ne toiminnan sisältä vai asetetaanko ne ulkoapäin.

Ulkoisen ja sisäisen arvioinnin toteuttajat ovat erilaisessa suhteessa arvioinnin kohteeseen; sisäinen arviointi suoritetaan omin voimin yhteisön sisällä, kun taas ulkoisen arvioinnin tekijä ei ole osa arvioinnin kohdetta eikä osallistu sen toimintaan. Sisäinen arviointi on usein itsearviointia ja prosessien arviointia.

Luottamus ja sen vastapoolina **kontrolli** näyttäytyvät eri tavoin tilivelvollisuusevaluaatiossa kuin osallistavassa arvioinnissa sekä eri tavoin kokoavassa järjestelmätason arvioinnissa kuin opiskelijan arvioinnissa. Kontrolli liitetään julkiseen valvontaan, joka usein on tilivelvollisuusevaluaatiota. Arvioinnin seuraamukset tai vaikutukset kohderyhmään vaikuttavat siihen, miten kontrolli ja luottamus ilmenevät. Mitä enemmän arvioinnin tulokset vaikuttavat ar-

vioitavan tulevaisuuteen (ns. high-stake evaluation), sitä epäoikeudenmukaisemmaksi ulkoinen kontrolli voidaan kokea (esim. rankinglistat), ja tämä puolestaan voi heikentää luottamusta. Luottamus heikkenee myös, jos arvioinnilla ei ole minkäänlaisia seuraamuksia eikä sen tuloksia voi hyödyntää eikä käyttää kehittämisessä. Arvioitavalle kohteelle asetetut standardit ja kriteerit lisäävät kontrollin ulottuvuutta arvioinneissa. Luottamukseen liittyy kontekstin huomioiminen sekä prosessiarviointi ja osallistumista painottavat menetelmät (ks. esim. Crooks 2002, 2003; Gipps 2003; Kuusela 2003; Linn 2000; O'Neil 2002; Walberg & Haerthel 1990).

Näyttöperusteisen arvioinnin kontekstuaalisuus ja eri tehtävät

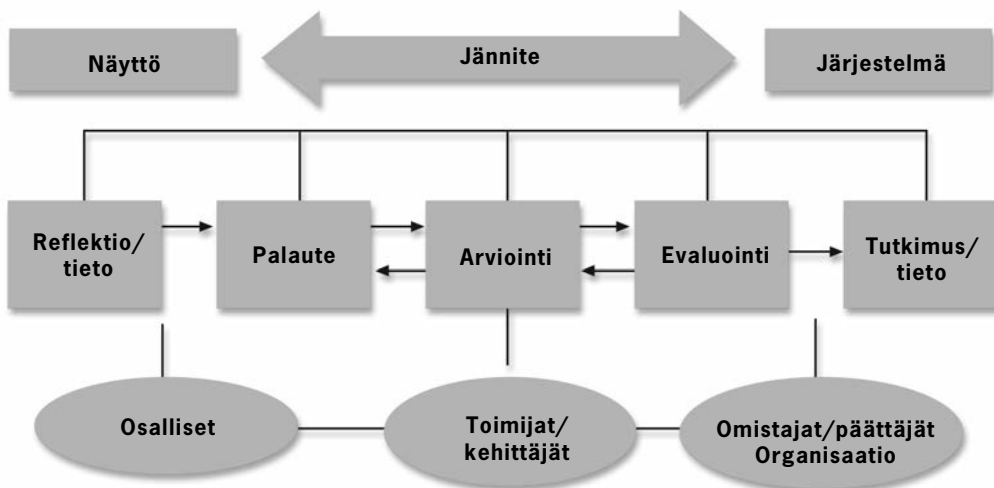
Näyttöperusteinen osaamisen arviointi (ammattiosaamisen näytöt) muuttaa aiempaa opiskelijan arviointia kontekstiperusteiseksi ja toiminnalliseksi. Näyttöperusteisessa arvioinnissa arvioidaan toteennäytettyä osaamista (demonstrated competence). Näyttö tarkoittaa suoritusta tai esitystä (performance-based) ”oikeassa elämässä”. Kontekstiperusteisessa arvioinnissa kiinnitetään huomio toimintaprosessiin, toimijoihin ja prosessin ”omistajiin” sekä tietojen ja taitojen soveltamiseen. Näyttöperusteinen osaamisen arviointi määritetään myös autenttiseksi arvioinniksi tai suoritusarvioinniksi. Tällöin ajatellaan, että suoritus itsessään on osoitus oppimisesta. Suoritus on myös mahdollista arvioida yhteisarviointina, jolloin arviointitieto syntyy neuvottelun tuloksena. Samoin kuin luottamusta edistävässä tilivelvollisuudessakin, myös opiskelijan arvioinnissa luottamus edellyttää palautetta, reflektion mahdollisuutta ja opiskelijan osallisuutta arvioinnissa (ks. esim.

Ecclestone 1996, 1999, 2005; Virta 1999; Wolff 1996).

Malli näyttöperusteisesta arviointiprosessista auttaa tarkastelemaan näyttöperusteista arviointia eri konteksteissaan ja jännitteiden ilmenemistä niiden välillä (ks. kuvio 2). Jäsennyksen pohjalta on käsitys arvioinnin monitasoisuudesta sekä teoreettinen näkemys arviointitiedon, reflektion ja oppimisen kontekstuaalisuudesta ja monimutkaisten prosessien luonteesta (ks. Fliedner 2001; Karjalainen 2007; Van de Ven & Poole 1995).

Eri kontekstit on nimetty arviointiprosessissa palautteen, arvioinnin ja evaluoinnin käsitteillä. Reflektio liittyy oppimiseen, olemassa olevien asioiden opetteluun ja uuden tiedon tuottamiseen näyttöperusteisessa arviointiprosessissa. Reflektion tärkein väline on palaute. Yksilötasolla muodostunut palautetieto opiskelijan osaamisesta muovautuu näytön yhteisessä arviointikeskustelussa (opettaja – työpaikkaohjaaja – opiskelija) arviointitiedoksi, jota kokoomalla organisaatioiden tasolla (koulutuksen järjestäjät, oppilaitokset, opetushallinto) tuotetaan yhteiskunnallista evaluointitietoa koulutuksen tuloksista (arviointijärjestelmä) (Poikela & Rökköläinen 2006; Rökköläinen 2011).

Näyttöperusteisessa arviointiprosessissa näytöt ovat osana opiskelijan arviointia (assessment) ja samalla perustana kansalliselle oppimistulosten arvioinnille, joka on julkista tuloksellisuuden arviointia (evaluation). Sekä opiskelijan arviointi kahdenkeskisessä vuorovaikutuksessa että näytön yhteisarviointi ja arviointikeskustelu ovat sisäistä arviointia ja usein formatiivista, joskus diagnostista, jolloin ”yleisönä”



Kuvio 2. Näyttöperusteinen arviointiprosessi (Räkköläinen 2011).

(audience) ovat opiskelijat, opettajat, työelämän edustajat, oppilaitokset ja työpaikat (yhteisöt). Näyttöperusteisessa arviointiprosessissa nämä tahot ovat osallisia ja toimijoita, jotka itse asiassa ovat sekä arviointitiedon tuottajia että käyttäjiä (Räkköläinen 2011).

Silloin kun arviointitietoa tai palautetta ryhdytään dokumentoimaan yhdenmukaisella tavalla ja antamaan suorituksesta todistuksia (esimerkiksi näyttötodistus), siirrytään arvioinnin ulottuvuuden dimensioilla ulkoisen arvioinnin suuntaan. Tällaisen julkisen arviointitiedon intressitahoja ovat hallinto, toimeksiantajat, veronmaksajat, ”suuri yleisö”. Näyttöperusteisen arviointiprosessin kontekstissa ne ovat omistajia ja päättäjiä. Tuotettava tieto on evaluointia, saavutuksista ja tuloksista kertovaa summatiivista arviointitietoa, jota tosin voidaan käyttää myös formatiivisessa tarkoituksessa (kehittävä evaluointi, ks. Stern 2005).

Arvioinnin laatuvaatimukset ja luotettavuuden kriteerit

Luotettavuuden määrittämisestä

Arvioinnin perinteisinä laatuvaatimuksina pidetään arviointien **toistettavuutta** eli reliabiliteettia sekä arviointien **osuvuutta** eli validiteettia (esim. Patton 2002; Rossi, Lipsey & Freeman 2004). Arvioinnilta edellytetään sekä toistettavuutta, pysyvyyttä ja ei-sattumanvaraisuutta (reliabiliteetti) että osuvuutta ja pätevyttä (validiteetti). Arviointi on osuvaa, jos on tuotettu tietoa siitä, mistä on ollut tarkoituskin. Arvioinnin on oltava vertailukelpoista ja kohdealueen kattavaa sekä tuotetun tiedon käyttökelpoista eri tarkoituksissa. Validiteetti- ja reliabiliteettivaatimukset liitetään tavallisesti määrällisiin ja mitaamiseen perustuviin arviointeihin, mutta termit ovat käytössä myös laadullisissa arvioinneissa. Arvioinnin validiteetin ja reliabiliteetin vaatimusten välillä on teoreettinen jännite; ne eivät voi esiintyä täysimääräisesti yhtä aikaa.

Myös aivan uusia käsitteitä luotettavuuden määrittämiseksi on luotu. Gordon Stobart (2008) on tuonut arviointien luotettavuuskeskusteluun mielenkiintoisen uuden käsitteen ”dependability” (luotettavuus). ”Depend on” merkitsee johonkin luottamista, mutta ”depend” (it depends...) voi tarkoittaa myös jonkin asian riippumista, johtumista jostakin, esimerkiksi olosuhteista. Tämä käsite yhdistää validiteetin ja reliabiliteetin toisiinsa. Käsitteen avulla voi teoreettisesti yrittää löytää tasapainoa reliabiliteetti- ja validiteettivaatimusten välillä. Arvioinnin luotettavuuden varmistamisessa tulisi huomioida arvioinnin eri kontekstit ja eri tarkoitukset. Joissakin konteksteissa validiteetti eli arvioinnin osuvuus on tärkeämpää kuin reliabiliteetti. Toisissa konteksteissa taas reliabiliteetin eli vertailtavuuden ja yleistettävyyden korostaminen on tarkoituksenmukaisempaa kuin korkean validiteetin säilyttäminen.

Usein arviointiaselmissa painotetaan reliabiliteettia validiteetin kustannuksella, vaikka tosinaan olisi otettava reliabiliteettiriskejä, jotta arvioinnilla todella voitaisiin vaikuttaa oppimiseen ja opetukseen. Tällöin arvioinnin autenttisuus on luotettavuuden kannalta merkittävintä. Arviointiaselmat voidaan muun muassa suunnitella paikallisiin olosuhteisiin sopiviksi ja varmistaa muulla tavoin mahdollisimman korkea luotettavuus, ”dependability”, arviointijärjestelmässä. Se voidaan saavuttaa painottamalla ensisijaisesti validiteettia usein valittujen halpojen, tehokkaiden ja ilman reliabiliteettiriskeä olevien arviointimenetelmien sijaan. Reliabiliteetissa tulosten toistettavuus nimittäin perustuu siihen, että arviointi on standardisoitu soveltuvaksi samanlaisena erilaisiin tilanteisiin, mikä voi kaventaa ar-

vioinnin kohdetta ja heikentää validiteettia. Stobardin kanta on, ettei ole mahdollista saavuttaa täysin oikeudenmukaista arviointia, mutta arvioinnista voi aina yrittää tehdä oikeudenmukaisempaa.

Uusimmissa (viidennen sukupolven) kontekstuaalisuutta ja osallistumista painottavissa arvioinneissa korostuu arvioinnin hyödyllisyyden ja käyttökelpoisuuden vaatimukset sekä arvioinnin eettiset kysymykset, jolloin tarkastellaan kriittisesti arvioinnin motiiveja ja tavoitteita. On ryhdytty korostamaan, että arvioijien on kannettava vastuuta myös arviointien hyödyllisyydestä (ks. Patton 1997, utilization-focused evaluation). Osallistavassa ja kehittävässä arvioinnissa korostuu erityisesti arvioinnin merkitys arvioinnin kohteelle, arvioitavalle asialle (esimerkiksi oppimiselle) sekä arvioinnin seuraamukset. Yleistäminen tai kausaalisuhteiden etsiminen ei silloin ole keskeisin tavoite, vaan tietyn tapauksen ymmärtäminen tietyssä kontekstissa mahdollisimman totuudenmukaisesti (thrustworthiness). Reliabiliteetti ja arvioinnin uskottavuus (credibility) varmistetaan tällöin suunnittelemalla ja toteuttamalla systemaattinen arviointiprosessi. Arviointiprosessia ei voi kuitenkaan kokonaan kontrolloida, vaan arvioijan on siedettävä epävarmuutta (esim. Guba & Lincoln 1989; Patton 2002; Poikela 2004, 2005a, 2005b).

Tutkimuksen tuloksia

Näytöillä myönteisiä vaikutuksia

Tutkimukseni mukaan näytöt lisäävät osallisuutta arvioinnissa, ne edistävät oppimista ja johtavat kehittämistyöhön. Tutkimus osoittaa, että erityisesti työpaikoilla toteutetut

Työpaikoilla toteutetut näytöt antavat osuvaa arviointitietoa.

näytöt antavat osuvaa arviointitietoa opiskelijan osaamisesta ja arvosanat koetaan oikeudenmukaisiksi. Työelämän kanssa yhteistyössä suunnitellut näytöt kehittävät opetusta ja sen vastaavuutta työelämän vaatimuksiin. Näyttöjä käytetään hyvin kehittävästi; opiskelijoita ohjataan ja valmennetaan ja heidän osaamisensa varmistetaan ennen näyttöä, mikä edistää oppimista. Uusi näyttöihin perustuva oppimistulosten kansallinen arviointi on koettu hyödylliseksi oppilaitoksissa ja arviointiraportteja, etenkin oppilaitoskohtaisia palautteita, on käytetty oman toiminnan kehittämisessä.

Keskeisin haaste näyttöjärjestelmän kehittämisessä on, miten samaa arvioinnin välinettä, näyttöä, käytetään eri tarkoituksissa ja onko laadunvarmennuksessa pääpaino paikallisella vai valtakunnallisella tasolla. Lopullisista näyttöä koskevista säädöksistä tuli tasapainoilua jännitteiden välillä, mikä ilmenee kaikissa näyttöperusteisen arvioinnin konteksteissa erityisesti kontrollin ja

luottamuksen välillä. Kansallisen arvioinnin tehtävän ja tarkoituksen määrittäminen jäi epäselväksi.

Opiskelijan arviointia koskevat säädökset ja ohjeet ovat lisääntyneet ammattiosaamisen näyttöjen myötä, näyttöjen arviointiin luotiin kriteerit, arvioinnista ja arvioinnin dokumentoinnista annettiin uusia määräyksiä ja opetussuunnitelman perusteet (sittemmin tutkinnon perusteet) vahvistivat järjestelmätason ohjausta opiskelijan arvioinnissa. Näyttöjen arviointi vaikuttaa merkittävästi opiskelijaan: näytöt ovat pakollisia, jokaisen opintokokonaisuuden keskeinen osaaminen arvioidaan ja näytöistä annetaan erillinen näyttötodistus. Kaiken kaikkiaan näyttöjärjestelmästä tuli hyvin kattava; se koskee kaikkia tutkintoja ja kaikkia koulutuksen järjestäjiä ja kaikkien on otettava näytöt käyttöön, niiden toteuttaminen ei ole vapaaehtoista.

Tutkimukseni mukaan opettajat kokivat säädösten lisääntymisen ja työelämän edustajan oikeuden päättää yksin arvioinnista epäluottamuksena ja kontrollina, joka kaventaa heidän ammatillista vastuutaan. Työelämän edustajat puolestaan kokivat epävarmuutta arvioinnissa, eivätkä luottaneet omiin arviointitaitoihinsa. He kokivat vastuunsa näytöistä suureksi ja toivoivat arvioinnissa kiinteämpää yhteistyötä opettajien kanssa. Kaikissa tutkinnoissa työelämän edustaja arvioi näytön erittäin harvoin yksin.

Näyttöihin luotetaan, mutta niiden vertailukelpoisuutta epäillään

Näyttöjen toteuttamistavat, suorituspaikat ja arviointikäytännöt vaihtelevat paljon tutkinnoittain ja oppilaitoksittain.

Arviointikriteereitä sovelletaan ja ymmärretään arvioijasta riippuen hyvin eri tavoin. Kriteeriperusteinen arviointi koetaan pääsääntöisesti vaikeaksi. Säädösten lisääntymisestä huolimatta näyttöjen laatuvaatimukset ovat jääneet pääasiassa suosituksiksi ja oppilaitoksille on paljon joustoa näyttöjen käytännön toteuttamisessa. Näytöissä annetaan erittäin hyviä arvosanoja. Vaikka näyttöihin luotetaan, arvosanojen vertailukelpoisuutta epäillään.

Näytöt eivät tuota riittävän vertailukelpoista tietoa oppimistulosten seuranta varten. Kansallisen arvioinnin tulosten luotettavuuden kannalta sääntely ja näyttöjen laadunvarmistus on liian vähäistä. Luottamusta puolestaan heikentää ennen kaikkea epäselvyys siitä, mihin näytöistä saatavaa arviointitietoa ja oppimistuloksia käytetään. Kansallisen arvioinnin seuraamuksista ei käydä keskustelua eikä arvioinnin kohteella, opettajilla, oppilaitosjohdolla eikä opiskelijoilla, ole niistä tietoa.

Vertailtavuusongelmasta huolimatta näyttöihin arviointimenetelmänä luotetaan, koska näytöillä on paljon myönteisiä seuraamuksia opiskelijan oppimiseen ja koulutuksen kehittämiseen. Näyttöjä pidetään osuvana arvioinnin menetelmänä ja arvosanoja pääsääntöisesti oikeudenmukaisina. Toisaalta näyttöjen toteuttaminen erilaisissa olosuhteissa, erilaisin tehtävin ja yksilöllisten suunnitelmien mukaan koetaan riskiksi arvioinnin vertailtavuudelle.

Luotettavuuden ja luottamuksen välillä yhteys

Tutkimus osoittaa, että samat asiat, jotka varmistavat näyttöperusteisen arvioinnin laatua ja siten oppimistulosten

luotettavuutta, lisäävät myös luottamusta näyttöperusteiseen arviointiin. *Yhteisarviointi* herättää voimakkaimmin luottamusta, samoin opettajien ja työelämän edustajien yhteistyö näyttöjen suunnittelussa, näyttöjen toteuttaminen työpaikoilla ja kriteeriperusteinen arviointi. Luotettavuutta voidaan lisätä kouluttamalla arvioijat arviointitehtävään entistä paremmin, selkiyttämällä arviointikriteereitä ja vahvistamalla arvioijien yhteistä käsitystä kriteereistä.

Luottamusta edistää myös vertaisarviointi näyttötoiminnan laadunvarmistuksen menetelmänä. Yhteisarvioinnissa ja vertaisarvioinnissa arviointivastuu jaetaan. Tämä edistää *osallisuutta* arvioinnin toteuttamisessa ja lisää vaikutusmahdollisuuksia omassa asiassa, mikä on luottamuksen perusta. Vertaisarviointiin suhtaudutaan hyvin myönteisesti; opettajat luottavat vertaisilta saatuun palautteeseen. Vertailutieto muiden oppilaitosten tuloksista ja taustatieto tulosten tulkinnan tueksi edistää luottamusta. Luottamuksen tärkein edellytys on kuitenkin, että kansallisen arvioinnin tarkoituksesta ja seuraamuksista sovitaan selkeästi ja avoimesti. Arvioinnin tarkoituksen selkiyttäminen mahdollistaa myös arvioinnin luotettavuuden parantamisen.

Luotettavuuden varmistaminen näyttöperusteisessa arviointiprosessissa

Opiskelijan arvioinnin luotettavuutta ja vertailukelpoisuutta on näytöissä pyritty lisäämään siirtymällä normiperusteisesta arvioinnista (norm-referenced) kriteeriperusteiseen arviointiin (criterion-referenced). Kansallisen arvioinnin kannalta ammattiosaamisen näyttöjen luotettavuusongelmat liittyvät nimenomaan

arvioinnin reliabiliteettiin, koska näytöt suunniteltiin lisäämään ennen kaikkea arvioinnin osuvuutta ja relevanssia, eikä kehittämistyössä tai säädöksissä niinkään korostettu niiden yhdenmukaisuutta ja toistettavuutta. Näytöille asetetut tavoitteet osana opiskelijan arviointia ja oppilaitosten oman toiminnan kehittämistä saavutetaan paremmin kuin järjestelmätason tavoitteet.

*Yhteisarviointi
herättää
voimakkaimmin
luottamusta.*

Näyttöperusteisessa arvioinnissa on etsitty tasapainoa arvioinnin laatuvaatimusten välille. Arvioinnin autenttisuus on näyttöperusteisen arvioinnin luotettavuuden kannalta merkittävää ja se voidaan useimmiten saavuttaa painottamalla validiteettia. Ratkaisuna on ollut myös reliabiliteettiriskien ottaminen ja validiteetin painottaminen luotettavuutta koskevilla linjauksissa. Näytöissä arvioidaan laaja-alaista ammatillista osaamista, vaikka se hankaloittaa tulosten vertailemista kansallisessa oppimistulosten arvioinnissa. Validiteetin pai-

nottaminen mahdollistaa laaja-alaiset ammattiosaamisen näytöt. Reliabiliteetin painottaminen olisi kaventanut ammatillisesta osaamisesta saatavaa arviointitietoa näytöissä.

Näytöt kehittävät opiskelijan arviointia lisäämällä sen osuvuutta ja autenttisuutta. Ne parantavat näin oppimistulosten arvioinnin validiteettia myös kansallisella tasolla. Stobart (2008) korostaa, että arviointi tulisi pitää niin autenttisenä kuin vain mahdollista, jolloin se on kaikkien hyödyllisintä opetuksen ja oppimisen kannalta. Myös oman tutkimukseni tulokset vastaavat tätä näkemystä; näytöt ovat autenttista arviointia ja ne ovat hyödyllisiä oppimisen ja opettamisen kannalta. Johtopäätöksenä tästä voisi todeta, että mikäli luottamusta pidetään tavoiteltavana asiana, tulisi näytöt pitää mahdollisimman autenttisina jopa kansallisen vertailtavuuden kustannuksella.

Tutkimukseni perusteella näyttöperusteisen arvioinnin validiteetti on merkityksellisempää toimijoiden (yksilöt, yhteisö) kannalta ja paikallisella tasolla, reliabiliteetin tärkeys puolestaan on suurempaa kansallisen järjestelmän (poliittinen päätöksenteko, hallinto) kannalta. Ero johtuu siitä, että arvioinnin funktiot ovat erilaisia ja näytöt ovat arviointivälineinä eri tarkoituksessa. Stobart (2008) korostaa, että asetettaessa standardeja arvioinnille tulee aina samalla pohtia, tuotetaanko näin sellaista tietoa, jota tarvitaan (fitness -for -purpose) ja mitkä ovat arvioinnin seuraamukset opetuksessa ja oppimisessa. Liiallisen yhdenmukaistamisen (reliabiliteetin painottaminen) vaarana on näyttöjen ja opetuksen kaventuminen; yhdenmukaisuuden korostaminen voi kaventaa näyttöjä yksittäisiksi tehtäviksi ja testeiksi.

Tällöin ne eivät enää ole niin osuvia ammatillisen osaamiseen ”mittareita” jollaisiksi ne tutkimuksessa osoittautuivat. Vertailtavuuden korostaminen merkitsi muun muassa, että joudutaan tinkimään laaja-alaisen ammatillisen osaamisen arvioinnista.

Yhteisarviointi, kriteeriperusteinen arviointi ja näyttöjen suunnittelu oppilaitosten ja työpaikkojen yhteistyönä lisäävät näyttöjen autenttisuutta ja siten validiteettia. Reliabiliteettia ja validiteettia ei tulisi erottaa toisistaan liian jyrkästi (Stobart 2008). Vertaisarviointi on yksi mahdollisuus yhdistää arvioinnin luotettavuuskriteereitä ja löytää tasapaino reliabiliteetti- ja validiteettivaatimusten välille (dependability). Ulkoisen ja sisäisen arvioinnin yhdistäminen on toisinaan parempi ratkaisu kuin niiden erillään pitäminen (Patton 1997). Laadunvarmennusmenetelmässä käytetyssä vertaisarvioinnissa sisäinen ja ulkoinen arviointi yhdistyvät.

Kohti uutta arviointiajattelua

Vaikka näyttöperusteisen arviointijärjestelmän kehittäminen on vasta aluillaan, voidaan siinä nähdä useita piirteitä kehittävästä ja osallistavasta arvioinnista. Etenkin opiskelijan arvioinnin kontekstissa kehittävä tavoite tuntuu toteutuvan, mutta kansallisen arvioinnin tasolla olisi määritettävä, millainen osallistumisen aste edistää luottamusta. Arvioinnin vieminen lähemmäksi toimijoita ja varsinaista arvioinnin kohdetta on lähempänä neljännen sukupolven konstruktivistista arviointinäkemystä kuin aiemmat erilliset koejärjestelyt.

Näyttöperusteisen kansallisen arviointijärjestelmän kehittyminen riippuu paljon siitä, miten muotoutuu arvioijan ja arvioinnin kohteen välinen vuorovaikutus. Mitä useammassa arviointiprosessin vaiheessa arvioinnin kohde on mukana, sitä osallistavampaa arviointi on. Osallistumista pidetään moderneissa arviointinäkemyksissä arvioinnin yhtenä keskeisimpänä tavoitteena ja laatuvaatimuksena. Mikäli halutaan painottaa kehittämistä ja osallistumista, tulisi osallisten ja toimijoiden (koulutuksen järjestäjien, opettajien, työelämän edustajien ja opiskelijoiden) ääntä kuulla myös määritettäessä näyttöperusteisen kansallisen arvioinnin arviointikysymyksiä ja suunniteltaessa arviointiasetelmia. Realistisen arvioinnin suuntaan mentäessä tulisi yhä tietoisemmin tunnistaa kontekstiperusteinen arviointi ja tiedon käyttökelpoisuus eri konteksteissa.

Tutkimuksen perusteella voi todeta, että kansallisen arvioinnin ja ammatti-osaamisen näyttöjen integrointi edellyttää

- arvioinnin tarkoituksen kirkastamista
- linjausta siitä, mikä on hyväksyttävä taso sääntelyssä
- laadunvarmennuksen menetelmän valintaa
- arvioinnin hyödyllisyyden ja käyttökelpoisuuden varmistavaa palautejärjestelmän kehittämistä.

Tutkimus osoitti myös, että luotettavuuden ja luottamuksen välillä on yhteys; arvioinnin tulee täyttää tietyt vähimmäistason luotettavuuskriteerit, jotta tuloksiin voi luottaa. Kontrollin ja sääntelyn keskeneräisyys, epämääräisyys tai liiallinen joustavuus heikentävät

luottamusta. Laadunvarmennus edellyttäsikin näin rohkeutta valita sitovien normien ja suositusten välillä, jolloin määritettäisiin vähintään arvioinnin minimilaatuvaatimukset. Tällöin on kysymys myös linjanvedosta sääntelyn ja paikallisen joustavuuden välillä.

Tutkimukseni mukaan samat asiat, jotka varmistivat arvioinnin laatua, edistivät myös luottamuksen syntyä. Kaiken kaikkiaan näyttöperusteinen arviointi koetaan hyödylliseksi ja motivoivaksi, mikä on perusta luottamukselle.

Lähteet

Asetus ammatillisesta koulutuksesta annetun asetuksen muuttamisesta 2005. 603/21.7.2005.

Cronbach, L. J. 1982. Designing Evaluations of Educational and Social programs. San Francisco: Jossey Bass.

Crooks, T.J. 2002. Educational Assessment in New Zealand Schools. Profiles of Educational Assessment Systems World-Wide. Assessment in Education 9 (2), 237–253.

Ecclestone, K. 1996. How to Assess the Vocational Curriculum. London: Kogan Page.

Ecclestone, K. 1999. Empowering or Ensnaring? The Implications of Outcome-based Assessment in higher Education. Higher Education Quarterly 53 (1), 29–48.

Ecclestone, K. 2005. Understanding Assessment and Qualifications in Post-Compulsory Education and Training. Principles, Politics and Practice. Leicester: NIACE (National Institute of Adult Continuing Education).

Fetterman, D. 1994. Empowerment Evaluation. Evaluation Practice. 15 (1), 1–15.

Fetterman, D. 1996. Empowerment Evaluation: An Introduction to Theory and Practice. In D. Fetterman, S. Kaftarian & A. Wandersman (Eds.) Empowerment Evaluation. Knowledge and Tools for Self-Assessment and Accountability. London: Sage Publications, 1–46.

Fetterman, D. 1998. Empowerment Evaluation and Accreditation in Higher Education. In E. Chelmsky and W. Shadish (Eds.) Evaluation for the 21st century: A Handbook. Thousand Oaks, CA: Sage, 381–395.

Fetterman, D. 2001. Foundations of Empowerment Evaluation. London: Sage Publications.

Fliedner, D. 2001. Six Levels of Complexity; A Typology of Process and Systems. Journal of

Artificial Societies and Social Simulation 4 (1). Luettu 28.11.2010 osoitteesta <http://jass.soc.surrey.ac.uk/4/1/4.html>.

Guba, G.E. & Lincoln, Y.S. 1989. Fourth Generation Evaluation. Newbury Park, CA: Sage.

House, E.R. & Howe, K.R. 1999. Values in Evaluation and Social Research. Thousand Oaks, CA: Sage.

Jakku-Sihvonen, R. & Heinonen, S. (toim.) 2001. Johdatus koulutuksen uudistuvaan arviointikulttuuriin. Arviointi 2/2001. Helsinki: Opetushallitus.

Karjalainen, A. 2007. Koulutusorganisaation prosessit. Ammattikorkeakoulujen verkkolehti KeVer 2/2007. Luettu 28.11.2010 osoitteesta <http://www.osaaja.fi./index.php/kever/article/viewArticle/1016/8>.

Kuusela, J. 2003. Koulujen paremmuusjärjestyksestä. Opetushallitus. Moniste 7/2993. Helsinki: Edita Prima Oy Ab.

Laki ammatillisesta koulutuksesta annetun lain muuttamisesta 2005. 601/15.7.2005.

Linn, R.L. 2000. Assessments and Accountability. Educational Researcher 29 (2), 4–16.

O’Neil O. 2002. A Question of Trust. The BBC Reith Lectures 2002. Cambridge: Cambridge University Press.

Opetushallituksen määräys 2005. Ammatillisen koulutuksen opetus-suunnitelman ja näyttötutkinnon perusteet. M 32/011/2005/30.9.2005.

Opetusministeriö. 2004. Koulutus ja tutkimus 2003–2008. Kehittämisen suunnitelma. Opetusministeriön julkaisuja 2004:6. Helsinki: Yliopistopaino Oy.

Opetusministeriön päätös (212/430/98) 23.6.1998 ammatillisten perustutkintojen kehittämisestä.

Opetusministeriön päätös (3/502/2007) 12.2.2007. Ammattiosaamisen näyttöjen kehittäminen: oppimistulostiedon tuottaminen.

Opetusministeriön työryhmien muistioita (14:1999) 3.6.1999. Ammatillisen peruskoulutuksen näyttöjen toteuttaminen.

Patton, M.Q. 1997. Utilization – Focused Evaluation: The new Century Text. 3.painos. Thousand Oaks, CA: Sage Publications.

Patton, M.Q. 2002. Qualitative Research and Evaluation Methods. 3.painos. London: Sage Publications.

Pawson, R. & Tilley, T. 1997. Realistic Evaluation. London: Sage Publication.

Poikela, E. 2004. Developing Criteria for Knowing and Learning at Work: Towards Context-Based Assessment. Journal of Workplace Learning 16 (5), 267–274.

Poikela, E. 2005a. Työssä oppimisen prosessimalli. Teoksessa E. Poikela (toim.)

Osaaminen ja kokemus – työ, oppiminen ja kasvatus. Tampere: Tampere University Press, 21–41.

Poikela, E. 2005b. Luottamusta luova kehityskeskustelu. Teoksessa E. Poikela (toim.) *Oppiminen ja sosiaalinen pääoma*. Tampere: Tampereen yliopisto-paino oy-Juvenes Print, 31–52.

Poikela, E. & Rökköläinen, M. 2006. ”Intelligent Accountability” – konteksti-perusteisen arvioinnin lähtökohtia. *Ammattikasvatuksen aikakaus-kirja 8 (2)*, 6–18.

Rossiey, P.H., Lipsey, M.W. & Freeman H.E. 2004. *Evaluation. A Systematic approach*. 7. painos. Thousand Oaks: Sage Publications.

Räisänen, A. 2005. Kehittävään arviointiin. Teoksessa K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnissa. Koulutuksen arviointi-neuvoston julkaisuja 6*. Jyväskylä: Jyväskylän yliopistopaino, 109–127.

Rökköläinen, M. 2005a. Kansallisen näyttöperusteisen oppimistulosten arviointijärjestelmän kehittäminen ammatillisiin perustutkintoihin. Arviointi-kokeilusta kohti käytäntöä. Arviointi 3/2005. Opetushallitus. Helsinki: Hakapaino Oy.

Rökköläinen, M. 2005b. Kontrollista luottamukseen – tilivelvollisuus arvioinnissa. Teoksessa E. Poikela (toim.) *Oppiminen ja sosiaalinen pääoma*. Tampere: Tampereen Yliopistopaino Oy – Juvenes Print, 71–100.

Rökköläinen, M. & Ecclestone, K. 2005. The Implications of Using Skills Tests as Basis for a National Evaluation System in Finland. Outcomes from a Pilot Evaluation in 2002–2003 in Finland. *Evaluation 1/2005*. Finnish National Board of Education. Helsinki: Hakapaino Oy.

Rökköläinen, M. 2011. Mitä näytöt näyttävät? Luotettavuus ja luottamus ammatillisten perustutkintojen näyttöperusteisessa arviointiprosessissa. *Acta Universitatis Tamperensis 1636*. Tampere: Tampere University Press.

Salmio, K. 2004. Esimerkkejä peruskoulun valtakunnallisista arviointihankkeista kestävä kehityksen didaktiikan näkökulmasta. Vuosien 1993–1995 valtakunnalliset kokeet ja vuoden 1998 luonnontieteiden oppimistulosten arviointi. Joensuun yliopisto. *Kasvatustieteellisiä julkaisuja 98*. Joensuu: Joensuun yliopistopaino.

Saunders, M. 2006a. Do we Hear the Voices? The “Presence“ of Evaluation Theory and Practice in Social Development. *Evaluation. The International Journal of Theory, Research and Practice 12 (2)*, 251–264.

Scriven, M. 1967. The Methodology of Evaluation. In R.W. Tyler (Ed.) *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally, 39–83.

Stern, E. (Ed.) 2005. *Evaluation research methods. Volume I–IV. Sage Benchmarks in Social Research Methods*. London: Sage Publications.

Stobart, G. 2008. *Testing Times. The uses and abuses of assessment*. Lontoo: Routledge.

Tang, C. & TiWari, A. 2003. From Process to Outcome: the Effect of Portfolio Assessment on Student Learning. *Nurse Education Today 16 (4)*, 269–277.

Walberg, H. & Haertel, G. 1990. *the International Encyclopedia of Educational Evaluation*. Oxford: Pergamon Press.

Van de Ven, A. & Poole, M. 1995. Explaining Development and Change in Organizations. *Academy of Management Review 20 (3)*, 510–540.

Virta, A. 1999. Uudistuva oppimisen arviointi. Mahdollisuuksia ja varauksia. Turun yliopiston kasvatustieteiden tiedekunta. *Julkaisusarja B:65*. Turun opettajakoulutuslaitos. Turku: Painosalama Oy.

Wolff, A. 1996. *Competence-based Assessment*. Buckingham: Open University Press.

