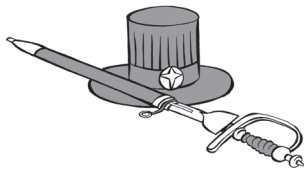


## Najat Ouakrim-Soivio



## Toimiiko päättöarviointi yhdenvertaisesti?

### Kohokohdat

1. Päättöarvosanojen antaminen perustuu nykyisessä peruskoulujärjestelmässä siihen, että oppiainekohdaisilla päättöarvioinnin kriteereillä (arvosanalle 8) pyritään määrittämään yhdenmukaisesti se tieto- ja taitotaso, joka oppilaalla oletetaan olevan perusopetuksen päättövaiheessa hänen siirtyessään jatko-opintoihin tai työelämään.
2. Väitöstutkimuksessa analysoitiin opettajien antamien päättöarvosanojen luotettavuutta mitata perusopetuksen opetussuunnitelman perusteissa määriteltyä osaamista. Tulokset osoittivat, että yhteiskuntaopissa ja historiassa oli systemaattisia koulujen välisiä eroja seuranta-arvioinnissa osoitetun osaamisen ja koulujen antamien arvosanojen välillä.
3. Näyttää siltä, että opettajat suhteuttavat antamansa arvosanat valtakunnallisten kriteerien sijasta muiden oppilaidensa osaamistasoon, vaikka suhteellinen arviointi on poistunut normeista jo vuonna 1985.
4. Kansainvälisestikin mainetta niittänyt suomalainen peruskoulujärjestelmä, jonka lähtökohtin on koko sen voimassaoloajan kuulunut koulutuksellinen tasa-arvo ja yhdenvertaisuus, joutuu koeteltavaksi, kun oppiaineesta annettujen arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen välillä on suuria koulujen välisiä eroja. Yhdenmukaisen arvioinnin varmistamiseksi opetussuunnitelman perusteiden olisi tarjottava aiempaa enemmän tukea opettajan arviointityöhön.

Najat Ouakrim-Soivion kasvatustieteen väitöskirja ”Toimivatko päättöarvioinnin kriteerit? Opetushallituksen oppimistulosten seuranta-arviointi koulujen välisten osaamiserojen mittareina” tarkastettiin Helsingin yliopiston käyttäytymistieteellisessä tiedekunnassa 8. marraskuuta 2013. Vastaväittäjänä toimi professori Arja Virta Turun yliopistosta ja kustoksena professori Jukka Rantala Helsingin yliopistosta.

Väitöstutkimuksessa analysoitiin koulujen väli-

siä eroja oppilaiden saamien arvosanojen ja heidän oppimistulosten arvioinnissa osoittamansa osaamisen suhteen; tutkimuksen kohteena olivat erot oppilaiden saamien kouluarvosanojen ja seuranta-arviointimenestyksen välillä. Tutkimuksessa tarkasteltiin opettajien antamien arvosanojen luotettavuutta osaamisen mittarina. Oppimistuloksia arvioitiin otospohjaisesti ja tuloksia hyödynnetään arvioitaessa opetussuunnitelman perusteissa määriteltyjen tavoitteiden saavuttamista. Tutkimuksen otos käsitti 3 613 suomenkielisten koulujen oppilasta ja 545

ruotsinkielisten koulujen oppilasta eli yhteensä 4 158 oppilasta. Kouluja oli yhteensä 109.

Yhteiskuntaopista ja historiasta kerätyllä aineistolla kyettiin osoittamaan säännönmukaisuus, joka muodostuu koulujen osaamistason ja oppilasarvioinnin välille: jos koulun osaamistaso on seuranta-arvioinnin perusteella hyvä, oppilaat saavat kouluaikana osamistaan heikompia arvosanoja. Kahden eri oppiaineen samanaikaisella arvioinnilla voitiin ratkaista kahden mittauksen eli arvosanojen ja seuranta-arvioinnin validiusongelma. Kouluisa, joissa yhteiskuntaopin arvosanat annettiin esimerkiksi tiukemmin perustein ja arvosanat olivat heikompia kuin seuranta-arvioinnissa osoitettu keskimääräinen osaaminen olisi antanut olettaa, annettiin tiukemmin perustein myös historian päättöarvosana. Siksi koulujen välillä tuntuu olevan sellaisia oppiaineesta riippumattomia arviointikäytänteiden eroja, joiden perusteella voidaan tulkita, että arvosanat eivät mittaa validisti oppilaiden osaamista, vaan samalla osaamisella voi eri kouluissa saada erilaisen arvosanan. Näyttää siltä, että opettajat yhä suhteuttavat antamansa arvosanat valtakunnallisten kriteerien sijasta muiden oppilaidensa osaamistasoon. Koska perusopetuksen opetussuunnitelman perusteissa määritellyt kriteerit eivät koulutasolla toimi arvioinnin perustana, päättöarvioinnin yhdenvertaisuus ei toteudu.

Väitöskirja e-julkaisuna osoitteessa <https://helda.helsinki.fi/handle/10138/41026>

**Asiasanat: perusopetus, oppimistulosten arviointi, päättöarvioinnin kriteerit, arvosanat, validius**

Oppimistulosten mittaaminen saattaa käsitteenä herättää ihmisissä voimakkaita tunteita ja vahvoja mielipiteitä. Taustalla voi olla mielikuvia oppilaiden testaamisesta, heidän asettamisestaan testitulosten

mukaan paremmuusjärjestykseen, päättökokeista ja koulujen ranking-listoista, joissa kouluja jaetaan saavutettujen tulosten mukaan ”hyviin” ja ”huonoihin”. Taustalla saattaa olla myös omakohtaisia kokemuksia siitä, että on tullut arvioitua väärin tai epäreilusti.

Oppimistuloksia voidaan arvioida erilaisilla mittareilla, kuten arvosanoilla, joista meillä kaikilla lienee kokemusta, tai esimerkiksi seuranta-arvioinneilla, jotka osaamisen mittarina ovat suurimmalle osalle vieraampia. Arvosanojen antaminen perustuu nykyisessä peruskoulujärjestelmässä siihen, että oppilaan saavuttamia tietoja ja taitoja verrataan perusopetuksen opetussuunnitelman perusteissa (POPS 2004) määriteltyyn hyvän osaamiseen kuvaukseen tai perusopetuksen päättövaiheessa päättöarvioinnin kriteereihin eli arvosanalle 8 kirjoitettuun kuvaukseen (POPS, 2004). Oppiainekohtaisilla päättöarvioinnin kriteereillä määritellään kansallisesti se tieto- ja taitotaso, joka oppilaalla oletetaan olevan perusopetuksen päättövaiheessa, kun hän siirtyy jatko-opintoihin tai työelämään. Tätä arviointitapaa kutsutaan kriteeriperusteiseksi arviointiksi. Oppilaan suorituksia ei siis enää verrata muiden oppilaiden suorituksiin, kuten suhteellisissa arvioinnissa aiemmin tehtiin, vaan arviointi kohdistuu siihen, miten oppilas on suoriutunut suhteessa ennalta asetettuihin tavoitteisiin. Suhteellinen arviointi ja Gaussin käyrän käyttäminen istuvat kuitenkin tiukasti suomalaisessa arviointikulttuurissa, vaikka suhteellinen arviointitapa poistettiin peruskoulun opetussuunnitelmaperusteista jo vuonna 1985.

Toinen oppimistulosten mittaamismenetelmä ovat seuranta-arvioinnit, jotka suomalaisessa koulutusjärjestelmässä

kuuluvat vielä toistaiseksi Opetushallituksen tehtäviin. Opetushallitus on tehnyt, raportoinut ja seurannut oppimistulosten arviointia yli 15 vuoden ajan. Systemaattisimmin Suomessa on seurattu matematiikan, äidinkielen sekä modersmålet och litteratur -oppiaineen oppimistuloksia, mutta tähän mennessä kaikkien perusopetuksessa opetettavien aineiden oppimistuloksia on arvioitu ainakin yhden kerran. Oppimistulosten arviointien lähtöoletuksena on, että opetussuunnitelman perusteisiin on kirjattu ne tavoitteet ja oppiainekohtaiset tiedot ja taidot, jotka kaikkien oppilaiden tulee perusopetuksen päättövaiheessa hallita. Opetushallituksen tekemien oppimistulosarviointien tarkoituksena on saada tietoa siitä, miten hyvin opetussuunnitelman perusteissa määritellyt koulutuksen tavoitteet on saavutettu. Niiden tehtävänä on myös antaa koulutuksen järjestäjille, esimerkiksi kunnille ja kouluille, tietoa siitä, millaisia oppilaiden keskimääräinen osaaminen ja koulun oppilasarvioinnin yleinen linja todistusarvosanoissa ovat kansalliseen tasoon verrattuna. Oppimistulosten arvioinnissa ei siis ole kyseessä koko ikäluokan päättökoe, eikä niiden tarkoituksena ole määrittää yksittäisten koulujen tai oppilaiden keskinäistä järjestystä.

Koulutuksellisen tasa-arvon toteuttaminen on noussut pääkriteeriksi arvioitaessa peruskoulun ja opetussuunnitelman perusteiden toimivuutta oppimistulosten kannalta. Käytännössä tämä tarkoittaa sitä, että kulloisenkin oppiaineen oppimistuloksista tarkastellaan, toteutuuko niissä alueellinen, kielellinen sekä sukupuolten ja sosiaaliryhmien välinen tasa-arvo. Koulutuksen tasa-arvon lisäksi oppimistuloksissa arvioidaan myös koulutuksen saatavuutta, jolla tarkoitetaan oppilaiden

yhdenvertaisia mahdollisuuksia jatkaa opintojaan peruskoulun jälkeen. Koulutuksellisen tasa-arvon voidaan katsoa toteutuneen hyvin, mikäli systemaattisia eroja ei ilmene näiden tarkasteltavien osa-alueiden suhteen.

Opetushallitus julkaisi ensimmäistä kertaa yhteiskunnallisten aineiden eli historian ja yhteiskuntaopin oppimistulosten seuranta-arvioinnin tulokset. Niiden mukaan sekä historiassa että yhteiskuntaopissa oli systemaattisia eroja seuranta-arvioinnissa osoitetun osaamisen ja koulujen antamien arvosanojen välillä.

Koulut jaettiin neljään ryhmään sen mukaan, millainen niiden keskimääräinen yhteiskuntaopin osaamistaso oli ollut seuranta-arvioinnissa. Kuviossa 1 ilmenee, että kouluissa, joissa yhteiskuntaopin seuranta-arvioinneissa on menestytty keskimäärin hyvin (Q4), on annettu samanlaisista suorituksista alempia arvosanoja kuin niissä kouluissa, joissa seuranta-arvioinneissa on pärjätty heikosti (Q1). Käytännössä tämä tarkoittaa sitä, että yhteiskuntaopissa 55 prosentin ratkaisuosuudella sai keskimääräisen osaamisensa puolesta parhaimpaan neljännekseen kuuluvassa koulussa (Q4) keskimäärin arvosanan 5 ja heikoimmin menestyneeseen neljännekseen kuuluvassa koulussa (Q1) arvosanan 7. Kuvion 1 tulos on tehty yhteiskuntaopin aineistosta, mutta myös toisessa oppiaineessa eli historiassa tulos oli samanlainen. Tämä sai pohtimaan perusteellisemmin kysymystä siitä, toimivatko päätösarvioinnin kriteerit yhdenvertaisesti.

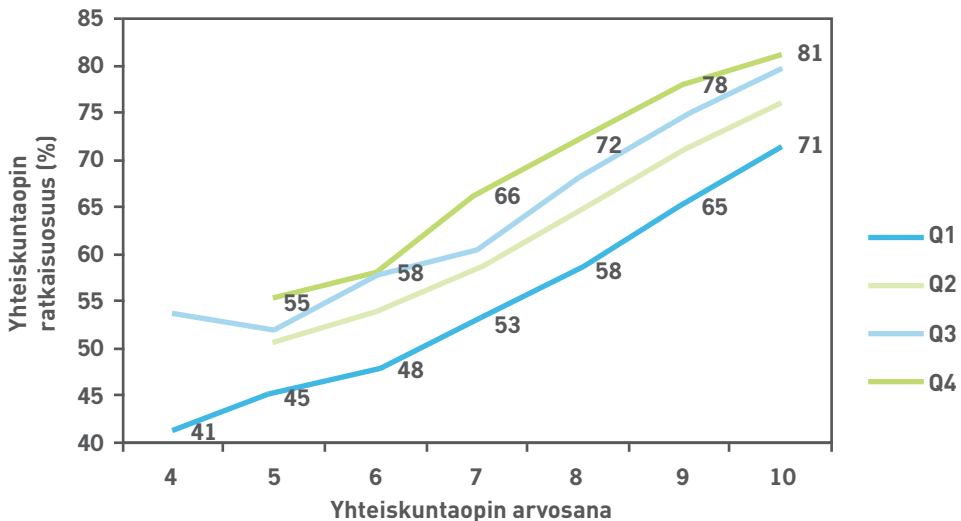
Edellä esitetyissä tutkimustuloksissa ei ole kyse siitä, ovatko oppilaat saaneet ”oikean arvosanan”, sillä sitä on lähes mahdotonta arvioida. Kyse on ennen kaikkea siitä, saavatko oppilaat arvosanansa

ja päättöarvosanansa yhdenvertaisin perustein ja samanlaisesta osaamistasosta. Osaamistasolla tarkoitetaan oppilaiden suoritusta seuranta-arvioinnissa eli ratkaisuprosenttia. Seuranta-arvioinnin ilmaiseva osaamistaso sekä arvosanat ovat kaksi erilaista mittaria, mutta ne mittaavat samaa asiaa eli Perusopetuksen opetussuunnitelman perusteissa (2004) määritettyjen tietojen ja taitojen saavuttamista.

Kansainvälisestikin mainetta nauttanut suomalainen peruskoulujärjestelmä, jonka lähtökohtiin on koko ajan kuulunut kouluksellinen tasa-arvo ja yhdenvertaisuus, joutuu koeteltavaksi, kun oppiaineesta annettujen arvosanojen ja seuranta-arvioinnissa osoitetun osaamisen välillä on suuria koulujen välisiä eroja. Kuviossa 1 esitetty tutkimustulos ei nimittäin tue käsitystä siitä, että yhdenvertaisuus toteutuisi tällä hetkellä oppilaan arvioinnissa. Perusperiaatteena arvosanoja ja päättöarvosanoja annettaessa on aina ollut se, että

oppilas saa samanlaisesta osaamisesta saman arvosanan riippumatta siitä, missä hän opintonsa suorittaa. Kaikissa oppilaan arviointia ohjaavissa säädösteksteissä on ilmaistu vahvasti periaate oppilaan arvioinnin yhdenmukaisuudesta, joka käytännössä tarkoittaa sitä, että arvosanojen on oltava valtakunnallisesti vertailukelpoisia jatko-opintoihin hakeuduttaessa.

Koulutusjärjestelmästä ja arvosanaasteikoista riippumatta arvosanan keskeinen tehtävä on kertoa oppilaalle ja hänen taustaryhmilleen, että oppilas on saavuttanut opintojen tavoitteet ja voi siirtyä opinnoissaan eteenpäin (Guskey, 2011). Näin on myös suomalaisessa koulutusjärjestelmässä, jossa oppilaat pyrkivät peruskoulun jälkeisiin jatko-opintoihin tai työelämään saamallaan päättöarvosanoilla. Arvosanat ovat usein ainoa tieto, jonka oppilaitos tai työnantaja saa nuoren osaamistasosta. Arvosanoilla voidaan myös motivoida, palkita tai rangaista oppilasta



Kuvio 1. Koulujen osaamistaso arvosanoittain yhteiskuntaopissa. Lähde: Ouakrim-Soivio 2013, 175.

(esim. Hills, 1991), ja tätä kautta niillä on suuri merkitys yksilötasolla, kun oppilaat muodostavat käsitystä omasta suoriutumisestaan ja itsestään oppijana.

Arvosanojen vertailtavuus ja päätösarvosanojen antaminen yhdenvertaisin periaattein on yhteiskunnallisesti ja koulutuspoliittisesti merkittävä asia. Opintojen aikaisten arvosanojen ja päätösarvosanojen validius kytkeytyy mitä suurimmassa määrin paitsi oppilaiden oikeusturvaan myös siihen, kuinka yhdenvertaisin periaattein heidän osaamistaan arvioidaan ja kuinka vertailtavia päätösarvosanat valtakunnallisesti ovat.

Tämän tutkimuksen keskeisimpiä käsitteitä on validius, jolla yleensä tarkoitetaan pätevyyttä ja luotettavuutta, mittaustarkkuutta ja systemaattisten virheiden puuttumista. Nykyisin validiuteen katsotaan kuuluvaksi myös mittaustuloksista tehtävien päätelmien sopivuus, mielekyys ja käyttökelpoisuus (Nummenmaa, 1997). Modernia validius-käsitettä alkoivat 1980-luvulta lähtien kehittää sellaiset teoreetikot kuin Cronbach, Messick ja Kane, joiden mukaan validiuden tarkastelussa tulee ottaa huomioon paitsi mittari(t), myös dataan tai aineistoon perustuva tulkinta (Cronbach, 1971). Messick lähestyi validiutta kahden kysymyksen avulla, jotka ovat sovellettavissa myös arvioinnin kontekstiin:

1. mittaako testi tai koe riittävän hyvin niitä ominaisuuksia, joita sen tulisi mitata, ja
2. soveltuuko testi tai koe käytettäväksi esitettyyn tarkoitukseen esitetyllä tavalla?

Ensimmäiseen kysymykseen vastaus löytyy testien ominaisuuksien avulla ja tarkastelemalla esimerkiksi tutkimusasetelmaa käsitevalidiuden näkökulmasta. Toiseen kysymykseen vastaaminen edellyttää eettistä pohdintaa ja sosiaalisten arvojen kytkemistä vastaukseen (Messick, 1980). Myönteisestä vaikutuksesta voisi olla esimerkkinä kansainvälisten oppimistulosvertailujen pohjalta kansalliseen koulutuspolitiikkaan tehtävät uudistukset, kuten kansainväliset Education at Glance 2010- ja ICCS-tutkimukset. Kyseiset tutkimukset osoittivat, että kansainvälisesti vertailtuna yhteiskuntaoppia tai sen sisältöjä opetetaan Suomessa varsin vähän ja peruskoulun alaluokilla ei juuri lainkaan. (Education at Glance, 2010; ICCS, 2011.) Sittemmin kesäkuussa 2012 valtioneuvoston antama asetus perusopetuksen uudesta tuntijaosta muutti tilannetta niin, että yhteiskuntaoppia aletaan opettaa myös alaluokilla.

Kielteisenä esimerkkinä tulosten vaikutuksesta voisi puolestaan olla tilanne, jossa tuloksia käytetään muuhun kuin siihen, mihin niitä on alun perin aiottu. Esimerkkinä tästä voisi olla Yhdysvalloissa 2000-luvun alussa säädetty No Child Left Behind -lainsäädäntö, jolla alun perin pyrittiin yhdenmukaistamaan sitä, mitä oppilaiden tulisi esimerkiksi matematiikassa, luonnontieteissä ja englannin kielessä oppia ja miten oppimista tulisi mitata (Mathinson, Ross & Vinson, 2006). Kyseinen lainsäädäntö on kuitenkin johtanut siihen, että oppilaiden oppimistuloksia mitataan siksi, että voitaisiin arvioida koulujen niin sanottua tilivelvollisuutta (accountability) (Pinar, 2004; Koretz & Hamilton, 2006; Mathinson, Ross & Vinson, 2006; Abrams, 2007).

Arviointi ja validius kytkeytyvät käsitteinä tiiviisti toisiinsa. Arvioinnin on oltava vertailtavaa, oikeudenmukaista, reliaabelia ja validia (Messick, 1994; 1995). Oikeudenmukaisuus johtaa vaatimukseen yhdenvertaisuudesta, jolla tarkoitetaan tuloksen ja siitä tehtävien tulkintojen arvovaputta ja puolueettomuutta. Edellä mainitut periaatteet eivät liity vain mittaamiseen vaan ne ovat tärkeitä arvoja siinä vaiheessa, kun arvioinnin perusteella tehdään päätöksiä. (Messick, 1995.) Aina, kun tuloksia analysoidaan tai tulkitaan tai tuloksia käytetään johonkin tarkoitukseen, on kyse myös vallasta ja siitä, kenen lähtökohdista käsin arvotetaan, mitataan ja määritellään.

Sir Francis Baconin toteamus ”tieto itsessään on valtaa” voitaisiin muuttaa muotoon ”arviointi on valtaa”. Arvioijalla oleva tieto tekee hänet vaikutusvaltaiseksi (Atjonen, 2007), sillä arvioijan asema ja auktoriteetti suhteessa arvioitavaan sisältävät aina vallankäytön mahdollisuuden. Tyypillinen esimerkki arvioijan valta- asemasta on opettaja–oppilas-suhde, jossa opettajalle on annettu valtaa suhteessa oppilaaseen esimerkiksi siinä, että opettaja voi määritellä, millä kriteereillä (Heinonen & Jakku-Sihvonen, 2001) ja mitä hän arvioi. Opettajalle tai arvioijalle annettuun valtaan kuuluu kuitenkin myös vastuu siitä, että hän arvottavia päätöksiä tehdessään ymmärtää arvioida myös tekemiensä ratkaisujen seurauksia ja vaikutuksia (Atjonen, 2007).

On selvää, että oppilaan arvioinnin ja arvosanan antamisen tulee olla oikeudenmukaista eikä arviointi saa perustua esimerkiksi sukupuoleen, etniseen taustaan, sosioekonomiseen asemaan tai muihin oppilaan ominaisuuksiin (Loyd & Loyd, 1997). Vaikka opettajat arviointityössään

pyrkivät objektiivisuuteen, tiedetään, että arviointi ja arvosanojen antaminen on aina jossain määrin subjektiivista. Opettajalla on arvioijana valtaa koota tietoa ja tehdä niiden perusteella päätöksiä (Atjonen, 2007), jotka vaikuttavat oppilaaseen.

Brookhartin (1994) tekemässä meta-analysissä ilmeni, että opettajat pyrkivät parhaan kykynsä mukaan antamaan arvosanoja oikeudenmukaisin periaattein ja selvittämään oppilailleen arvosanan muodostumisperiaatteet. Kuitenkin tutkimukset osoittivat, että opettajat voivat pyrkiä palkitsemaan paremmalla arvosanalla oppilaita, jotka yrittivät kovasti (Brookhart, 1991; Brookhart, 2007), tai oppilaita, joiden sosiaalinen tausta oli muuta ryhmää heikompi (Cross & Frary, 1990). Klapp Lekholm esitti samansuuntaisia tutkimustuloksia Ruotsissa kerätyllä aineistolla kuin Cross ja Frary. Klapp Lekholm arveli, että arvosanojen nostaminen heikosti menestyvissä ryhmissä tai kouluissa saattaa liittyä esimerkiksi koulun yleiseen arviointikulttuuriin siten, että opettajat haluavat mahdollisimman monen oppilaan ylittävän hyväksytyt suorituksen rajan (Klapp Lekholm, 2008). Suomessa esimerkiksi Atjonen ja Jakku-Sihvonen ovat korostaneet kriteeriperusteisen oppilaan arvioinnin ”reiluutta ja oikeudenmukaisuutta”, koska kriteerien tavoitteet on määriteltävä etukäteen ja ne perustuvat opetussuunnitelmaan (Jakku-Sihvonen, 2001; Atjonen, 2007), jolloin oppilailla ja heidän huoltajillaan on mahdollisuus saada etukäteen tietoa siitä, mitkä ovat asetetut tavoitteet ja miten tavoitteita arvioidaan.

Suomessa säädökset ohjaavat opettajaa arvioimaan oppilasta monipuolisesti, kuten esimerkiksi seuraamaan myös luokassa tapahtuvaa tuntityöskent-

telyä. Luokkatyöskentelyn validiutta on alettu pohtia järjestelmällisemmin vasta 2000-luvulla. Näyttää siltä, että luokkatyöskentelyn arviointiin liittyy validiusergelmiä, jotka kytkeytyvät opettajan valtaan arvioitsijana. Opettajat voivat valita varsin vapaasti, millä arviointimenetelmillä opetus suunnitelmassa olevien, kaikille yhteisten tavoitteiden saavuttamista arvioidaan, jolloin opettaja on usein viime kädessä se henkilö, joka määrittelee, mitkä tiedot, taidot, oppiminen ja osaaminen vaikuttavat arvosanaan (esim. Citzek, Fritzgerald & Rachor, 1995/1996). Opettajilla on helposti kaksoisrooli sekä arvioijana että havainnoijana. Koska havainnointi on kaksisuuntaista, opettaja saattaa omalla käyttäytymisellään tahtomattaan vahvistaa joidenkin oppilaiden työskentelyä ja antaa heille esimerkiksi enemmän vastausaikaa tai vastauksia. (Bulterman-Bos ym., 2002.) Opettajien omat kokemukset ja mielikuvat aiemmin koulua käyneiden oppilaiden toiminnasta vastaavassa tilanteessa voivat vaikuttaa senhetkisten oppilaiden arviointiin (Brookhart, 2007), jolloin aiemmat oppilassukupolvet toimivat ikään kuin normina, johon senhetkisen oppilasryhmän työskentelyä, tietoja ja taitoja verrataan. Opettajien omat arviointitiedot ja taito käyttää eri arviointimenetelmiä vaihtelevat suuresti (Phye, 1997). Lisäksi osa opettajista saattaa kokea arviointityön ikäväksi velvollisuudeksi (Green & Emerson, 2007).

Tutkimusten esille tuomat oppilaan arvioinnin ongelmakohdat saavat pohtimaan sitä, onko opettajille annettu arvioinnin ja arvosanojen antamisen suhteen paljon valtaa, mutta vain vähän työkaluja. Opettajan työhön kuuluu olennaisesti sekä opettaminen että arviointi. Pitkään opetus-

työtä tehneenä minusta kuitenkin näyttää siltä, että kiinnitämme sekä opettajankoulutuksessa että koulun arjessa yhä edelleen paljon huomiota opettamiseen, mutta vain vähän arviointiin.

Turusen mukaan arviointia ei voi olla ilman perusteita, jolloin ydinkysymykseksi nousee se, kenellä on oikeus tai valta-asema asettaa arviointiperusteita (Turunen, 1999). Arviointiin liittyy aina myös valintojen tekemistä, jolloin valtaa käytetään esimerkiksi siinä, kuka tai ketkä valintoja tekevät ja millä perusteella niitä tehdään (Atjonen, 2007). Suomalaisessa koulutusjärjestelmässä myös Opetushallitukselle on annettu paljon arviointiin liittyvää valtaa, sillä sen tehtäviin kuuluu toisaalta mitata ja arvioida oppimistulosten seuranta-arviointien kautta opetus suunnitelmassa asetettujen tavoitteiden toteutumista, toisaalta kehittää opetus suunnitelmien perusteita. Tällä hetkellä Opetushallitus tekee valintoja ja käyttää sille annettua valtaa, kun se valmistelee perusopetuksen opetus suunnitelman perusteita, joiden pohjalta opettajat tekevät käytännön arviointityötä. Opetushallituksen antamien normien avulla myös määritellään, kuinka yhdenvertaisin ja yhteismittallisin periaattein oppilaiden päättöarvosanat annetaan.

Väitöstutkimukseni keskeisin tulos on, että päättöarvosanat eivät kerro validisti siitä, millainen oppilaiden keskimääräinen osaamistaso on verrattuna suomalaisten oppilaiden seuranta-arvioinnissa osoittamaan keskimääräiseen osaamiseen, koska perusopetuksen opetus suunnitelman perusteissa (2004) määritellyt päättöarvioinnin kriteerit arvosanalle 8 eivät tue riittävän hyvin opettajien arviointityötä. Käsitemitallisuuden näkökulmasta tutkimuksen todellinen arvo mitataan siinä,



kuinka hyvin tulokset palvelevat niitä käytäviä tahoja, kuten poliitikkoja (Messick, 1980; Messick, 1995). Myös tämän väitöstutkimuksen validius ja arvo mitattaneen siinä, kuinka hyvin tutkimustulokset puhuttelevat päätöksentekijöitä.

### Kirjoittajatiedot:

Najat Ouakrim-Soivio toimii opetus- ja kulttuuriministeriössä opetusneuvoksena. Ennen perusopetuksen ja lukiokoulutuksen asiantuntijatehtäviä hän on työskennellyt historian ja yhteiskuntaopin opettajana, rehtorina ja opettajankouluttajana. Lisäksi Najat Ouakrim-Soivio on toiminut historian ja yhteiskuntaopin oppimistulosten arvioinnin projektipäällikkönä Helsingin yliopiston koulutuksen arviointikeskuksessa ja Opetushallituksessa (2011–2012), kun kyseisten oppiaineiden oppimistulokset arvioitiin ensimmäisen kerran. Najat Ouakrim-Soivio väitteli 8.11.2013 sen aineiston pohjalta, jota OPH ei raportoinut ja julkaissut vuonna 2012, kun yhteiskunnallisten aineiden oppimistulokset julkaistiin (Ouakrim-Soivio, N & Kuusela, J., 2012. Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011. Koulutuksen seurantaraportit 2012:3. Helsinki: Opetushallitus).

## LÄHTEET

- Abrams, L. M. (2007). Implications of high stakes testing for the use of formative classroom assessment. Teoksessa J. McMillan (toim.), *Formative classroom achievement* (s. 79–98). New York: Teachers College Press.
- Atjonen, P. (2007). Hyvä, paha arviointi. Helsinki: Kustannusosakeyhtiö Tammi.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10, 35–36.
- Brookhart, S. M. (1994). Teacher's grading: Theory and practice. *Applied measurement in Education* 7, 279–301.
- Brookhart, S. (2007). Expanding views about formative classroom assessment: A review of the Literature. Teoksessa J. McMillan (toim.), *Formative Classroom Achievement* (s. 43–62). New York: Teachers College Press.
- Bulterman-Bos, J., Terwel, J., Verloop, N. & Wardekker, W. (2002). Observation in teaching: Toward a practice of objectivity. *Teachers College record*, 104, 1069–1100.
- Cizek, G. C., Fitzgerald, S. M. & Rachor, R. E. (1995/1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment* 3, 159–179.
- Cronbach, L. J. (1971). Test validation. Teoksessa R. L. Thorndike (toim.), *Educational measurement* (toinen painos.) (s. 443–507). Washington DC: American Council on Education and Praeger Publishers,
- Cross, L. H. & Frary, R. B. (1999). Hodgebodge grading: Endorsed by students and teachers alike. *Applied measurement in Education*, 12, 53–72.
- Green, K. H. & Emerson, A. (2007). A New framework for grading. *Assessment and Evaluation in Higher Education* 32, 496–511.
- Guskey, T. R. (2011). Stability and change in high school grades. *National Association of Secondary School Principals bulletin*, 95, 85–98.
- Heinonen, S. & Jakku-Sihvonen, R. (2001). Koulutuksen arvioinnin etiikka. Teoksessa R. Jakku-Sihvonen & S. Heinonen, *Johdatus koulutuksen uudistuvaan arviointikulttuuriin* (s. 95–110). Arviointi 2/2001. Helsinki: Opetushallitus.
- Hills, J. R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan*, 72, 540–545.
- Jakku-Sihvonen, R. (2001). Arviointitiedon luotettavuuden osoittaminen. Teoksessa R. Jakku-Sihvonen & S. Heinonen (toim.), *Johdatus koulutuksen uudistuvaan arviointikulttuuriin* (s. 111–135). Arviointi 2/2001. Helsinki: Opetushallitus.
- Klapp Leckholm, A. (2008). Grades and grade assignment: effects of student and school characteristics. Göteborg: Göteborgs Universitetet. Haettu 29.11.2012 osoitteesta <https://gupea.ub.gu.se/handle/2077/18673>
- Koretz, D. M. & Hamilton, L. S. (2006). Testing for Accountability in K-12. Teoksessa R. L. Brennan (toim.), *Educational Measurement* (neljäs painos) (s. 531–578). Westport, CT: American Council on Education and Praeger Publishers.
- Loyd, B. H. & Loyd, D.E. (1997). Kindergarten through Grade 12 Standards: A philosophy of grading. Teoksessa G.D. Phye (toim.), *Handbook of classroom assessment, learning, adjustment and achievement* (s. 481–489). San Diego: Academic Press.
- Mathinson, S., Ross, E. W. & Vinson K. D. (2006). Defining the social studies curriculum: Influence of resistance to curriculum standards and testing in social studies. Teoksessa
- E.W. Ross (toim.), *Social studies curriculum purposes, problems and possibilities* (s. 99–114). Kolmas painos. Albany: State University of New York Press.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012–1027.
- Messick, S. (1994). The interplay of evidence



- and consequences in the validation of performance assessment. *Educational Researcher*, 23, 13–23.
- Messick, S. (1995). Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Nummenmaa T., Konttinen R., Kuusinen J. & Leskinen E. (1997). Tutkimusaineiston analyysi. Porvoo: WSOY.
- OECD (2010). *Education at Glance 2010*.
- Ouakrim-Soivio N. (2013). Toimivatko päättöarvioinnin kriteerit? Raportit ja selvitykset 2013:9 Helsinki: Opetushallitus.
- Phye, G. D. (1997). Classroom assessment: A multidimensional perspective. Teoksessa G. D. Phye (toim.), *Handbook of classroom assessment. Learning, adjustment and achievement* (s. 33–51). San Diego: Academic Press.
- Pinar, E.F. (2004). *What is Curriculum Theory?* Mahwah: Lawrence Erlbaum Associated Publishers.
- POPS (2004). *Perusopetuksen opetussuunnitelman perusteet*. Helsinki: Opetushallitus.
- Turunen, K. E. (1999). *Opetustyön perusteet*. Jyväskylä: Atena Kustannus.