

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 27–43.

Henna Tossavainen

Jyväskylän yliopisto

Kielitestien eettisyydestä ja oikeudenmukaisuudesta

In the field of language testing and assessment, questions relating to ethics and fairness have been under discussion and debate since they were explicitly raised in the 1990s. Focusing on social responsibility and equality in institutional standardized tests, this paper describes the two concepts and explores some of the discussion in the field since the emergence of the topic. Connecting ethics to other developments in the field and in particular to professionalism, the paper outlines some concerns of ethical assessment in standardized language tests.

Keywords: ethics, fairness, social responsibility, professionalism, language assessment

Asiasanat: eettisyys, oikeudenmukaisuus, asiantuntijuus, kielitaidon arviointi

1 Johdanto

Eettisyyteen ja oikeudenmukaisuuteen liittyvät kysymykset ovat olleet vahvasti esillä kielitaidon arvioinnin ja kielentestauksen alalla ja kielitestejä koskevassa joskus varsin negatiivisessa kritiikissä 1990-luvulta lähtien. Eettisten näkökulmien mukanaolo tarkoittaa, että alalla tunnustetaan testauksen sosiaalinen rooli – kielitesteillä on eettinen dimensio ns. siksi, että ne vaikuttavat ihmisten elämään. Eettisyys ja oikeudenmukaisuus (*ethics, fairness*) liittyvät siihen, miten testitulokset syntyvät ja mihin ja miten tuloksia käytetään sekä millaisia vaikutuksia tuloksilla on yksilöille ja yhteiskuntaan. Eettisyyden yhteydessä on tapana puhua myös sosiaalisesta vastuunkannosta. (McNamara 2000: 71–72.)

Alan eettistä keskustelua on käyty etupäässä kahdesta näkökulmasta. Toisaalta kielitestit nähdään sosiaalisina ja poliittisina rakenteina ja instituutioina, joiden kriittinen tarkastelu kohdistuu poliittisen vallankäytön paljastamiseen ja vääristyneinä pidettyjen yhteiskunnallisten ja koulutusrakenteiden avaamiseen. Tällöin näkökulma tulee usein kielitestien käytännön toiminnan ulkopuolelta. Toisaalta eettisyyttä voidaan tarkastella kielitestien sisäisinä keinoina tavoitella mahdollisimman tasa-arvoista, tasalaatuaista ja oikeudenmukaista toimintaa. Tämän näkökulman mukaan kielitestejä on mahdollista kehittää eettisempään suuntaan omaksumalla eettisesti hyväksytyjä käytänteitä. Molemmat näkökulmat ovat laajentaneet alalla käytyä keskustelua eettisyyden roolista sekä asettaneet kysymyksiä testauksessa käytettyjen prosessien ja koko toiminnan oikeutuksesta. Alan kansainvälisissä yhteenliittymissä on laadittu esimerkiksi eettinen koodisto (*Code of Ethics* ILTA 2000) ja käytännön toimintaohjeistoja (*Code of Practice* esim. ALTE 2001, EALTA 2007), joiden tehtävänä on turvata testiin osallistujan oikeudet ja määritellä testaajan ja testitulosten käyttäjän vastuut. (McNamara 2000; Lynch 2001; Shohamy 2001b.)

Tarkastelen tässä artikkelissa eettisyyteen ja oikeudenmukaisuuteen liittyviä kysymyksiä. Määrittelen käsitteitä ja kuvaan eettisyyden roolia kielitaidon arvioinnin ja kielentestauksen historiassa sekä alalla käytyä keskustelua eettisyydestä ja ammattilaisuudesta. Lopuksi tuon esiin ja pohdin standardoiduissa kielitesteissä tehtäviä eettisyyteen ja oikeudenmukaisuuteen tähtääviä toimenpiteitä. Käytän termiä 'kielitaidon arviointi' (*language assessment*) alaan liittyvänä yläkäsitteenä, termiä 'kielentestaus' (*language testing*) testauksen toimintoihin ja toimijoihin liittyvänä käsitteenä sekä termiä 'mittaaminen' tai 'mittaustulos' (*measurement*) viittaamaan alalla tehtävään laskennalliseen työhön. Käytän tekstissä käsitteestä *fairness* vastinetta 'oikeudenmukaisuus'. Käsitteestä käytetään myös käänkösvastinetta 'reiluus' (esim. Atjonen 2007).

2 Eettisyyden ja oikeudenmukaisuuden määrittelyä

Kielitaidon arviointi, kuten arviointi yleensä, on arvottamista, toiminnan suhteuttamista hyväksytyinä pidettyihin arvoihin. Kielitesteissä voidaan esimerkiksi sopia eettisesti hyvänä pidettävän arvioinnin perusteista. Yleisesti etiikan voidaan sanoa tarkoittavan jakoa hyvään ja pahaan, oikeaan ja väärään, hyväksyttävään ja ei-hyväksyttävään toimintaan sekä näiden selvittämistä. Taylorin (2013: 1) mukaan eettisyys viittaa hyväksytyinä pidettäviin uskomuksiin ja käytänteisiin, joiden tavoitteena on jonkin käyttäytymisen rajoittaminen yhteisen hyvän nimissä. Etiikka on täten luonteeltaan vuorovaikutteista (Atjonen 2007: 11). Yhteiseen hyvään voidaan liittää ajatus toiminnan eettisistä seurauksista. Usein on tarpeen määritellä eettiset normit ja toimintaohjeet siksi, että toiminnan rajat ovat selvät kaikille siinä toimiville sekä tärkeille sidosryhmille. (Atjonen 2007: 228; Linnankylä ja Atjonen 2008: 47; ks. Tarnanen 2014.)

Standardoitujen kielitestien yhteydessä voidaan ajatella ainakin seuraavia etiikan lähestymistapoja. Jos testit mielletään pelkästään todisteiden avulla kerättäväksi mittauksilokseksi, jossa erotetaan toisistaan testattava tieto ja taustalla vaikuttavat arvot, voidaan puhua tieteellis-rationaalisesta lähestymistavasta ja psykometriikan paradigmasta. Tavoitteena on tällöin testeissä saatavan tiedon luotettavuuden varmistaminen (Linnankylä & Atjonen 2008: 47). Etiikan yksi tarkastelunäkökulma on konsekventalistinen etiikka, jossa kyse on tekojen seurausten tarkastelusta. Yksi tämän etiikan muoto on utilitarismi, joka tähdentää toiminnasta saatavaa suurinta mahdollista hyötyä ja siitä aiheutuvaa vähintään mahdollista haittaa (Taylor 2013: 1; Atjonen 2007: 228; Spooft 2007: 8). Eettisyyden voidaan sanoa kielentestauksessa tarkoittavan sitä, että testaaja kantaa testien seurausten sosiaalista vastuuta ja käyttää kaikkia mahdollisia keinoja oikeudenmukaisuuden ja tasa-arvon lisäämiseen. Tällä eettisyyden näkemyksellä on juurensa sosiaalisen oikeudenmukaisuuden teorioissa. (Ks. Taylor 2013; Atjonen 2007.) Kielitaidon arvioinnin alalla tähdennetään nykyisin sosiaalieettistä lähestymistapaa (Taylor 2013, Davies 2004; McNamara & Roever 2006; Weir 2005).

Etiikan yhteydessä puhutaan usein **moraalista**, ja etiikkaa myös kutsutaan moraalifilosofiaksi. (Ks. Atjonen 2007: 12.) Esimerkiksi Davies (2004: 98) painottaa sitä, että eettisesti hyväksyttävät ratkaisut eivät voi perustua yksittäisten ihmisten uskomuksiin, vaan kyseessä ovat yhteisesti sovitut moraaliset periaatteet. Kielentestaus tapahtuu yleisimmin institutionaalisissa kehyksissä, ja toiminnan toteutus voidaan määritellä ammattimaiseksi. Eettisyydellä on selkeä tehtävänsä määriteltäessä institutionaalisen toiminnan rajoja, toisin sanoen toimintaan kuuluvia oikeuksia, vastuita ja velvollisuuksia. Fulcher ja Davidson (2010: 273–274) määrittelevät institutionaalisen etiikan ja moraalien olevan julkisen ja yksityisen moraalien välissä. Yksityinen moraalit on esimerkiksi uskonnon aikaansaama ja liittyy usein käsitteeseen omatunto. Ammattieettiset ratkaisut puo-

lestaan pohjaavat ammattialojen moraaliiin, jonka mukaan määritellään ammattiin kuuluvat toimintaohjeet, sopimukset, ammatillinen koulutus, eettiset normit ja standardit.

Etiikkaan liittyvät kiinteästi **arvot**, joiden perusteella esimerkiksi määritellään, mitkä asiat kielitaidon arvioinnissa ovat tärkeitä ja mitä kaikkea tarvitaan oikeudenmukaisuuden varmentamiseen. Bachmanin (1990: 279) mukaan kielitestit eivät ole arvovapaita, koska niitä käytetään palvelemaan jotain yhteiskunnallista tai koulutuksellista tavoitetta. Arvot ovat olleet kielitaidon arvioinnin teorioissa jo kauan mukana; esimerkiksi Messick (1980, 1981) toi arvot mukaan kielitaidon arvioinnin keskusteluun jo 1980-luvulla. Psykometriikassa Cronbach (1988: 6–7) painotti, että kielientestaaajan tulisi toimia yhteistyössä päättäjien kanssa ja ilmaista tarkasti, mitä testin avulla pystytään saavuttamaan ja mitä ei, eli mitkä ovat siis testin taustalla vallitsevat arvot. Arvoja ei pystytä laskennallisesti määrittelemään, mistä voi johtua se, että niistä on alalla tärkeänä pidetyn validiteetin yhteydessä puhuttu vain vähän (Fulcher 2015: 88).

Yksi keskeisiä eettisiä kysymyksiä alalla on ollut **oikeudenmukaisuus** ja sen varmistamiseen tähtäävät toimet kielitesteissä. Oikeudenmukaisuus tarkoittaa testaaajan velvollisuutta varmistaa, että kaikkia mahdollisia keinoja käytetään mahdollisimman tasapuolisen ja tasalaatuisen testin sekä testitulosten saavuttamiseksi (Hamp-Lyons 2001: 124). Nämä liittyvät standardoiduissa kielitesteissä tärkeinä pidettyihin reliabiliteettiin ja validiteettiin, mutta myös testin saatavuuteen ja kaikkien osallistujien mahdollisimman tasapuoliseen kohteluun. Weirin (2005: 1) mukaan kielitestien oikeudenmukaisuuden varmistaminen on välttämätöntä, sillä testeissä pitää pystyä osoittamaan kielitestien eettiset valinnat koskien niitä kriteereitä, joilla arvioidaan ja niitä keinoja, joilla testeissä operoidaan. Voidaan sanoa, että ilman oikeudenmukaisuuteen ja tasa-arvoiseen kohteluun tähtäviä toimia mikään kielitesti ei voi olla kovin vahvoilla eettisillä perusteilla toteutettua toimintaa.

Oikeudenmukaisuuden saavuttamiseen on usein liitetty tarkastelu testin reliabiliteetista ja validiteetista, toisin sanoen niistä toimista, jotka tähtäävät sen selvittämiseen, ovatko käytetyt testauskeinot oikeutettuja kuvaamaan sitä kykyä tai taitoa, jota ollaan arvioimassa (validiteetti), ja onko testi johdonmukainen ja toistettava (reliabiliteetti). Testin oikeudenmukaisuutta tarkastellaan useimmiten tilastollisin keinoin. Voidaan mitata, suosivatko tai syrjivätkö testitehtävät ja tehtävien osiot, arviointikäytänteet tai testin järjestelyt joitakin henkilöitä tai henkilöryhmiä perusteettomasti ja testiin kuulumattoman ominaisuuden kautta. Testiaineistoa analysoimalla voidaan saada selville eroja esimerkiksi sukupuolen, iän, ensikielen, etnisen taustan tai koulutuksen mukaisissa ryhmissä. Analyysien perusteella voidaan osoittaa, onko jokin näistä ryhmistä saanut perusteetonta hyötyä tai haittaa. Oikeudenmukaisuuden näkökulmasta on myös tärkeää, että testitulokset heijastavat, mitä osallistujat ovat kielestä opiskelleet ja oppineet. Saadulla todistuksella on myös oltava käyttöarvoa testiin osallistujalle. Xin (2010) mukaan kaikki, mikä heikentää testin oikeudenmukaisuutta, vaarantaa testin

validiteettia. Kyseessä on paitsi itse testin oikeudenmukaisuus, myös testitulosten tulkintaan liittyvä oikeudenmukaisuus, eli kysymys siitä, mihin ja miten testituloksia käytetään. (Davies, Brown, Elder, Hill, Lumley & McNamara 1999: 199; Van Avermaet, Kuijper & Saville 2004: 140.)

3.1 Eettisyyden näkyminen kielitaidon arvioinnin ja kielentestauksen alalla

Alan ensimmäisenä varsinaisena eettisten kysymysten äänenä Bachman (2000:16) ja monet muut pitävät Spolskya (1981), joka toi esiin testien mahdolliset väärinkäytöt ja negatiiviset vaikutukset. Stevenson (1981) puolestaan peräänkuulutti testaaajan vastuuta ja yhteisten kansainvälisten normistojen kehittämistä. Tärkeänä alkutekijänä voidaan pitää myös Canalen (1988) painottamaa tiedonkerääjän vastuuta. Lisäksi Messick (1989a, 1989b) ulotti validiteetin käsittämään testeissä annettujen tulosten sosiaaliset seuraukset ja testitulosten tulkintaan liittyvät arvolataukset ja siten kytki sosiaalisen oikeudenmukaisuuden validiteetin käsitteeseen. Erityisesti näitä näkökulmia on pidetty merkittävänä eettisen keskustelun alkuunpanijoita aikana, jolloin kielentestauksessa keskityttiin laskennallisesti määriteltäviin ja teknisiin keinoihin testien reliabiliteetin saavuttamiseksi (Hamp-Lyons 1997: 323).

Eettisyyden voidaan kuitenkin sanoa olleen läsnä siinä, että reliabiliteetin ja validiteetin mittaamisen avulla on pyritty oikeudenmukaiseen toimintaan ja arvioitavien tasapuoliseen kohteluun. Lisäksi on mahdollista Tayloria (2013: 1) myötäillen todeta, että ammattimaisesti toteutetuissa kielitesteissä on ymmärretty testien merkitys osallistujien elämässä ja että arvioijien tavoitteena voidaan sanoa olleen mahdollisimman oikeudenmukainen toiminta. Kielentestaajilla ei voi sanoa olleen epäeettisiä tavoitteita, kun toiminta on perustunut oman aikansa hyvinä pidettyihin toimintaperiaatteisiin.

Eettiset kysymykset tulivat alalle näkyvästi mukaan 1990-luvulla erityisesti sen myötä, että sosiaalietieteistä tuttu kriittinen teoria alettiin omaksua soveltavaan kielitieteeseen ja sitä myötä myös kielentestaukseen (Lynch 2001: 351). Shohamy (1993, 1997) toi alan keskusteluun testien valtaan ja ihmisten kontrollointiin liittyvät kysymykset. Testien ohjauvaikutusta (*washback*) kouluopetukseen ja yhteiskuntaan tarkastelivat esimerkiksi Alderson ja Wall (1993) sekä Wall ja Alderson (1993). Samaan aikaan alettiin hakea näkökulmia myös laskennallisesti mitattavien menetelmien ulkopuolelta. Kielitaidon arvioinnin kenttä laajeni ja muuttui tieteenalana, ja alan ammattilaisten tieto ja tietoisuus kehittyivät (Hamp-Lyons 2000: 160–162).

Kun kielitaidon arviointi kehittyi tieteenalana, muuttuivat myös käsitykset kielitesteistä ja kielitaidosta. Kielitesteissä ei enää korostettu kieliopin rakenteiden merkitystä, vaan kielitaito nähtiin viestinnällisenä ja vuorovaikutukseen perustuvana. Myös testien standardointi luotiin toteuttamaan oikeudenmukaisempaa ja tasapuolisempaa

kohtelua kaikille testiin osallistujille. Ennen standardointia yksittäiset henkilöt virkan-
sa perusteella päättivät ja toteuttivat arvioinnit. Standardoinnissa pyrittiin luomaan
samanlaiset testiolosuhteet kaikille kielitestiin osallistuville eliminoimalla testin kon-
struktion ja perusteisiin kuulumattomat tekijät. Myös arvioijien kouluttaminen sovel-
tamaan arvioinnissa käytettäviä kriteereitä ja pisteytyksiä samansuuntaisesti kuuluu
standardoidun testin oikeudenmukaisuuden vaatimuksiin. (Huhta 2014: 160–162.)

Kriittisen soveltavan kielentutkimuksen myötä kehittyi 1990-luvun lopulla kriit-
tinen kielentestaus (*Critical Language Testing*, CLT). Lynch (2001: 362) käyttää nimitys-
tä kriittinen arviointi (*critical assessment*) ja mainitsee tärkeimmäksi kehittäjäksi Elana
Shohamyn (esim. 1999). CLT voidaan tiivistää kolmeen suureen alueeseen, joita ovat
kielitestien vaikutukset yksilöön ja yhteiskuntaan, kielitesteihin liittyvät poliittiset ja
ideologiset kytkennät ja vallan diskurssi sekä kielitestien oikeudenmukaisuuteen täh-
täävät toimet. Tärkeänä alueena on ollut myös saada kielitesteihin osallistuvien ääni
kuuluviin. Kriittisestä näkökulmasta kysytään usein, kenen ja mitä ideologiaa kielitesti
ajaa. Kriittisimmillään CLT näkee testit vallan ja kontrollin välikappaleina, tahallisesti tai
tahattomasti puolueellisina, epädemokraattisina ja epäoikeudenmukaisina. Testauksen
tavoitteina pidetään rajoitteiden asettamista, opintosisältöjen kaventamista, poliittisen
agendan ajamista ja mekaanisten opetusmenetelmien ja lähestymistapojen edistämistä
(Weir 2005: 38). Shohamyn (1993a, 1993b, 1997) mukaan kielitestit palvelivat eetti-
sesti kyseenalaisia päämääriä, jotka yleensä vielä pidettiin piilossa. Testit eivät siis vält-
tämättä ole, mitä sanovat olevansa, joten niiden validius voidaan kyseenalaistaa. (Lynch
2001; Shohamy 1999, 2001b.)

Kritiikin kohteina ovat erityisesti olleet instituutioiksi kehittyneet, kaupalliset yli-
kansalliset kielitestit, joiden testitulokset ovat osallistujille merkittäviä. Kaupallisuutta
itsessään ei voida pitää ongelmana, vaan sitä, että liiketaloudelliset tavoitteet sanele-
vat testin toteutusta kielitaidon arvioinnin asiantuntemuksen kustannuksella. Suurten
ihmisryhmien testauksella voidaan sanoa olevan moraalisesti arveluttava historia,
minkä esimerkiksi Fulcher ja Davidson (2010: 273) tuovat esiin. Kriittisessä tarkastelussa
kyse on näiden rakenteiden näkyväksi tekemisestä. Näistä näkökulmista CLT ei keskity
kielitestien prosessien parantamiseen, vaan etabloituneiden testijärjestelmien ja niiden
ideologisten perusteiden paljastamiseen (McNamara 2000: 76).

Osin kriittisen näkökulman synnyttämän keskustelun vuoksi eettisyys tuotiin
1990-luvulla teemaksi myös alan tapaamisiin ja julkaisuihin. Soveltavan kielitieteen
maailmankonferenssissa (AILA) Jyväskylässä 1996 pidettiin symposium eettisestä kie-
lentestauksesta, ja vuonna 1997 eettiset kysymykset olivat pääaiheena *Language Testing*
-lehden numerossa. Tärkeintä ei tuolloin ollut niinkään määritellä, mitä pidetään eetti-
senä tai antaa eettisiä ohjeita kielitestien järjestäjille, vaan haluttiin tuoda esiin sellai-
sia eettisiä vastakkainasetteluja, joita kielentestaajat olivat kohdanneet testauksen eri
alueilla. Esimerkiksi Davies (1997: 236) nosti esiin sen, että eettinen dilemma testaajan

ja vallanpitäjän välille voi syntyä, kun kielitestejä käytetään poliittisen vallankäytön välineenä. Myös alan kansainvälisissä konferensseissa eettiset teemat ovat olleet vahvasti edustettuina. Esimerkiksi vuonna 1996 *Language Testing Research Colloquium* (LTRC) -konferenssin teemana oli oikeudenmukaisuus, ja myöhemmin myös testien vaikutukset (*impact*) ovat olleet useissa eri konferensseissa esillä.

Eettisyyttä on kielitaidon arvioinnin alalla kansainvälisissä yhteyksissä käsitelty kovinkin paljon, ja esimerkiksi Taylor (2013: 1) pitää eettistä näkökulmaa alalle vakiintuneena. Oikeudenmukaisuutta on tarkasteltu laajemmasta näkökulmasta kuin aikaisemmin. Esimerkiksi Kunnan (2000, 2004) yhdistää perinteiset validiteetin ja reliabiliteetin käsitteet testin osallistujaryhmien välisiin eroihin (*bias*). Oikeudenmukaisuuden ja validiteetin yhdistämistä tarkastelee Kane (2010). Xi (2010) on tarkastellut osallistujien välillä arvioinnissa syntyneitä eroja ja niiden vaikutuksia oikeudenmukaisuuteen.

Suomessa eettisiä teemoja on käsitelty erityisesti arvioinnin näkökulmasta ja lisäksi etupäässä kouluopetukseen sovellettuna. Arvioinnin eettisyyttä ovat käsitelleet esimerkiksi Atjonen (2007) ja Linnankylä ja Atjonen (2008) sekä opetukseen sovellettuna esimerkiksi Spoof (2008). Testiarvioinnin yhteydessä eettisiä teemoja on sivunut Tarnanen (2002, 2014), ja Ahola (2014) on lyhyesti käsitellyt eettistä näkökulmaa kielitestin tehtävänlaadinnassa. Salmenkivi (2013) on tarkastellut ylioppilastutkinnon ohjausvaikutusta eli sitä, miten tutkinto ja siinä käytettävät tehtävät vaikuttavat opetus-, opiskelu- ja oppimisprosesseihin. Aina eettisyys ei eksplisiittisesti teksteissä esiinny, vaikka ne käsittelisivät eettisesti merkittäviä asioita, kuten oikeudenmukaisuuden mittaamista tai testin vaikutuksia. Varsinaisesti standardoitujen testien etiikkaa en ole suomalaisessa alan kirjallisuudessa havainnut käsiteltävän.

3.2 Ammattilaistuminen ja ammattieettiset ohjeet

Näkyviksi teemoiksi 1990-luvulta lähtien nousivat kielitaidon arvioinnin ja kielentestauksen alalla samanaikaisesti eettisyys ja alan ammattilaistuminen ja vakiintuminen omaksi tieteenalaksi (Bachman 2000; McNamara ja Roever 2006; Spolsky 2008). Perustettiin kielitaidon arvioinnin kattojärjestöjä ja yhteenliittymiä, joissa tapaamisten ja konferenssien yhteydessä on käyty läpi testien sisältöön ja toteutukseen liittyviä kysymyksiä. Yhteenliittymien perustarkoituksena voidaan sanoa oman ammattialan yhteisten ohjeiden ja standardien tuottaminen. Suomen oloissa vaikuttavimmat järjestöt ovat olleet *International Language Testing Association* (ILTA), *Association of Language Testers in Europe* (ALTE) ja *European Association of Language Testing and Assessment* (EALTA).

ILTA (2000) tuotti jäsenilleen eettisen koodiston (*Code of Ethics*) ja toimintaohjeiston (*Guidelines for Practice*, ILTA 2007), joiden tavoitteena on määritellä hyvinä pidettäviä käytänteitä utilitarismin hengessä. Pidettiin tarpeellisena määritellä yhteiset ammatilliset normit ja säännöt, joita jokaisen jäsenen ympäri maailmaa odotetaan nou-

dattavan. Eettinen koodisto sisältää yhdeksän periaatetta, näiden lisäselvityksiä ja väärinkäytösten sanktiot. Kolme ensimmäistä periaatetta liittyvät testaajan ja testattavan väliseen suhteeseen ja testattavan ja hänen henkilötietojensa suojelemiseen. Ohjeissa esimerkiksi kielletään eri osallistujaryhmien syrjiminen ja loukkaava kohtelu. Kuusi seuraavaa periaatetta liittyvät testaajan vastuisiin sekä ammattialaa että yhteiskuntaa kohtaan. Näissä tuodaan tärkeänä esiin esimerkiksi testaajan oman ammattitaidon ylläpito ja kehittäminen. Toimintaohjeiston tavoitteena on puolestaan testiprosessin eri osien eettisesti hyväksyttävän toiminnan ohjaus. Ohjeissa painottuvat testin perusteiden tunteminen (konstruktio, validiteetti, reliabiliteetti) ja testin kehittäjän ja koostajan vastuut. Erikseen määritellään vastuut sellaisille testin järjestäjille, joiden testillä on suuri merkitys osallistujille (*high-stakes tests*) sekä mainitaan vastuut testituloksista ja erityisjärjestelyistä. Ohjeissa määritellään myös testiin osallistuvien oikeudet ja velvollisuudet. (Ks. Davies 2004: 99–104.)

Eurooppalaiset kielientestaus- ja arviointialan jäsenjärjestöt ALTE ja EALTA ovat myös tuottaneet jäsenistölleen toimintaohjeistot (ALTE 1994, 2001; EALTA 2006). ALTE:n toimintaohjeistossa määritellään sekä testin tuottajien että testitulosten käyttäjien vastuut ja erikseen laatustandardien minimivaatimukset, joilla jäsenjärjestöt arvioivat omien ja toistensa tuottamien testien laatua. (Ks. Van Avermaet ym. 2004.) EALTA:n (2006) ohjeistossa annetaan jäsenille ohjeita hyviksi kielitaidontestauksen ja arvioinnin käytänteiksi. EALTA:n missiona on ollut edistää sekä alan teoreettisten perusteiden että hyvien testaus- ja arviointikäytänteiden tuntemista.

Oikeudenmukaisuuteen on USA:ssa kehitetty oma normisto, *Standards of Educational and Psychological Testing* (viimeisimmät vuodelta 2014, useita aikaisempia), jossa on 21 erilaista oikeudenmukaisuuteen liittyvää periaatetta. Tämän ohjeiston tavoitteena on havaita ja poistaa testikonstruktioon kuulumattomia esteitä kaikkien testiin osallistuvien suorituksen maksimoimiseksi. Oikeudenmukaisuuden mittaaminen oli erityisen näkyvässä asemassa 1990-luvulla, jolloin esimerkiksi ALTE:n tapaamisissa pyrittiin kokoamaan keinoja ja menetelmiä kielitestien oikeudenmukaisuuden määrittämiseksi. (Verhelst, Banerjee ja McLain 2016: 244; ALTE:sta ks. Van Avermaet ym. 2004: 139.)

Yhteiset normit ja ohjeet ovat kohdanneet alalla sekä tunnustusta että kritiikkiä. McNamara ja Roever (2006: 146) toteavat, että koodistot ja ohjeistot ovat alan eettisen toiminnan kulmakivinä tärkeitä, koska kielitestien käyttö ja vaikutukset ovat jatkuvasti lisääntyneet. Voidaan ajatella, että ILTA:n eettinen koodisto antaa käsityksen ammattialan moraalista, ihanteista ja arvoista, kun taas toimintaohjeistot tarkastelevat enemmänkin hyvän toiminnan vaatimuksia ja antavat myös konkreettisia ohjeita väärän ja epäammattimaisen toiminnan suhteen. Taylor (2012: 3) toteaa kuitenkin, että näiden kahden ohjeistopankin välinen raja ei ole selvä, vaan monet ilmaistut asiat eettisessä koodistossa ovat lähinnä hyvän toiminnan käytännön ohjausta ja useat käytännön toi-

mintaohjeista puolestaan enemmän eettistä ohjausta. Voidaan Tayloria (2013: 3) myönteillen peräänkuuluttaa ohjeistojen päivittämistä. Käytännön ohjeistoja on myös arvosteltu siitä, että hyvänä pidettävään toimintaan ja ongelmien välttämiseen tähtäävät yleivät ja yleistettävät eettiset tavoitteet kääntyvät huonosti käytännön toimenpiteiksi (Taylor 2013: 3). (Ks. myös Davies 2004.)

Omalta osaltaan normit yhtenäistävät ammattieettistä arvomaailmaa ja toiminnan laillisuutta (Fulcher & Davidson 2010: 273–274). Ammattilaisuuden määrittäminen ei kuitenkaan ole alalla kovin tarkkaan normitettua. Esimerkiksi McNamara ja Roever (2007: 147) tähdentävät, että kielentestauksen alan eettiset koodistot ja toimintaohjeet eivät voi olla samoja kuin ns. kovien tieteiden (esim. lääketiede, psykologia), vaan eettisten kysymysten ratkaiseminen liittyy väijäämättä jäsentensä omatuntoon, joka rakentuu heidän oman moraalinsa ja eettisten periaatteidensa varaan.

4 Vastuulliseen ja oikeudenmukaiseen toimintaan pyrkimistä

Keskityn tässä luvussa sellaisiin toimiin, joita standardoiduissa kielitesteissä voidaan pitää eettisinä sekä pohdin muutamia ongelmakohtia. Keskityn ammatilliseen yhteistyöhön ja tutkimuksen merkitykseen, testien oikeudenmukaisuuden ja arvioinnin eettisiin perusteisiin liittyviin kysymyksiin sekä näkökulmaan osallistujista. Poimin myös joitakin näkökohtia kielitesteihin osallistuvien oikeuksista.

Voidaan ajatella, että eettiset kysymykset liittyvät kielitestien lähes jokaiseen valintaan. Jo itse standardointi on valinta, jolla on arvoperusta. McNamara ja Roever (2007: 146) painottavat sitä, että ottaakseen huomioon kaikki sidos- ja viiteryhmät, kielentestaajan on tiedostettava testin sosiaalinen ja poliittinen konteksti ja lisäksi taidettava sekä psykometriikan antamat mahdollisuudet testien oikeudenmukaisuuden mittaamiseen että tiedot ja teoriat vieraan kielen oppimisen kehittämisestä. Testiä on tarkasteltava entistä laajemmasta perspektiivistä, ja tarkastelussa on otettava huomioon testien seuraukset ja myös ymmärrettävä syyt, joiden vuoksi testi ylipäättään tuotetaan (Shohamy 2011: 420).

Kielentestaus voidaan käsittää **tieteenalaksi** ja **ammattilaisuudeksi**. Tieteenalana sitä ohjaavat samat eettiset ohjeet, jotka ohjaavat muitakin tieteenaloja, laajasti ottaen tutkittavien suojaamiseen ja tiedonhankintaan sekä tutkijan vastuusiin liittyvät ohjeet (Tutkimuseettinen neuvottelukunta 2013). Kielentestaus voidaan määritellä Lynchin (2001: 359) mukaisesti järjestelmälliseksi tiedon keruuksi, jonka perusteella tehdään ihmisiä koskevia päätelmiä. Tieteenalana kielitaidon arviointi on kielitestien eettisen toiminnan kehittäjä ja toimintaohjeiden luoja. Kielentestaajan ammatillinen

etiikka liittyy siis siihen, miten hyvin ammatti pohjaa käytänteensä alan tieteelliseen työhön ja millaisia laillisia velvoitteita ammattiin liittyy.

Kielitestejä ja niissä tehtävää eettistä työtä voidaan tarkastella kolmenlaisen sidonnaisuuden kautta. Toisaalta virallisia kansallisia testejä ohjaavat lait ja asetukset, joiden määrittelemiä toimia testin järjestäjän kuuluu noudattaa. Toisaalta testeillä on itsellään kirjalliset perusteet, standardeja ja spesifikaatioita, joiden avulla laaditaan testeissä käytettäviä tehtäviä ja toteutetaan arviointiprosessi. Lisäksi tulevat kansainväliset ammattikunnan ja tieteenalan sisäiset, hyväksi todetut käytänteet, joiden avulla kielitestejä auditoidaan, niiden sisältöjä validoidaan ja niiden kehittämistä tarvittaessa konsultoidaan. Vaikka testin toiminta ei perustuisikaan mihinkään pakkoon tai tiukkaan valvontaan, se on kuitenkin normitettua, ja voidaan sanoa, että ammattieettiset vaatimukset on syytä sisäistää jo ennen toiminnan aloittamista. Tätä voidaan pitää testaajan moraalisen velvoitteena.

Monet kielitestien tehtävät vaativat **asiantuntijuutta**, johon ulkopuolisilla ei yleensä ole asiaa, vaikka testien tulos olisi heitä itseään koskettava. Kielitestin sidosryhmät eivät välttämättä ymmärrä teknistä ammattikieltä tai psykometriikan menetelmiä, vaikka olisivat esimerkiksi kielenopetuksen asiantuntijoita. Ongelmaksi voi muodostua se, että ala ikään kuin erikoistuu liikaa ja sitä johtaa muista soveltavan kielentutkimuksen alueista eristäytynyt joukko asiantuntijoita. Taylorin (2013: 5) mukaan alalla vallinnut kiinnostus mittaamiseen ja testin teknisiin ominaisuuksiin sekä ammatillinen erikoistuminen ovat saaneet aikaan sen, että laajemman yleisön tietoisuutta arvioinnin periaatteista ja käytänteistä ja sen eettisistä ulottuvuuksista ei ole onnistuttu saavuttamaan. Tämä voi tarkoittaa sitä, että kielitesteihin liittyy edelleen uskomuksia, joita voi hälventää vain toiminnan mahdollisimman suurella avoimuudella. (Taylor 2013; Spolsky 2008.)

Standardoitujen kielitestien arvioinneissa syntyvän tuloksen ei pitäisi olla kenenkään arvioijan yksin päättämä kokemuksesta ja asiantuntijuudesta huolimatta. Asiantuntijuus tarkoittaa sen ymmärtämistä, että esimerkiksi arviointipäätöksiin tarvitaan tilastollisten menetelmien tuomaa tietoa ja yleensä useita eri arviointeja. Esimerkiksi Atjonen (2007: 231) painottaa kollektiivisen asiantuntijuuden merkitystä arviointipäätöksissä. Myös arvioijien kouluttaminen ja heidän toimintansa seuraaminen ovat eettisesti kestäviä keinoja arviointipäätöksissä. Institutionaalisen arvioinnin luonteeseen kuuluu systemaattinen arvioinnin luotettavuuden varmistaminen esimerkiksi tarkastelemalla arvioijien johdonmukaisuutta ja yhdenmukaisuutta. (Tarnanen 2014: 116.)

Alalla **yhteisesti hyväksytyt standardit** ja toimintaohjeet voivat helpottaa kielentestaajaa ymmärtämään, missä eettisesti hyvinä pidettyjen toimintojen rajat kulkevat. Vaikka tällaiset ohjeet on tarkoitettu kaikille, niiden soveltaminen voi kuitenkin olla hankalaa. Alan eettinen koodisto on kirjoitettu edistämään eettisyyttä ympäri maail-

maa, mutta sitä on noudatettava kunkin lain sallimissa puitteissa (Davies 2004: 103). Näin ollen, vaikka yhdessä laadittuja kansainvälisiä koodistoja ja toimintaohjeistoja on hyvä pitää perustana, olisi paikallisten eettisten ohjeiden pohtiminen tärkeää kielitestiä toimintaa ja toiminnan ulottuvuuksia määriteltäessä. Erityisen tärkeää se on siksi, että kielitestiä määrä ja käyttötarkoitukset näyttävät lisääntyvän. Arviointikiinnostuksen kasvun myötä myös eettiset kysymykset tuntuvat voimistuvan (Atjonen 2007: 10). Tärkeää olisi esimerkiksi kartoittaa eettisiä haasteita, joita teknologian jatkuva kehittyminen, testiä digitalisointi ja lisääntynyt käyttö esimerkiksi kansalaisuuden määrittelyssä tuovat mukanaan.

Yhteisesti hyväksytyt kansainväliset ja kansalliset standardit auttavat kielentestaa tekemään ratkaisuja näissä eettisissä valintoja koskevissa kysymyksissä. Tällaisia voivat olla esimerkiksi kysymykset mittaamisen käytetyistä työkaluista tai arvioinnissa käytetyistä kriteereistä. Yksi laajalti yhteisessä käytössä oleva työkalu on ollut Eurooppalainen viitekehys (EVK 2003), joka omalta osaltaan on yhtenäistänyt varsin kirjavaa kielentestauksen käytäntöä erityisesti Suomessa ja muualla Euroopassa. Viitekehys on Suomessa ollut käytössä sekä opetuksessa, opetussuunnittelussa että kielitesteissä jo varsin kauan, joten sillä on tuttuusarvoa. Vaikka sen käyttö ei ole ongelmaton, omalta osaltaan se auttaa osallistujia ymmärtämään oman kielenoppimisen kehittymistä. Testin tulosta voi helpommin käyttää oppimisprosessin tukena, kun testin käyttämät arviointikriteerit ja taustalla olevat arvioinnin perusteet ovat testattavalle tuttuja.

On selvää, että standardoituihin kielitesteihin kytkeytyvät valta ja politiikka (Van Avermaet ym. 2004: 150). Näin on esimerkiksi silloin, kun kielitestiä tuottaja toimii poliittisten päättäjien ja muiden valtaapitävien heille suomalla mandaatilla. Yksilöiden ja instituutioiden toimintaa rajoittavat säännöt vaativat laajamittaista sosiaalista yhteistyötä, että pystyttäisiin saavuttamaan molemminpuoliset hyödyt ja edut (Taylor 2013: 5). Tämä tarkoittaa sosiaalisen vastuunkannon näkökulmaa, joka ei valitettavasti aina toteudu, vaan testejä usein toteutetaan ainoastaan yhteiskunnan vaatimusten vuoksi. Fulcher ja Davidson (2015: 188) peräänkuuluttavat testiä sosiaalisten tavoitteiden huomioonottamista. On pystyttävä kantamaan vastuu kielitestiä käytänteiden vaikutuksesta sekä ennakoimaan ja reagoimaan tulosten poliittiseen käyttöön ja väärinkäyttöön (Hamp-Lyons 2001: 117). Eettisestä näkökulmasta esiin nousevat myös esimerkiksi kysymykset siitä, kenen ja mitä näkemystä kielestä, kielitaidosta, testiä sisällöistä ja testissä tehtävästä toiminnasta noudatetaan (Shohamy 2009: 56). Kielitestiä keskittyvät yhteen kieleen kerrallaan, mitä monikielisessä maailmassa voidaan pitää ongelmana, mutta testauksen näkökulmasta vaikeasti ratkaistavana.

Yksi keskeinen toimenpide kielitesteissä on **tutkimuksen tekeminen ja lisääminen** niistä alueista, joita eettiset ja oikeudenmukaisuuteen liittyvät kysymykset nostavat esille. Esimerkiksi Xi (2010: 148) on sitä mieltä, että empiiristä tutkimusta oikeudenmukaisuudesta on liian vähän. Tämän hän arvelee johtuvan siitä, että aikaisemmat

tutkimukset ovat kyllä tuoneet esille tutkimuksen kohteita, mutta eivät käytännön ohjausta siitä, miten eri paradigmojen näkökulmia oikeudenmukaisuuteen yhdistetään. Oikeudenmukaisuuden varmistamiseksi testeissä tarkastellaan sitä, onko eri osallistujaryhmien välillä arvioinnissa syntyneitä eroja. Tavoitteena on varmistaa, että osallistujia ei syrjitä esimerkiksi heidän sukupuolensa, ikänsä, etnisen taustansa tai ensikielensä perusteella. Xi (2010: 162) mainitsee useita eron aiheuttajia, joista testiryhmien välille voi syntyä eturistiriitaa. Ristiriitoja voivat aiheuttaa esimerkiksi epäsovivat ja epäjohdonmukaiset testikäytänteet, huonot mahdollisuudet päästä testiin sekä testin käsitemateriaaliin ja perusteisiin sopimattomat tehtävät, tehtävänaiheet ja sisällöt. Esimerkiksi arvioinnissa voidaan suosia jotain ryhmää vaikkapa käsialan tai ääntämisessä kuuluvan lähtökielen perusteella.

Alalla useimmin käytetyt oikeudenmukaisuuteen tähtäävät analyysikeinot kulkevat nimellä *Differential Item Functioning* (DIF). DIF-analyysit antavat esimerkiksi tuloksen siitä, vääristyvätkö testin osiot vaikkapa iän mukaan. Mikäli tuloksissa havaitaan poikkeamia, jotka voidaan liittää testiin kuulumattomina pidettyihin tekijöihin, voidaan eriarvoisuutta tuottavat osiot poistaa. Tässä on kuitenkin ongelmana se, että testiin valituista tehtävistä on vaikea poistaa osioita vaurioittamatta testin kokonaisuutta. Analyysit sopivatkin siten parhaiten testeihin, joiden osiot voidaan esitellä. Standardoiduissa kielitesteissä voi esitellä hankalaa esimerkiksi esitestiryhmien vähäisen määrän vuoksi. Arvioinnin vääristymien ja poikkeamien sekä oikeudenmukaisuuden suhdetta pidetäänkin alalla ongelmallisena, koska yhteisesti hyväksytyjä keinoja ei vielä ole (Verhelst, Banerjee & McLain 2016: 241).

Eettiset ratkaisut standardoiduissa kielitesteissä liittyvät siihen, miten laajana validointi ja siihen liittyvät toimenpiteet ymmärretään. Kielitestien keskiössä on itse arviointi, jonka tarkasteluun on syytä mukauttaa sekä mitattavia että laadullisia menetelmiä (Tarnanen 2014: 121). Validointi itsessään ei voi olla vain luottamista reliabiliteettiin tai siihen, että kaikki psykometrisesti toteutettavissa oleva riittää oikeudenmukaisuuden määrittämiseksi. Esimerkiksi Fulcher (2015: 172) toteaa, että mittaukseen ja mittaustulokseen käytettäviin menetelmiin ei pitäisi liian dogmaattisesti uskoa, vaan olisi tarkasteltava kaikkia näkökohtia, jotka vaikuttavat testituloksen syntymiseen ja tuloksista tehtäviin tulkintoihin. On tärkeä tarkastella paitsi testin kehittämisen, kokoamisen, tehtävien laatimisen ja arvioinnin prosesseja myös testien todellisia ja mahdollisia seurauksia. Tarvitaan monenlaista empiiristä näyttöä, jotta testituloksen ja siitä tehtävien tulkintojen voidaan sanoa olevan niin oikeudenmukaisia kuin mahdollista.

Eettisestä näkökulmasta on olennaista ottaa huomioon **testiin osallistujat**. Esimerkiksi Linnankylä ja Atjonen (2008: 49) toteavat, että arvioitaville on etukäteen tiedotettava arvioinnin tarkoitus ja mahdolliset seuraukset sekä esitettävät arvioinnissa käytettävät kriteerit. Sosiaaliseen oikeudenmukaisuuteen kuuluu se, että testattava tietää kielitestin perusteet ja että nämä ovat avoimesti saatavilla ja ymmärrettävissä

muodossa. Arvioitavien oikeuteen luetaan kuuluvaksi myös palaute, jonka avulla testiin osallistuva voi paitsi tarkastella kielitaitonsa kehittymistä, myös liittää testituloksen omaan oppimisprosessiinsa (Atjonen 2007: 228). Standardoidussa kielitestissä testaa-ajan rooli on yleensä toiminnan kohteesta etäinen ja siinä mielessä esimerkiksi opettajan suorittamasta arvioinnista poikkeava, että toiminta ei ole läheisessä yhteydessä arvioitavan kohteen kanssa. Etäisyys mahdollistaa testattavan näkymättömyyden, joten testaa-ajan rooliksi jää tukeutuminen testin konstruktion kielenoppijan omien näkemysten kustannuksella. Kielitestin on siis syytä käyttää sellaisia tehtäviä ja testimuotoja, jotka ovat oppijalle merkityksellisiä ja ymmärrettäviä. Tärkeänä voidaan pitää tehtävien kiinnostavuutta ja mielekkyyttä (Ahola 2014: 64). (Ks. Shohamy 2009: 57.)

Hamp-Lyons (2001: 124) muistuttaa, että eettiset näkökohdat täytyy ottaa huomioon esimerkiksi silloin, kun teknologia otetaan mukaan kielitesteihin. Teknologia on erinomainen apu ja sitä on mahdotonta sivuuttaa nykypäivän testauksessa, mutta sen avulla on mahdollista myös luoda lisää etäisyyttä arvioitavaan. Vaikka ei ole eettisesti oikein testata läheistä sukulaista tai ystävää, liian suuri etäisyys arvioitavaan saattaa aiheuttaa sen, että on helppo unohtaa, että arvioitavat kohteet ovat ihmisiä, joiden elämään arvioinnilla on aina jokin merkitys, joskus hyvinkin ratkaiseva. Testiin osallistuvien lisäksi on myös hyvä saada kuuluviin kaikkien niiden äänet, joita voidaan laskea kielitestien sidosryhmiin kuuluviksi. Sidoryhmistä esimerkiksi Hamp-Lyons (1997: 299) mainitsee kielenoppijat, heidän sukulaisensa, opettajat, hallitukset, viralliset lautakunnat ja kielitestejä järjestävät tahot. Testattavat ovat kielitestien suurin ja tärkein sidosryhmä. He panostavat testiin eniten ja heillä on eniten menetettävää ja saavutettavaa. Vaikka he ovat tärkein joukko, näyttää edelleen siltä, että heidän äänensä ei kuulu oikeudenmukaisuuden ja eettisyyden kysymyksissä. (Ks. Hamp-Lyons 2000.)

Yksi keskeinen alue oikeudenmukaisuudessa on se, miten **erityisryhmien** (esim. vammaiset) **tarpeet** otetaan testissä huomioon. Sosiaalisen vastuunkannon ja ihmisoi-keuksien näkökulmasta tämä on tärkeä alue. Kielitesteissä erityisjärjestelyt tarkoittavat esimerkiksi testirutiinien mukauttamista sekä testisisältöjen muokkaamista. Tavoitteena on, että testistä vähennetään esimerkiksi vamman laadun perusteella mahdollisimman paljon sellaista, joka ei olennaisesti kuulu testissä arvioitavan kyvyn luonteeseen. Vammaisuuden tai vamman ei kuulu estää kielitestiin osallistumista. Usein mukauttamiset ovat tapauskohtaisia ja perustuvat pikaisiinkin ammattilaispäätöksiin ja käytännön sallimiin mahdollisuuksiin. Erityisjärjestelyjen vaikutuksista ei kuitenkaan ole tarpeeksi tutkimustietoa (Taylor 2013: 5).

Kielitestejä on entistä enemmän Suomessa ja maailmanlaajuisesti, ja niiden parissa työskentelee entistä useampi henkilö. Testejä myös kehitetään yhä uusiin tarpeisiin – meillä esimerkiksi mittaamaan hoitohenkilökunnan, sotilaiden ja maahanmuuttajien kielitaitoa. Spolsky (2008: 297) pitää kielentestausta laajamittaisena liiketoimintana, jonka vaikutukset ulottuvat laajoille alueille koulutuksessa, työelämässä ja

koko yhteiskunnassa. (Ks. myös Shohamy 2008: xiv.) Tämä tuo mukanaan lisää eettisiä haasteita sekä ammattikunnalle, koulutukselle että yhteiskunnalle. Alaa on pystyttävä kehittämään, ja alalta saatava tutkimustieto ja eettisiä kysymyksiä koskeva tieto on hyvä saada yhteiskunnassa laajemman yleisön tietoon.

5 Päätäntö

Eettiset valinnat voidaan liittää laajalle alueelle standardoitujen kielitestien toiminnassa alkaen kieleen ja kielitaiton liittyvistä kysymyksistä. Koska eettisyys on kovin laaja käsite, olen pyrkinyt keskittymään niihin näkökulmiin, joita itse pidän tärkeinä erityisesti testien toimivuuden näkökulmasta. Tämä vuoksi monia keskeisiä eettisesti merkittäviä asioita jää käsittelemättä. En ole tuonut esiin testin mahdollisen näennäisvaliditeetin (*face validity*) kysymyksiä, jotka voidaan hyvin nostaa eettiseen keskusteluun. Myös testien vaikutukset ja niistä tehty tutkimus jäävät melko vähäiselle käsittelylle ja yleiselle tasolle samoin kuin itse testin kohde, osallistuja. Lisäksi oman käsittelynsä tarvinnevat vaihtoehtoinen, joustava ja oppijakeskeinen arviointi. Lisäkeskustelu standardoitujen kielitestien tulevaisuudesta ja testien mahdollisen roolin muuttumisesta olisi mielestäni paikallaan myös eettisestä näkökulmasta.

Eettisesti hyväksyttävän toiminnan tausta-ajatuksena voidaan pitää sitä, että kielitesteissä toimitaan hyväntahtoisesti ja myötätuntoisesti vahingoittamatta toista ihmistä. Oikeudenmukaisuuteen tähtäävien toimien lisäksi on hyvä pitää mielessä testin kehittäminen ja tutkiminen. Tiedon ja tietoisuuden lisääminen on osa eettisyyttä edistävää toimintaa. Testeissä ei aina voida toimia kaikkia tyydyttävällä tavalla, mutta eettisesti on riittävää, että tavoittelemme oikeudenmukaisia ratkaisuja kaikkia käytössä olevia keinoja hyödyntäen.

Kirjallisuus

- Ahola, S. 2014. Tehtävien elinkaari yleisissä kielitutkinnoissa. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. : Opetushallitus, 63–71.
- Alderson, J. C. & D. Wall. 1993: Does washback exist? *Applied Linguistics* 14, 115–29.
- ALTE, 2001. Toimintaohjeisto. Saatavilla osoitteessa http://www.alte.org/attachments/files/code_practice_fi.pdf [luettu 5.11.2016].
- American Educational Research Association, American Psychological Association & The National Council of Measurement in Education, 2014. *Standards of Educational and Psychological Testing*. Saatavilla osoitteessa <http://www.apa.org/science/programs/testing/standards.aspx> [luettu 13.1.2017].
- Atjonen, P. 2007. *Hyvä, paha arviointi*. Helsinki: Tammi.
- Bachman, L. F. 2000. Modern language testing at the turn of the century: assuring that what we count counts. *Language Testing* 17 (1), 1–42. DOI:10.1177/026553220001700101.

- Bachman, L.F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Canale, M. 1987. The measurement of communicative competence. *Annual Review of Applied Linguistics*, 8, 67–84. DOI: 10.1017/S0267190500001033.
- Davies, A. 2004. Introduction: Language testing and the golden rule. *Language Assessment Quarterly*, 1 (2 & 3), 97–107. DOI:10.1080/15434303.2004.9671778.
- Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley & T. McNamara. 1999. *Dictionary of language testing*. Cambridge: Cambridge University Press.
- EVK 2003. *Eurooppalainen viitekehys: kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys*. Helsinki: WSOY.
- Fulcher, G. 2015. *Re-examining language testing: a philosophical and social inquiry*. London & New York: Routledge.
- Fulcher, G. & F. Davidson. 2010. *Language testing and assessment: an advanced resource book*. London & New York: Routledge.
- Hamp-Lyons, L. 2001. Fourth generation writing assessment. Teoksessa T. Silva & P. K. Matsuda (toim.) *On second language writing*. Mahwah, NJ: Lawrence Erlbaum Associates, 117–127.
- Hamp-Lyons, L. 2000. Social, professional and individual responsibility in language testing. *System* 28 (4), 579–591.
- Hamp-Lyons, L. 1997. Washback, impact and validity: ethical concerns. *Language Testing*, 14 (3), 295–303. DOI: 10.1177/026553229701400306.
- Huhta, A. 2014. Yleiset kielitutkinnot kansainvälisessä testauskentässä. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen. *Yleiset kielitutkinnot 20 vuotta*. Helsinki: Opetushallitus, 159–168.
- Kane, M. 2010. Validity and fairness. *Language Testing*, 27 (2), 177–82.
- Kunnan, A. J. 2004. Test fairness. Teoksessa M. Milanovic & C. Weir (toim.) *European Language Testing in a Global Context*. Cambridge: Cambridge University Press, 27–48.
- Kunnan, A.J. 2000. *Fairness and validation in language assessment*. Cambridge: Cambridge University Press.
- Linnankylä, P. & P. Atjonen. 2008. Arvioinnin eettisyys. Teoksessa E. Korkeakoski & H. Silvennoinen (toim.) *Avaimia koulutuksen arvioinnin kehittämiseen*. Koulutuksen arviointineuvoston julkaisuja 31. Jyväskylä, 47–64.
- Lynch, B. 2001. Rethinking assessment from a critical perspective. *Language Testing*, 18 (4), 351–372. DOI: 10.1177/026553220101800403
- McNamara, T. & C. Roever. 2006. *Language testing: the social dimension*. Malden, MA: Blackwell Publishing.
- McNamara, T. 2000. *Language testing*. Oxford: Oxford University Press.
- Messick, S. 1989a. Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, 18 (2), 5–11.
- Messick, S. 1989b. Validity. Teoksessa R. Linn (toim.) *Educational measurement*. New York, NY: Macmillan, 13–103.
- Messick, S. 1981. Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10 (9), 9–20.
- Messick, S. 1980. Test validity and the ethics of assessment. *American Psychologist*, 35 (11), 1012–1027. DOI: 10.1002/j.2333-8504.1981.tb01244.x.
- Salmenkivi, E. 2013. Ylioppilastutkinnon rakenne- ja reaaliokoeuudistusten vaikutuksia: miten lisääntynyt valinnaisuus ohjaa lukiolaisia. *Kasvatus & aika* 3/2013. Saatavilla osoitteessa: http://www.kasvatus-ja-aika.fi/site/?page_id=565. [luettu 13.11.2016].

- Shohamy, E. 2011. Assessing multilingual competencies: adopting construct valid assessment policies. *The Modern Language Journal*, 95 (3), 418–429. DOI: 10.1111/j.1540-4781.2011.01210.x 0026-7902/11/.
- Shohamy, E. 2009. Language tests for immigrants. Why language? Why tests? Why citizenship? Teoksessa G. Hogan-Brun, C. Mar-Molinero & P. Stevenson (toim.) *Discourses on language and integration*. Amsterdam, Philadelphia: John Benjamins, 45–60.
- Shohamy, E. 2001a. Democratic assessment as an alternative. *Language Testing*, 18 (4), 373–345.
- Shohamy, E. 2001b. *The power of tests: a critical perspective on the uses of language tests*. Harlow, England: Pearson Education.
- Shohamy, E. 1999. Critical language testing: uses and consequences of tests, responsibilities of testers and rights of test-takers. Paper presented at the 21st Annual Language Testing Research Colloquium, Tsukuba, Japan.
- Shohamy, E. 1997. Testing methods, testing consequences: are they ethical? *Language Testing*, 14 (4), 340–49.
- Shohamy, E. 1993a. The exercise of power and control in the rhetorics of testing. Teoksessa A. Huhta, K. Sajavaara & S. Takala. (toim.) *Language testing: new openings*. Jyväskylä: Koulutuksen tutkimuslaitos, 23–38.
- Spolsky, B. 2008. Introduction – Language Testing at 25: maturity and responsibility? *Language Testing*, 25 (3), 297–305.
- Spolsky, B. 1981. Some ethical questions about language testing. Teoksessa C. Klein-Braley & D. K. Stevenson (toim.) *Practice and problems in language testing*. Frankfurt: Peter Lang, 5–21.
- Spoof, M. 2007. ”Pintaa syvemmälle”: eettinen pohdinta opettajan voimavarana luokanopettajan ja erityisopettajan silmin. Tutkimuksia 277. Helsingin yliopisto. Saatavilla osoitteessa: <http://hdl.handle.net/10138/20028>.
- Stevenson, D. K. 1981. Language testing and academic accountability: on redefining the role of language testing in language teaching. *International Review of Applied Linguistics*, 19, 15–30.
- Tarnanen, M. 2014. Arvioija taidon arvottajana. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Helsinki: Opetushallitus, 115–125.
- Tarnanen, M. 2002. *Arvioija valokeilassa: suomi toisena kielenä arviointia*. Jyväskylä: Soveltavan kielentutkimuksen keskus.
- Taylor, L. 2013. Ethics in language assessment. Teoksessa C. A. Chapelle (toim.) *The Encyclopedia of applied linguistics*. Blackwell, 1–7. DOI:10.1002/9781405198431.wbeal0393/pdf.
- Tutkimuseettinen neuvottelukunta 2012. *Hyvä tieteellinen käytäntö ja sen loukkausepäilyjen käsitleminen Suomessa*. Saatavilla osoitteessa http://www.tenk.fi/sites/tenk.fi/files/HTK_ohje_2012.pdf [luettu 7.12.2016].
- Van Avermaet, P., H. Kuijper & N. Saville. 2004. A code of practice and quality management system for international language examinations. *Language Assessment Quarterly* 1: 2–3, 137–150. DOI: 10.1080/15434303.2004.967178.
- Verhelst, N., J. Banerjee & P. McLain. 2016. Fairness and bias in language assessment. Teoksessa D. Tsagari & J. Banerjee (toim.) *Contemporary second language assessment*. Contemporary Applied Linguistics 4. London: Bloomsbury, 243–260.
- Wall, D. & J.C. Alderson. 1993. Examining washback: the Sri Lankan impact study. *Language Testing*, 10 (1), 41–69.
- Weir, C. J. 2005. *Language Testing and Validation: an Evidence-based Approach*. New York: Palgrave Macmillan.
- Xi, X. 2010. How do we go about investigating test fairness? *Language Testing*, 27 (2), 147–70. Saatavilla osoitteessa <http://ltj.sagepub.com/cgi/content/abstract/27/2/147>.

Verkkolähteet

EALTA 2016. Ealtan ohjeisto hyväksi kielitaidon testauksen ja arvioinnin käytänteiksi. 2006 <http://www.ealta.eu.org/documents/archive/guidelines/Finnish.pdf>.

ILTA 2000. *Code of Ethics; Guidelines for Practice*. <http://www.iltaonline.com/>.

Tutkimuseettinen neuvottelukunta 2013. *Hyvä tieteellinen käytäntö*. <http://www.tenk.fi/fi/htk-ohje/hyva-tieteellinen-kaytanto>.