

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 89–109.

Sari Ahola

Jyväskylän yliopisto

Puhetta arvioinnista – Yleisten kielitutkintojen arvioijien käsityksiä arvioinnista

This paper focuses on raters' beliefs and views of assessment and the factors influencing their judgments and performance as raters in the National Certificates of Language Proficiency (NCLP) Finnish language test. Data were collected using semi-structured interviews (N=18) and content analysis was used to classify the responses. Findings suggest that raters' experience and training have positive effects on rater performance and the use of assessment criteria. Experienced raters seem to be more familiar with different language proficiency levels and their view of language is more holistic today than during the first years of their rating experience. There are some external factors and linguistic features that affect raters' performance, such as pronunciation, handwriting and the test-taker's personality. In these cases, self-reflection helps raters to strive for objectivity in their evaluations.

Keywords: language testing, raters, assessment, rater training, rater experience

Asiasanat: kielitaidon arviointi, arvioijat, arvioijakoulutus, arviointikokemus

1 Johdanto

Artikkelin aineisto ja sisältö liittyvät jatkotutkimukseeni, jossa tutkin arvioijien käsityksiä kielitaidosta ja arvioinnista. Esittelen artikkelissa, miten arvioijina toimivien käsitykset omasta arviointilinjasta ja arviointi-identiteetistä syntyvät ja mikä merkitys arvioijan taustatekijöillä (arviointi- ja työkokemus) on niiden syntymiselle. Lisäksi tarkastelen sitä, mikä merkitys arviointikoulutuksella ja arvioinnissa käytettävillä kriteereillä on arviointiprosessissa ja mitkä arvioinnin ”ulkopuoliset tekijät” vaikuttavat arviointiin. Tutkimuksen tarkoituksena on tuottaa lisätietoa arviointiin vaikuttavista tekijöistä ja arvioijien toiminnasta. Tietoa voidaan käyttää hyväksi kehitettäessä Yleisten kielitutkintojen arviointiprosessia ja arvioijien koulutusta, sillä arvioijat ovat keskeisiä toimijoita testiarvioinnissa määriteltäessä testiin osallistuneiden kielitaidon tasoa.

Käsitykset arvioinnista ovat usein kiinteästi yhteydessä omiin arviointikokemuksiimme, sillä meillä jokaisella on kokemusta siitä, kuinka on tullut arvioiduksi tai kuinka itse on arvioinut jotain henkilöä tai tapahtumaa. Tyypillisimmin arviointi yhdistetään kuitenkin opetustyötä tekeviin, joiden työhön usein olennaisena osana kuuluu arviointi. Tässä artikkelissa tarkastelen lähemmin valtakunnallisen kielitutkinnon, Yleisten kielitutkintojen, suomen kielen arvioijia ja heidän käsityksiään arvioinnista, sillä testijärjestelmässä toimivilla arvioijilla on kielitaidon arvioinnista varsin laaja kokemus. Arvioijat toimivat päivätyössään opettajina, joten opetukseen liittyvä arviointi on osa heidän jokapäiväistä työtänsä, ja Yleisten kielitutkintojen arvioijina he taas ovat saaneet kokemusta testiarvioinnista. Näissä kahdessa arviointitapahtumassa on paljon yhteistä, mutta niissä on myös eroavaisuuksia. Suurin ero liittyy siihen, että testiarviointi on kertaluonteinen tapahtuma, jossa kielenkäyttäjän taidoista tehdään päätelmiä valikoidujen testitehtävien avulla, kun taas opetukseen liittyvässä arvioinnissa opettajalla on tavallisesti mahdollisuuksia seurata arvioitavan taitoja jatkuvasti erilaisin menetelmin. (Esim. Tarnanen 2014.)

Koska testiarvioinnissa arvio perustuu suhteellisen pieneen näytteeseen, vaaditaan arvioijalta hyvää näkemystä kielitaidon kehittymisestä ja perehtyneisyyttä käytettävissä oleviin arviointikriteereihin sekä pyrkimystä objektiivisuuteen erilaisista arvioitavista huolimatta. Pyrkimystä objektiivisuuteen toki helpottaa se, että arvioijat eivät tunne arvioitavia, mutta se ei kuitenkaan tarkoita sitä, että testiarviointi on täysin objektiivista arviointia. Arvioinnissa on aina ripaus subjektiivisuutta, koska kyse on ihmillisestä toiminnasta, jossa arvioijan omat tuntemukset, mieltymykset, kokemukset ja käsitykset vaikuttavat arvioinnin taustalla. Näiden tekijöiden havainnointi arviointiprosessin aikana on kuitenkin vaikeaa sillä se, mitä arvioija ajattelee arviointihetkellä ja minkä asioiden perusteella hän tekee kielitaidosta päätelmiä, eivät välttämättä ole tiedostettuja arvioijalle itsellekään.

2 Tutkimusaineisto ja tutkimuskysymykset

Artikkelissa käyttämäni aineisto on kerätty vuosina 2015–2016, ja se koostuu 19 pseudonyymein nimetyn arvioijan avoimesta teemahaastattelusta. Haastattelujen runkona toimivat ennalta mietityt kysymykset, mutta haastatteluissa edettiin kuitenkin pääosin sen mukaan, mistä asioista haastateltavalla itsellään oli sanottavaa. (Ks. teemahaastattelu: Metsämuuronen 2005; Hirsjärvi & Hurme 1985.) Haastatellut arvioijat toimivat aktiivisesti Yleisten kielitutkintojen suomen kielen arvioijina, ja he kaikki ovat opettaneet tai opettavat aikuisille kielenoppijoille suomea toisena kielenä. Tutkimukseen valittujen arvioijien opiskelutausta oli varsin samankaltainen, sillä heillä kaikilla oli opintoja suomen kielestä joko äidinkielenä tai/ja toisena ja vieraana kielenä. Muissa taustamuutujissa oli eroja, sillä aineiston keruussa huomioitiin se, että haastateltaviksi valikoitui eri-ikäisiä (30–70-vuotiaita) ja eri koulutusasteilla (kotoutumiskoulutus, aikuiskoulutus, toinen aste ja korkea-aste) toimivia opettajia, joilla oli eripituinen kokemus ajallisesti sekä opettamisesta (7–yli 25 vuotta) että Yleisten kielitutkintojen arvioijana toimimisesta (1–20 vuotta). Tällä tavoin saatiin mahdollisimman kattava otos erilaisista Yleisten kielitutkintojen suomen kielen arvioijista, joita tällä hetkellä Opetushallituksen ylläpitämässä arvioijarekisterissä on yli 100. Arviointityö edellyttää aina Opetushallituksen arvioijakoulutuksen suorittamista ja rekisteröintiä, sillä arvioijana toimiminen ja arvioijalle edellytetyt vaatimukset on säädetty asetuksella (1163/2004). (Ks. Ahola & Tossavainen 2014.)

Litteroitujen haastattelujen analysoinnissa käytettiin laadullisten aineistojen systemaattista analyysia tukevaa Atlas.ti-ohjelmaa, jonka avulla aineisto tyypiteltiin ja luokiteltiin sisällön perusteella useisiin eri luokkiin (esim. arvioijakokemus, työkokemus, arvioitavan persoonalliset piirteet, arviointikriteerit, arviointikoulutus, arviointilinjat). Näiden luokkien perusteella aineistosta saatiin eroteltua arvioijien arviointiin liittyvät käsitykset. Lähtökohtana analyysissa oli teoriaohjaava sisällönanalyysi, sillä haastatteluaineiston rinnalla analyysia ohjasivat olemassa oleva tutkimustieto arvioinnista ja arvioijana toimimisesta. (Ks. sisällönanalyysista: Tuomi & Sarajärvi 2004.)

3 Käsitykset omasta arviointilinjasta ja arvioija-identiteetistä

Arvioinnissa ankaruus ja lempeys voidaan kuvata janana, jonka toisessa ääripäässä on ankara arvioija ja vastakkaisessa päässä lempeä arvioija. Siihen, kuinka ankarana tai lempeänä arviointia tekevä pitää itseään ja mihin kohtaan hän itsensä tällä janalla sijoittaa, ei välttämättä ole olemassa yhtä syytä, sillä käsitykset omasta ankaruudesta/lempeydestä muodostuvat tavallisesti useiden eri tekijöiden lopputuloksena. Yhtenä

tekijänä esimerkiksi voivat olla muiden ihmisten arviot ja käsitykset henkilön ankaruudesta/lempeydestä.

- (1) Mä oon kuullut joskus, että mä olisin lempeä. Mä ehkä sanoisin, että mä olen lempeämpi. (Erika)

Arvioijien käsitys omasta arviointilinjastaan muodostui haastattelujen perusteella tavallisimmin kahdella eri tavalla. Käsitys itsestään lempeänä tai ankarana arvioijana oli lähtöisin usein joko oman arviointilinjan vertailuista muiden arvioijien linjaan tai testi-järjestelmästä saadun henkilökohtaisen arvioijapalautteen pohjalta tehdyistä päätelmistä. Tyypillisimmin oman arviointilinjan vertailu toisten arvioijien linjaan tapahtui ennen varsinaista arviointia pidettävässä arviointikoulutuksessa, jossa eritasoisten koulutusnäytteiden avulla pyritään löytämään yhdenmukainen käsitys arviointikriteereistä, tehtävien tulkinnasta ja koulutusnäytteinä olevista suorituksista. Arviointikoulutuksessa käydyn keskustelun aikana arvioijat kertoivat vertailevansa omia arvioitaan muiden arvioijien antamiin arvioihin.

- (2) Sanosinko, että ehkä mä olen löytänyt itsestäni joskus, kun ollaan oltu niissä keskusteluissa, että mä oon ollut pikkusen jopa lepsuhko, että se on pikkusen niinku menny sinne puolelle. (Heljä)

Koulutusten aikana tehdyt päätelmät eivät välttämättä kuitenkaan kerro totuutta todellisesta arviointilinjasta, vaan testijärjestelmän näkökulmasta tarvitaan luotettavampaa tilastollista tietoa arvioijien toiminnasta. Tästä syystä arviointilinjaa ja johdonmukaisuutta tarkkaillaan joka arviointikerran jälkeen kaksoisarviointien kautta¹. Tilastolliset analyysit edellyttävät aina tietyn määrän arvioituja suorituksia ja myös tietyn määrän muita arvioijia, joihin yksittäistä arvioijaa voidaan verrata. Tällaisten psykometriikassa kehitettyjen IRT (*Item Response Theory*) -pohjaisten mallien avulla voidaan tarkastella testin kannalta kolmea keskeistä ulottuvuutta samanaikaisesti: tehtäväosoiden vaikeustasoa ja suorittajien osaamisen tasoa sekä arvioijien ankaruutta ja lempeyttä. Mallin kautta syntyneen numeerisen tiedon avulla arvioijille annetaan palautetta heidän henkilökohtaisesta arviointilinjastaan. Malli mahdollistaa lisäksi arvioijien vertailun keskenään, ja siten saadaan myös tietoa siitä, kuinka arvioijat sijoittuvat suhteessa toisiinsa asteikolla lempeä–ankara. (Esim. Raschin malli: Wright & Stone 1979; Törmäkangas & Törmäkangas 2009.) Näihin tilastollisiin tietoihin myös haastatellut arvioijat usein viittasivat määritellessään omaa arviointilinjaansa.

1 Kaksi arvioijaa arvioi saman suorituksen. Arvioijat linkittyvät toisiinsa kaksoisarvion avulla mikä mahdollistaa heidän vertailunsa.

- (3) Mä olin sieltä tiukimmasta päästä arvioijia. Mä olin aivan hämmästynyt, koska mä olen mielestäni lempeä ja kiltti ja lepsu. Nyt mä oon sitten ruvennut omaksumaän sitä tiukan roolia. (Anneli)
- (4) Tietenkin pitää olla selkee linja, ja mähän oon sieltä yksi-arvioijista ankarammalla puolella. Mä aina pelkään sitä, että mä olen maailman tiukkapipoisin enkä ikinä päästäis koskaan mistään läpi, sillä ainahan siellä on jotain pientä viilattavaa, mutt sitten mä yritän hillitä sitä. (Maarit)

Tilastollinen tieto antaa testijärjestelmälle kvantitatiivista tietoa arvioijien toiminnasta, mutta testijärjestelmää kehitettäessä tarvitaan myös laadullista tietoa siitä, millaisina arvioijina arvioijat pitävät itseään ja mitkä asiat vaikuttavat heidän käsityksiinsä arvioinnista. Seuraavaksi tarkastelen tarkemmin muutamia arvioijien haastattelussa esiin noutamia tekijöitä, joilla on heidän mielestään ollut vaikutusta arvioijana kehittymiseen, oman arvioijaidentiteetin syntymiseen ja arviointilinjaan.

4 Arviointi- ja työkokemus tukemassa arvioijaksi kasvamista

Kokemukset arvioinnista ovat yhteydessä siihen, millainen arviointilinja kunkin arvioijan kohdalle muodostuu (Shahomy, Gordon & Kraemer 1992). Kokemus nähdään yhtenä arvioijaa luonnehtivana tekijänä, mutta se voidaan määritellä kahdella eri tavalla. Kokeneena voidaan pitää joko sellaista arvioijaa, joka on toiminut ajallisesti kauan arvioijina, tai sellaista arvioijaa, joka on määrällisesti arvioinut paljon (Lim 2011). Tämän määritelmän perusteella Yleisten kielitutkintojen suomen kielen arvioijia voidaan pitää varsin kokeneina, sillä arviointikertoja on useita vuodessa ja arvioitavien suoritusten määrä arviointikerroilla on varsin suuri. Haastatellut arvioijat arvostivat työssään ennen kaikkea arvioitavien suoritusten suurta määrää, sillä heistä juuri se kehitti heidän arviointitaitoaan.

- (5) Tulee kerralla niin paljon arvioitavaa. Niin sitten se on kehittävää sillä tavalla, että ei ole mikään kahenkympinen oppilaan testi vaan sadan. (Erika)
- (6) On saanut niin paljon arviointikokemusta. Sitten kuitenkin omassa työssä se arviointi on tietysti vähän erilaista – sitä tapahtuu koko ajan ja sitten kurssien päätteeksi, mutta yksissä sitä saa kertaheitolla paljon. (Martta)

Lim (2011) on myös tutkimuksissaan tullut siihen johtopäätökseen, että arviointityön säännöllisyys ja/tai riittävä määrä arvioitavia suorituksia ovat arvioinnin laatua parantavia tekijöitä ja että ongelmat johdonmukaisuudessa ovat yhteydessä tavallisesti liian

vähäiseen arviointimäärään. Arviointi edellyttää arviointivuosien tueksi siis aina riittävä määrä arvioitavia suorituksia.

Schoonen, Vergeer ja Eiting (1997: 158) pitävät lisäksi arvioijien kokemusta ja tietoa arvioinnista keskeisinä tekijöinä arvioinnin reliabiliteetille ja validiudelle. Kokeneita ja vähän aikaa arvioineita on tutkittu paljon (mm. Cumming 1990; Weigle 1999; Barkaoui 2010, 2011), etenkin kirjoittamisen arviointiin liittyen. Näiden tutkimusten perusteella kokemus tuo eroja ennen kaikkea arviointiprosessiin. Arviointityötä vasta vähän aikaa tehneet arvioijat turvautuvat usein päätöstä tehdessään muutamaa helpposti tunnistettavaan arviointikriteeriin, ja heillä on puutteita kirjoitelmien lukemisessa käytettävissä strategioissa. Kokeneiden arvioijien taas on havaittu käyttävän monipuolisemmin arviointikriteereitä ja hallitsevan hyvin lukemisessa tarvittavat taidot. Lisäksi heille on kehittynyt taito kontrolloida omia arviointitaitojaan. (Cumming 1990.)

Vaikka tutkimuksissa arviointikokemusta yleisesti pidetään arviointia helpottavana tekijänä, haastateltujen arvioijien mukaan kokemuksen vaikutus omaan arviointityöhön nähtiin myös negatiivisena. Kokemus muuttuu rutiiniksi, ja sitä ei pidetty arvioinnin luotettavuutta edistävänä asiana.

- (7) Kun sä oot tehnyt sitä esimerkiksi oikein paljon, sä luotat siihen, niin kyllä siitä sitten tulee rutiininomaisempaa. Tietyllä tavalla se rutiini on hyvä juttu, mutta toisaalta se voi olla myös mun mielestä huono juttu, että jos sä rutinoit taas liikaa. (Reetta)

Kokemuksen myötä syntynyt rutiini voi olla arvioinnin näkökulmasta ongelmallinen tekijä, jos se perustuu tiukasti arvioijien omiin luotuneisiin käsityksiin kielitaidosta. Tällöin arvioija esimerkiksi painottaa arvioinnissaan vain yhtä tiettyä kriteeriä, jota hän pitää tärkeänä oman kielitaitokäsityksensä perusteella. Koska rutiini liittyy kokemukseen, se ei arvioijien kohdalla välttämättä ole huono asia, sillä kokeneille arvioijille arviointikriteerit ovat tuttuja ja heille on syntynyt rutiineita käyttää niitä arvioinnissa laajasti hyväkseen (esim. Barkaoui 2010; Cumming 1990; Eckes 2008). Lisäksi kokeneiden ja rutinoituneiden arvioijien on todettu olevan arvioinnissaan vasta-alkajia johdonmukaisempia (mm. Barkaoui 2010; Lumley 2005; Cumming 1990; Wolfe 2006). Johdonmukaisuutta arvioijien saattaa olla itse vaikea tunnistaa, mutta haastattelujen perusteella arvioijat kokivat tunnistavansa lempeyden ja ankaruuden.

- (8) Musta on tullu paljon lempeempi. Mä olin silloin vielä paljon paljon tiukempi, ku mä tulin. Se johtuu siitä, että kun mä tulin sellasesta maailmasta missä vaaditaan paljon enemmän – jos on yksiki kirjain väärin niinku sanassa, niin se on niinku virhe tai se on nolla pistettä. (Maisa)
- (9) Niin mä oon varmaan aika lempee nykyisin. Musta tuntuu, että tuun sitä löysemmäks mitä enemmän mä oon niitä tehnyt, katsonut ihmisten kamppailuja kielen oppimi-

sen kentällä. Kyllä mä varmaan enemmän osaan kattoo enemmän niitä eri puolia kuin ennen kielitaidosta. Nyt huomaan niitä ja osaa arvostaa niitä. (Ritva)

Käsitykset lempeyden lisääntymisestä olivat todennäköisesti yhteydessä siihen, että arvioijista heidän käsityksensä kielitaidosta oli muuttunut arviointi- ja työkokemuksen myötä. He eivät enää omasta mielestään antaneet arvioinnissaan niin paljon painoarvoa yksittäisille kielen piirteille, vaan kielitaitoa tarkasteltiin alkuaikojaa holistisemmin. Arviointityön alussa vaikeiksi koetut taitotasojen kuvaukset ja tasojen väliset erot olivat arviointikokemuksen myötä selkeytyneet, ja sitä kautta varmuus arvioijana oli löytynyt. Tämä kuitenkin saattoi aiheuttaa arvioijalle mielikuvan, että hänen arviointilinjansa oli muuttunut lempeämmäksi.

- (10) Mä oon varmaan ollu aika tiukka arvioija aikasemmin ja sitten kun on tullu se holistisempi näkemys siihen kielitaitoon, niin sitten se on tarkottanu ehkä myös sitä, että sitten on hiukan niinku lempeämpi arvioija. Ei aseta niitä kriteerejä joka piirteessä niin korkeelle. (Pirkko)
- (11) Martta on löysämielinen... Kyllä mulla on nykyisin semmonen olo arvioitaessa, että kyllä mä tiedän, että minkä tasonen joku on. Mulla on selkeempi kuva siitä, että mikä on alle kolmonen, mikä on kolmonen ja mikä on nelonen. (Martta)

Vaikka haastatellut arvioijat kokivat suvaitsevaisuuden ja lempeyden lisääntyneen kokemuksen myötä, arvioijien lempeydestä ja ankaruudesta tehdyt tutkimukset eivät ole vahvistaneet sitä, että kokemuksella on vaikutusta lempeyden tai ankaruuden lisääntymiseen. Tutkimuksissa (esim. Leckie & Baird 2011) on lähinnä todettu kokemuksen vaikuttaneen lempeyttä/ankaruutta enemmän arvioijien keskinäiseen yhdenmukaisuuteen. Weiglen (1999) tutkimustulokset kuitenkin tukevat arvioijien käsityksiä lempeyden lisääntymisestä, sillä kirjoittamisen arvioijia tutkiessaan hän huomasi aloittelevien arvioijien olevan kokeneita arvioijia tiukempia, mutta arviointikokemuksen kartuttua ja oman arviointilinjan löydyttyä heidän arviointilinjansa muuttui lempeämmäksi.

Yleisten kielitutkintojen arviointikokemuksen lisäksi arvioijat mainitsivat myös päivätyöstä saadun arviointi- ja työkokemuksen vaikuttaneen heidän arviointilinjaansa ja arvioija-identiteetin syntymiseen. Haastateltavat kertoivat usein vertaavansa testiin osallistujia omiin opiskelijoihinsa ja omaa arviointilinjaansa työkollegoiden arviointilinjaan.

- (12) Meillä on näitä hyvän kielitaidon omaavia paljon ja me oltas tiukkoja sen B1:n alarajan kanssa. Mä oon huomannut, että ykissä mä saan palautteessa, että mä oon tiukemmasta päästä, mikä varmaan täytyy liittyä tähän omaan työhön. (Anneli)

- (13) Tässä työyhteisössä ja näissä muissa oppilaitoksissa on mun mielestä ollu semmonen ankarampi linja ja jotenkin niinku pihdataan niitä arvioita. Kaikki ei arvio kovin paljon ja ei ole kokemusta kaikista tasoista, että sitten sokeutuu siihen yhteen tasoon ja sitten se, että jotenkin se, että ei vain nähdä, että niitä tasoja on vain niin paljon. (Henni)
- (14) Esimerkiksi meidän kouluttaja, kun mä ajattelen mikä on se B1 meidän kielitaitotavoitetaso, niin mä koen, että sitä pidetään aivan liian ylhäällä, kun ajattelen sitä mitä minä ajattelen, mikä on B1. (Heljä)

Bonk ja Ockey (2003) pitävät arvioijien ammatillista työkokemusta tai henkilökohtaista kokemusta arvioitavista tärkeänä tekijänä arvioijana kehittymiselle. Positiivinen vaikutus toimii myös haastattelujen mukaan toisinpäin, sillä arvioijien päivätyön näkökulmasta laaja kokemus arvioinnista eri kielitaidon tasolla olevista oppijoista auttoi arvioijia opetustyössä, sillä se laajensi näkemystä kielitaidon kehittämisestä.

- (15) Tää tuo varmuutta siihen omaan työhön, lisäkoulutusta siihen, koska koen että me ollaan niin kovasti erilaisia saman koulun sisällä ja koulun välinen eroavuus se on ihan järkyttävää. (Sanni)
- (16) Saa niinku laajemman kuvan siitä ja laajemmasta ihmismäärästä. Että omat opiskelijat ovat aika homogeenista porukkaa loppujen lopuksi sitten. (Martta)

Kokemus arvioinnista vaikutti myös tutkimukseni mukaan siihen, että Yleisten kielitutkintojen arvioijat eivät enää nähneet arviointia erillisenä osana omassa työssään. Arviointi oli heistä osa jokapäiväistä työtä: osa opettamista ja osa oppimista.

- (17) Kyllä mä olen sellanen kokoaikainen arvioija. Mä haluan, että opiskelija näkee siinä oman edistymisen että mitä tapahtuu ja toisaalta mä näen sen sitten, että mitä täytyy tehdä lisää tai mikä täytyy tehdä toisin. (Susanna)

5 Arviointikoulutuksen merkitys arvioijien käsityksille arvioinnista

Tyypillisesti arviointikoulutukset ovat ajallisesti suhteellisen lyhyitä tilaisuuksia, joissa tutustutaan itse arviointiin liittyviin käytänteisiin, harjoitellaan arviointia ja keskustellaan arviointikriteereistä, tehtävistä ja suorituksista sekä saadaan mahdollisesti palautetta arvioinneista (Lane & Stone; Congdon & McQueen 2000; Weigle 1998). Arviointikoulutukset on nähty tärkeäksi prosessiksi arvioinnin validiuden näkökulmasta, sillä koulutuksilla on paljon myönteisiä vaikutuksia arvioijien toimintaan, etenkin arvioijien yhdenmukaisuuteen ja kriteereiden yhtenäisen tulkintaan (mm. Weigle 1994, 1998; Shohamy, Gordon ja Kraemer 1992; Fulcher 2003).

Yleisissä kielitutkinnoissa arvioijille järjestetään koulutus aina ennen varsinaisen arvioinnin aloittamista. Koulutuksessa arvioijat käyvät keskustelua ennalta valituista suorituserityksistä arviointia ohjaavien kriteereiden avulla. Haastattelujen perusteella koulutuksissa arvioijat saivat vahvistusta omille tulkinnoilleen suorituksista ja niiden taitotasosta. Lisäksi koulutukset auttoivat heitä hahmottamaan taitotasoa paremmin ja myös antoivat arvioijille viitteitä siitä, miten lähellä omat käsitykset taitotasosta olivat muiden arvioijien käsityksiä.

- (18) Kun keskustellaan, onko tämä vaikka kolmosta tai onko tää nyt nelosta. Ja sitten kun näkee, että se mielipidejakauma voi olla tosi iso, niin tavallaan sellainen tietynlainen armollisuus siihen arviointiin ja että se ei oo niin mustavalkoinen asia ja sillä lailla niin helposti ratkaistavissa oleva asia. (Heljä)
- (19) Mä tykkään hirveesti niistä meidän koulutuksista, koska niissä sillein päivittyä aina sit omakin osaamine. Ja voi myös ihmetellä sitä, että miten niinku eri ihmisten linjat poikkeaa niin paljon toisistaan. Että tota joskus tietysti ite on niinku no väärässä tai oikeassa mutta että se on vaan niinku niin iso se hajonta sitten kuitenkin. (Maisa)

Koulutuksen myönteisistä vaikutuksista huolimatta tyypillisiä tilanteita koulutuksissa olivat myös arvioija Heljän ja arvioija Maisan yllä kertomat tilanteet, joissa arvioijat ovat eri mieltä suorituksesta. Tähän on syynä se, että koulutuksella ei voida välttämättä vaikuttaa arvioijien ankaruuteen/lempeyteen, vaan arvioijat pitävät omasta arviointilinjastaan yleensä tiukasti kiinni (Lumley & McNamara 1995; Weigle 1998; Lim 2011).

Koulutuksen merkitystä tutkineet Shohamy, Gordon ja Kraemer (1992) vertasivat koulutettujen ja kouluttamattomien arvioijien arviointeja keskenään ja tulivat siihen johtopäätökseen, että arvioijat hyötyvät koulutuksesta, sillä koulutettujen arvioijien antamat arviot olivat luotettavampia ja arvioijien välinen reliabiliteetti oli korkea. Davis (2016) tuli tutkimuksissaan hyvin samankaltaiseen tulokseen, sillä hänestä arvioinnin johdonmukaisuus ja yhteisymmärrys arvioista liittyvät koulutettujen arvioijien piirteisiin. Hänestä koulutuksen ansiosta arvioijat kykenevät antamaan tarkempia arvioita suorituksista ja arvioijien keskinäinen yhdenmukaisuus paranee.

Koulutuksella on siis tutkimusten mukaan vaikutusta arvioijille, mutta tulokset ovat varsin ristiriitaisia siinä suhteessa, millaisille arvioijille koulutuksesta on eniten hyötyä ja miten koulutus vaikuttaa arvioijiin. Weigle (1998) esimerkiksi ei nähnyt koulutuksen vaikuttavan kokeneiden arvioijien arviointilinjaan, vaan suurin etu saavutetaan kokemattomien arvioijien tai sellaisten arvioijien kohdalla, jotka ovat äärimmäisen ankaria tai lempeitä (ks. myös Lim 2011; Engelhard 1992). Orrin (2002) ja Papajohnin (2002) tutkimukset arviointikoulutuksen merkityksestä arvioijille tukevat Weiglen tutkimusta, sillä myöskään heistä koulutuksella ei ollut vaikutusta kokeneiden arvioijien käyttäytymiseen, koska he toimivat varsin samankaltaisesti arviointikerrasta toiseen.

Vaikka tutkimuksissa on päädytty erilaisiin tulkintoihin koulutuksen merkityksestä arvioijien toiminnassa, koulutuksella näyttää kuitenkin jokaisessa tutkimuksessa olevan jonkinlaisia positiivisia vaikutuksia arviointiin. Limin (2011) tutkimusten perusteella molemmat arvioijaryhmät, sekä kokeneet että vasta uransa alussa olevat arvioijat, hyötyvät koulutuksesta. Hänestä kokeneiden arvioijien kohdalla koulutus vaikuttaa niin arvioijan sisäiseen kuin arvioijien väliseen johdonmukaisuuteen, kun taas kokeamattomien arvioijien arviointilinja lähestyy koulutuksen avulla luotettavuudeltaan kokeneiden arvioijien kaltaista arviointilinjaa tai jopa saavuttaa sen. Myös haastatellut arvioijat näkivät koulutuksen ja kokemuksen vaikuttaneen heidän kehitykseensä arvioijana.

- (20) Se on jonkinlainen semmonen epävarmuus, niin se on kyllä vähentynyt, että silläan luottaa, jos sitä voi sanoa kehittymiseksi niin luottaa siihen omaan linjaansa, että näin se on. (Kaija)
- (21) Se taito, jos sitä nyt on, niin se on ihan ykin ansiota, kun siellä on käynyt ja siellä on kokemusta tullut. Aina on sitä koulutusta ja kaikki uudet asiat kuulee siellä ja koko tämä kenttä tulee tutuksi sitä kautta. Että en mä usko, että mä muuten olisi näin kehittyneet kuin nyt. (Henni)
- (22) Taitotasot vahvistuneet, etenkin se kolmosen raja. Ne on hyviä ne koulutukset siellä arvioinnin alussa että siinä aina vähän palautuu maanpinnalle. (Maarit)

Arvioijat kokivat koulutuksesta olevan heille apua, ei ainoastaan arvioijina vaan myös opettajina. Opettaja on opetusmaailmassa usein yksin vastuussa arvioinnista, eivätkä keskustelut arvioinnista ja sen perusteista ole työpaikalla kovin yleisiä. Arvioijat näkivätkin koulutusten kehittävän heitä myös ammatillisesti.

- (23) Jos aattelee tuota opetuspuolen hommaa niin vaikuttaa myös, kun itse tekee arvioitavia tehtäviä sun muita, niin on tää tavallaan koulutusta. Ja kyllä sitä on paljon, sillä lailla jäsentänyt kun ajattelee, mitä arviointi ylipäätään on ja missä kohtaa sitä tapahtuu ja mikä on oleellista ja mitä tarkoitusta varten. (Ritva)
- (24) Se on itelle ammatillisesti tärkeä. Niin kun saa sen käsityksen siitä, että saa niinku isossa porukassa miettiä sitä, mikä se on se B1 ja mitä se on se B2. (Niina)

Yhteiset koulutushetket ja keskustelut arviointipäivinä toisten kollegoiden kanssa kehittivät arviointitaitoa, mutta niihin liittyi keskeisenä tekijänä myös yhteisöllisyys, sillä tilaisuuksissa saatiin kollegiaalista tukea omaan työhön suomen toisena ja vieraana kielinä -opettajana.

- (25) Ne on minusta ihan älyttömän hyviä ne koulutukset. Mä mietin niitä keskusteluja, se kollegiaalinen keskustelu, että siihen ei niinku pääse työelämässä, siihen ei oikeas-

taan oo aikaakaan ja sitten että siellä saa sen tuen sille omalle työlleen, et kun luottaa siihen systeemiin oikeesti mitä siinä on taustalla. (Heljä)

- (26) Mä oon oppinu sen EVK:n mukaisen taitotasoluokittelun ja oppinu arvioimaan kriteeripohjaisesti, ja se on tosi tärkeä ollu omassa työssä. Ja sitten se, että täällä tapaa kollegoja eri puolilta maata ilman että olis mitään työyhteisöpaineita niissä kokoon-tumisissa. Että hirveen opettavaista tämä on ollut. (Hannele)

6 Arviointikriteerit arvioinnin tukena ja kielikäsitusten muokkaajina

Haastatellut arvioijat pitivät arviointikriteereitä tärkeänä työkaluna arvioinnin aikana, sillä niiden olemassaolo toi varmuutta arviointitilanteissa, jolloin tarvittiin apua päätöksenteossa.

- (27) Mä tarkkailen nykyään enemmän kriteerejä. Että mä haen sellaisia kriteerejä, semmosia perusteita että miksi. (Sanni)

Testiin osallistuneiden tuotoksia ei kuitenkaan aina voida suoraan sijoittaa arviointiskaalalle, vaan arvioijien täytyy usein soveltaa ja tulkita arviointiskaalaa suoritusten kohdalla. Jos arviointikriteereissä esimerkiksi todetaan, että puhumisessa arvioitava *pystyy ilmaisemaan itseään ymmärrettävästi*, arvioijilla saattaa olla toisistaan poikkeavia käsityksiä siitä, miten ymmärrettävää tuotos oli. Arvioinnin luotettavuuden kannalta onkin tärkeää tietää, kuinka yksittäinen arvioija tulkitsee kriteerit (Pollitt & Murray 1996) ja mikä hänen mielestään on esimerkiksi ”sujuvaa puhetta”. Arviointikriteerit ovat työkalu, jonka tulkinnan suunnasta päättää aina arvioija itse (Lumley 2002: 266), mutta testin validiuden vuoksi arviointikriteereiden tulkinnassa on pyrittävä arvioijien väliinseen yhdenmukaisuuteen.

Arviointiin vaikuttaa myös se, millaista arviointiskaalaa arvioijat käyttävät: analyyttistä vai holistista. Analyttisessä arvioinnissa kieltä tarkastellaan pienempinä kokonaisuuksina, jolloin eri piirteille (esim. sujuvuus, johdonmukaisuus, tarkkuus, kohe-renssi, idiomaattisuus) on jokaiselle oma kuvauksensa ja jokainen kuvaus tai kriteeri otetaan arvioinnissa huomioon (Huhta 1993: 153). Holistisessa arvioinnissa kielitaidon tasokuvaus ei taas ole niin yksityiskohtainen, kuten katkelma Yleisten kielitutkintojen taitotason 3 arviointikriteeristä osoittaa:

Puhe voi olla melko hidasta, mutta epäluontevia katkoja ei esiinny kovin paljon. Tulee ymmärrettyä siitä huolimatta, että siirtää äidinkielen tai muiden kielten rakenteita ja sanastoa kohdekieleen ja ääntäminen saattaa olla selvästi ei-kohdekielenomaista.

Tällaisesta kuvauksesta arvioijan tulee valita ne piirteet, jotka hänestä parhaiten kuvaavat arvioitavan taitoa. Kuvaus ei välttämättä kokonaisuudessaan vastaa arvioitavan taitoa, sillä kielitaidon piirteet eivät kehity yhtäaikaisesti (Huhta 1993: 151–152).

Analyttisen arviointitavan on nähty soveltuvan arviointiuraansa alkaville, koska arvioinnissa voi keskittyä tiettyihin kielitaidon piirteisiin (Ks. Huot 1993; Cumming 1990; Barkaoui 2010.) Holistinen arviointi sopii taas paremmin kokeneille arvioijille, vaikkakin se mahdollistaa eri painotukset yksittäisten arviointikriteereissä olevien komponenttien kohdalla (Lumley 2005). Pollitt ja Murray (1996) ovat huomanneet, että vaikka lopullinen arvio suorituksesta onkin sama, arvioijat tekevät päätökset varsin erilaisen prosessin kautta, ja tämä tuottaa erilaisia käsityksiä suorituksesta.

Testijärjestelmässä arviointikoulutuksen tehtävänä onkin edellä mainituiden syiden vuoksi suunnata arvioijien huomio kriteereihin mahdollisimman monesta näkökulmasta ja opettaa heitä käyttämään kriteereitä kattavasti, jotta pystytään varmistamaan arvioinnin keskittyvän testijärjestelmän kielitaitonäkemyksen kannalta keskeisiin asioihin (Weigle 2002; Fulcher 2003). Jos arvioija käyttää selkeästi vain tiettyjä kriteereitä arviota tehdessään, hän todennäköisesti kiinnittää huomionsa myös vain pieneen osaan osallistujan tuotoksesta (Shohamy 1995; Reed & Cohen 2001). Tästä syystä arviointikriteereiden merkitystä arvioinnissa on hyvä painottaa ja koulutuksen avulla pyrkiä vaikuttamaan liian yksipuolisesti kriteereitä käyttävien arvioijien toimintaan (Eckes 2008).

Myös haastatellut arvioijat kertoivat arviointiuran alussa pyrkineensä helpottamaan epävarmuutta tukeutumalla tiukasti analyttisiin arviointikriteereihin. Kokemuksen myötä arviointikriteereiden tuntemus heistä selkeästi parani ja käsitys kielitaidosta laajeni sekä sitä myöten tarve tukeutua yksittäisiin kielen piirteitä kuvaaviin arviointikriteereihin väheni.

- (28) On oppinut tosi paljon laajemmin sitä, että mitä arvioit ja se että mihin asioihin kiinnitetään huomiota ja mihin kannattais ehkä kiinnittää huomiota. (Heljä)
- (29) Ei voi keskittyä yhteen asiaan, kriteeristö on kuitenkin aika laaja, että siinä täytyy pyörittää sitä koko palettia eikä nimenomaa, että tuijottaa sitä yhtä. Jos mä ajattelen ihan niitä alkuaikoja, niin sehän on ollut semmosta sirpaleista, että joku osaa tämän, tätä se ei osaa ja tämän se osaa. Nimenomaa kehittynyt siihen holistiseen suuntaan. (Susanna)

Jos arvioija ei pysty löytämään arviointikriteereistä selvää kuvausta suorituksesta tai arviointia ohjaavat kriteerit ovat liian monitulkintaisia, etenkin kokenut arvioija saattaa käyttää arvioinnissa omia ulkopuolisia kriteereitään (Orr 2002; Sato 2012). Ulkopuolisten kriteereiden käyttö on ongelmallista sellaisissa tapauksissa, kun arvioijan omat käsitykset, stereotypiat ja ennakko-odotukset osallistujan taidosta vaikuttavat arviointiin.

Ulkopuoliset kriteerit eivät aina kuitenkaan ole välttämättä arviointiin negatiivisesti vaikuttavia, vaan ne saattavat olla hyvinkin käyttökelpoisia, mutta arvioinnin luotettavuuden näkökulmasta arvioijien on oltava niistä tietoisia. (Barkaoui 2010; Leckie & Baird 2011.) Tämä tietoisuus voidaan saavuttaa omaa toimintaa refleктоimalla, ja tietoisuutta voidaan lisätä myös arvioinnista saatavan palautteen avulla. Arvioinnin luotettavuuden takia testijärjestelmän on kuitenkin hyvä varmistaa, että arvioijilla on yhteinen käsitys siitä, mitä arvioidaan ja miten arvioidaan, jotta arvioinnin ulkopuoliset seikat eivät vaikuta arviointiin negatiivisesti (Gamaroff 2000).

7 Muut arviointiin ja käsityksiin vaikuttavat tekijät

7.1 Osallistujan persoonalliset piirteet

Arvioijan henkilökohtaiset ominaisuudet, koulutustausta ja työkokemus sekä painotukset kriteereissä ovat suhteellisen helposti tutkittavissa olevia asioita arviointiprosessissa. Näiden lisäksi arvioinnin taustalta löytyy tekijöitä, joiden havainnointi on puolestaan vaikeaa. Tällaisia ovat mm. arvioijan henkilökohtaiset mieltymykset, asenteet ja stereotyyppiat, joiden tunnistaminen myös arvioijalle itselleen saattaa olla vaikeaa. Jos arvioija kuitenkin tunnistaa ne, se auttaa häntä pyrkimyksissä objektiivisuuteen arvioinnissa.

- (30) Totta kai ne jotain vaikuttaa varmasti aina alitajuisesti, että jos ihminen tekee jonkun positiivisen vaikutelman, niin ehkä siinä voi olla vaara, että sille antaa yhden pisteen enemmän kuin jollekin toiselle. Tietoisesti sitä koitetaan välttää. (Maarit)

Ymmärtämisen osakokeissa testiin osallistujan persoonalliset piirteet eivät ole niin hallitsevia testiarvioinnin näkökulmasta kuin tuottamisen taidoissa. Puhumisessa ja kirjoittamisessa testattavan persoona tulee herkemmin esiin positiivisella tai negatiivisella tavalla. Puhumisessa arvioitavan puheääni, sen voimakkuus, korkeus, laatu ja sävy ovat läsnä arvioinnissa, mutta myös puhujan mielipiteet käsiteltävästä asiasta tai hänen asenteensa testitilanteeseen saattavat vaikuttaa arviointiprosessissa.

- (31) Jos joku puhuu hirveen hiljaa tai hirveen monotonisesti tai epäluontevasti tai kimitää, se ärsyttää suunnattomasti. Ja tyyppi joka on luonteeltaan sellanen saarnaava, mutta sitten kuitenkin puhuu ihan väärin ja paistaa läpi, että on omasta mielestään tosi hyvä. Niin nää on sellaisia ärsyttäviä tekijöitä. Huumorin läpituominen on myös tärkeä osa kielitaitoa. Mutta tietysti jos joku on hauska, niin eihän se välttämättä tarkoita sitä, että sen kielitaito ois parempi tai että voi antaa pisteitä siitä, että oma huumorintaju osuu yksin jonkun arvioitavan huumorintajun kanssa. (Martta)

Haastatellut arvioijat kertoivatkin pohtivansa usein, arvioivatko he kielitaitoa vai henkilön persoonaa. Hitaasti puhuvan ja sanojaan harkitsevan henkilön tuotos herätti usein mielikuvan kielellisistä ongelmista, vaikka taustalla oli ehkä puhujan persoonallinen tapa toimia tilanteissa tai hänellä oli mahdollisesti jokin tuottamiseen liittyvä fysiologinen ongelma. Kirjoittamisessa puolestaan henkilön persoonallinen tapa viestiä saattoi ilmetä niukkasanaanaisena tuotoksena. Arvioijat joutuivat näissä tilanteissa ottamaan kantaa siihen, riittävätkö vähäiset lauseet kielitaidon osoittamiseen.

- (32) Just puhumisessa tai kirjoittamisessa on enemmän siihen ihmiseen, tai kun sitä arvioi, niin sitä mieltii sujuvuuttakin, mikä liittyy siihen persoonaan, että ihminen on vaan hiljainen ja harkitseva. (Heljä)
- (33) Että kun alkaa kuuntelemaan ja alkaa miettiä, että onkohan hänellä ihan oikeesti joku puhevika, että mikä on sitä kielitaitoo ja mikä on jotain fyysistä vammaa. (Hannele)

Testitilanteessa tuotoksen määrällä oli kuitenkin arvioijista merkitystä, sillä *jos näyttöä on hirveen vähän, niin aika hankala siitä on paljon sanoa mitään (Ritva)*. Osallistuja, joka tuottaa paljon ja on aktiivinen vuorovaikutusta vaativissa tehtävissä, antaa arvioijista yleensä positiivisen kuvan taidostaan.

- (34) Sen mä oon huomannu et mua helposti kyllä voi hämätä semmosella niinku puhe-
tulvalla. (Milja)

Aktiivisuuden ja tuotteliaisuuden lisäksi arvioitavan tuotokseen voi liittyä vielä näennäinen sujuvuus. Näennäisellä sujuvuudella viitataan tuotoksiin, joissa kielenkäyttäjällä on kyky käyttää kieltä sujuvan oloisesti yksittäisissä, hänelle hyvin tutuissa tilanteissa, mutta vieraammissa tilanteissa hän ei enää selviäkään oppimillaan fraaseilla. Myös virolaisten puhuma suomi saattaa kuulostaa meistä varsin sujuvalta, mutta lähemmin tarkasteltuna tuotoksessa saattaa olla paljon ei-suomalaisia sanoja ja rakenteita (Ahola & Tossavainen 2016). Etenkin tällaisten sujuvalta kuulostavien, mutta tarkkuudessa ja idiomaattisuudessa puutteellisten suoritusten taitotason määrittelyn arvioijat kokivat vaikeaksi.

- (35) Ihmisellä on asiaa, on monenlaista sanottavaa, mutta sitten se niinku peittyi. Yksi ylimmän tason puhuja, jolle sattui haastattelutehtävä omalta alalta, se tiesi siitä kaiken sanastollisesti. Mutt sitten kuitenkin tavallaan se tuotos ei vaikka sitä sisältöä sillä oli enemmän kuin kellään, mutta sit kuitenkin ne rakenteen heikkoudet. (Anneli)
- (36) On tämmösiä niin kuin ehkä sosiaalisia tyyppejä, joiden se puhumisen taito on parempi kuin muut osa-alueet. Joka ehkä myöskin liittyy enemmän siihen, että ne on taitavia viestijöitä kuin että niiden kielitaito tukis niiden puhumista paljon. (Martta)

Vähäisen tuottamisen tai runsaan tuottamisen taustalla saattaa olla yhtenä tekijänä persoonallisten piirteiden lisäksi osallistujan kulttuuritausta. Kulttuuri voi ilmetä mm. siten, että osallistujan on vaikea tuottaa vapaasti virheiden pelon vuoksi. Esimerkkinä tästä arvioijat mainitsivat usein venäläistaustaiset korkeastikoulutetut osallistujat, joiden kielitaito oli heistä tavallisesti kirjallisissa taidoissa hyvä, mutta jotka arastelivat puhumista, koska pyrkimyksenä oli tuottaa kieliopillisesti korrekti suoritus.

- (37) Semmonen akateeminen tyyppi, joka on vaan tottunut kynän kanssa opiskelemaan niin se oppii ne rakenteet niin kuin venäläiset. Ne oppii ne päätteet, mutta sitten kun pitäs sitä käytäntöön siirtää, niin sittenhän se ei taho mennä. Vaikee avata suunsa, jos siinä tulee kielioppivirhe. Kyllähän se riippuu siitä, mistä päin maailmaa tulee. (Anneli)
- (38) He (venäläiset) ovat siinä kynässä ja paperissa kiinni mutta heillä voi olla osalla sellainen että he eivät uskallakaan päästää sieltä suustaan mitään ennen kuin heillä on sellainen olo että he varmasti osaa sen sataprosenttisesti. (Reetta)

Kulttuuri myös vaikuttaa ihmisen tyyliin puhua ja viestiä (esim. Salo-Lee 1996). Jos puhetyyli vaikuttaa negatiiviselta tai kielenkäyttö ei sovi viestintätilanteeseen, se saattaa arvioijista vaikuttaa koko tuotoksen ymmärrettävyyteen, ja sitä kautta myös arvioon viestijän taidoista.

- (39) Tahaton epäkohteliaisuus tai sitten ollaan niinku liian korrekteja, että ollaan opittu käyttämään konditionaalia. Sitten ne käyttää sitä ja sitten se ei kuulostakaan yhtään niin kohteliaalta vaan se kuulostaa jotenkin käskevältä. (Erika)
- (40) Jos joku on vähän aggressiivinen niin sille antaa yhden pisteen vähemmän, vaikka se osaisikin kun vaikuttaa se tyyli. Että ihan semmosia henkilökohtaisia puhetapoja, miten niitä sitten arvioidaan mikä on hyvä mikä huono. (Maarit)

Lisäksi arvioijista meidän suomalaisten on usein helpompi ymmärtää kulttuurillisesti meitä lähellä olevia osallistujia.

- (41) Tietyllä tavalla erottuu kuitenkin, että siellä on esim. eurooppalaistaustaiset joille ne asiat on selviä että on jonkilainen yhteinen tietokanta että mistä puhutaan. Ja sitten taas nämä jotka ovat lähteneet ei vain opiskelemaan kieltä vaan myös sen kulttuurin sen takana, niin se teettää enemmän töitä. (Susanna)

6.2 Ääntämisen ja käsialan vaikutus arviointiin

Yksittäisistä arviointiin vaikuttavista kielellisistä piirteistä arvioijat usein mainitsivat osallistujan taidon ääntää suomea, sillä **ääntäminen on persoonakohtaista tai äidinkielikohtaista tai kuulijakohtaista tai että minkälaista on tottunut kuulemaan ja mitä haluaa**

kuulla (Milja). Ääntämisessä etenkin **aksentti** oli heistä merkityksellinen, sillä se vaikutti ymmärrettävyyteen mutta myös osallistujasta syntyvään mielikuvaan (lisää aksentin merkityksestä: Anderson-Hsieh & Koehler 1988; Flege, Munro & MacKay 1995; Piske, MacKay & Flege 2001.) Tuttu aksentti helpotti arviointia, sillä vastaavien puhetaapojen kuuntelemiseen harjaantuu ja *harjaantuminen voi tapahtua myös tosi nopeesti (Milja)*. Aksentti aiheutti joskus negatiivisia tunteitakin, sillä jos puhujan aksentti oli hyvin voimakas tai samankaltainen aksentti toistui samalla arviointikerralla useiden arvioitavien kohdalla, se myös ärsytti arvioijia.

- (42) Kun tulee kolmaskymmenes venäläinen nainen, jolla on liudennukset, niin ei sais vaikuttaa paljon, sitten kun se alkaa ärsyttää se piirre. (Hannele)
- (43) Jos on voimakas aksentti, saattaa ärsyttää, vaikka ei haluaisikaan sen antaa vaikuttaa. (Henni)

Arvioijat kuitenkin suhtautuivat puheessa kuuluvaan aksenttiin vakavasti, sillä sen vaikutus omaan arviointiin tunnistettiin ja sen vaikutusta pyrittiin minimoimaan arviointitilanteessa. Aksenttia pidettiin osana osallistujan persoonallisuutta, josta eroon pääseminen ei ollut arvioijista helppoa. Kielenoppijalla itsellään ei myöskään ehkä ole edes tarvetta päästä eroon aksentista, jos henkilö on puhunut samalla tavoin usean vuoden ajan ja on tullut siitä huolimatta hyvin ymmärretyksi.

- (44) Ihmisen on kyllä aksentistaan tosi vaikee päästä, jos se on voimakas. Milloin on se raja ylittynyt, että nyt tää on liikaa ja mikä vielä menee. Tämä on hirmu henkilökohtainen asia minusta sekä sille puhujalle, että arvioijalle. (Maarit)

Vaikeudet ääntämyksessä ja vieraan aksentin vaikutus ymmärrettävyyteen kuitenkin askarruttivat arvioijia. Useat heistä pohtivat, kuinka paljon enemmän he opettajina ymmärtävät kuin harvoin suomea vieraalla korostuksella puhuvien kielenkäyttäjien kanssa tekemisissä olevat.

- (45) Me ollaan tottuneita siihen, me ymmärretään ehkä enemmän kuin tavallinen vastaantulija, mutta kannattais ehkä kiinnittää kuitenkin enemmän huomiota siihen ääntämisen opetukseen, että se olis selkeä, koska mä luulen, että se herättää semmosen torjuntareaktion, että vastaanottaja ei välttämättä jaks keskittyä ja paneutua siihen, että mitä tämä tarkoittaa. (Susanna)

Tottumus ja tuttuus vaikuttavat arviointiin, sillä tutkimuksissa on todettu maahanmuuttajien puhetta tottuneiden kuuntelijoiden hyväksyvän enemmän virheitä ja antavan aksentista lievempiä arvioita kuin kielen kuunteluun vähemmän tottuneet arvioijat (Thompson 1991).

Tutkimukseeni osallistuneet suomen arvioijat mainitsivat haastatteluissa lisäksi tunnistavansa varsin helposti tietyt ensikieliset arvioitavansa ääntämisestä.

- (46) Kyllä sen kuulee ihan ensimmäisestä lauseesta, se on se ääntäminen ja aksentti ja intonaatio. (Maarit)

Ensikielen tunnistaminen aiheutti joskus myös odotuksia osallistujien taitoja kohtaan. Odotukset liittyivät etenkin tiettyjä ensikieliä puhuvien osallistujien suorituksiin: esimerkiksi suomalaisilta ruotsinkielisiltä odotettiin hyvää suoritusta, mutta myös viroa ensikielenään puhuvilta odotettiin lähikielisydestä johtuen hyvää taitoa puhua ja kirjoittaa suomen kieltä (Ks. myös Ahola & Tossavainen 2016). Jos odotukset eivät oletusta kielitaidosta täytyneetkään, se saattoi aiheuttaa arvioijalle pettymyksen.

- (47) Se suomenruotsalainen, joka puhuu niin huolitellusti ja kauniisti, niin siitäkään ei sitten taas voi sitten taas heti antaa nelosta vaan että pitää kuunnella kunnolla. (Linnea)
- (48) Että nimistä kattoo että mistä päin ja että minkälaista aksenttia mutta sitä täytyy niinku vaan yrittää päästä eroon. Sitten voi tulla kauhee pettymys että jollakin on vironkielinen nimi, ja sitten se onkin kauheen vaikee koska se puhuu puoliks viroa. (Hillevi)

Arviointi on luonteeltaan hyvin tilanneherkkää, ja tästä syystä myös mieltymykset osallistujan tuotosta kohtaan olivat hyvin tilannekohtaisia ja arvioijakohtaisia. Arvioija Reetasta *sitä vain toisilta jotenkin suodattaa vähemmän ja toisilta taas sitten odottaa enemmän*. Odotukset liittyivät usein siihen, että arvioidessaan arvioijat luovat suoritusta arvioidessaan oman mielikuvansa osallistujasta ja hänen taustastaan.

- (49) Jotenkin kun sitä kuulee minkälainen herrahenkilö tai minkälainen tyttönen siellä on, sitä yrittää että se ei saa vaikuttaa, että hyvä ihminen se on varmaan, kaksvitonen ja Venäjältä just Suomeen muuttanut. (Linnea)

Puhumisen osakokeessa ääntäminen synnytti mielikuvia arvioitavista, mutta kirjoittamisessa puolestaan arvioitavan **käsiala** herätti tuntemuksia, samoin kuin monet muut kirjoittamiseen liittyvät taidot. Arvioijat mainitsivat esimerkiksi että osallistujan kyky kirjoittaa jäseneltyä tekstiä ja käyttää välimerkkejä sekä hallita kirjoittamiseen liittyvät tekniset asiat, kuten kirjoittaminen tekstin viivan yläpuolelle (ei viivan päälle), vaikuttivat mielikuvaan kirjoittajan taidoista.

- (50) Esimerkiksi se käsiala, niin tietyillä ihmisillä on niin kamala käsiala, että musta tuntuu ennen kuin mä alan sitä lukemaan, että ai kauheeta nyt tää tulee, mä en ymmärrä tästä kuitenkaan mitään ja sitten pitää oikein skarpata, että nyt luet oikeesti ja kunnolla. Kyllähän sieltä niitä lauseita löytyy, mutta se käsialakin voi olla niin luotaantyöntävä. (Reetta)

- (51) Kirjottamisessa aina joutuu miettiä sitä että minkä verran rankaisee käsialasta. Sitähän ei kuuluisi mutta kun se aina vaikuttaa kirjottamisessa, että onko se kirjoitettu riville, minkälaisilla kirjaimilla ja onko isot alkukirjaimet ja välimerkit. (Hannele)
- (52) Jos on tosi vaikeaa saada selvää kirjottamisesta niin kyllä se väistämättä vaikuttaa. On tiettyjä jotka kirjoittaa esimerkiksi viivojen päälle, niin että se on työlästä sen tekstin lukeminen vaikka se on sisällöllisesti ja niinku kielellisesti hyvää. (Milja)

Hyvä käsiala on tutkimusten mukaan (esim. Hammerschmidt & Sudsawad 2004: 188–189) opettajana toimiville erittäin tärkeä tekijä kirjoittamisen taidossa. Jos käsiala ei ole luettava tai jos kirjoitussääntöjen hallinnassa on puutteita, kirjallinen tuotos arvioidaan huonommaksi kuin sisältö edellyttäisi. Yksistään käsialan perusteella saatetaan luoda mielikuva osallistujasta jo ennen varsinaisen tekstin lukemista, ja siitä tunnustetaan myös kirjoittajan ensikieli.

- (53) Ja mä en voi sille mitään että siellä taustalla on se että mua ärsyttää ne tietyt kansalaisuudet tai tietyt tyypit ja se tulee heti siitä käsialasta, ahaa tää on taas näitä, siitä on vaikeaa päästä eroon. (Hannele)
- (54) Harvoin tulee sillein eteen kirjoitettua tekstiä, että ei tietäisi. Oikeestaan melkeen aina tietää että mistä se teksti tulee. (Hillevi)

7 Lopuksi

Arvioijat näkivät arviointityöstä saadulla kokemuksella ja Yleisten kielitutkintojen arviointikoulutuksella olevan merkitystä heidän käsityksilleen arvioinnista ja kehittymiselleen arvioijana. Arviointikertojen ja arvioitavien suoritusten suuri määrä autoivat heitä hahmottamaan taitotasot ja löytämään oman arviointilinjan. Oma arvioijaidentiteetti muokkaantui kokemuksen myötä, ja se muodostui epävarmuuden sietämisestä ja erilaisten kielenkäyttäjien piirteiden hyväksymisestä sekä omien arvioitaviin liittyvien stereotyyppien ja ennako-odotusten tunnistamisesta.

Kokemus arvioinnista toi varmuutta arviointiin, mutta se ei poistanut sitä tosiasiaa, että arviointiin liittyy aina epävarmuus. Arvioitavien suoritukset olivat arvioijista usein epätasaisia, eikä niiden sijoittaminen arviointiasteikolle ollut helppoa, vaikka arvioinnista oli paljon kokemusta. Kun arvioitavien kielitaito ei kaikilta osin vastannut kriteereissä esitettyjä kuvauksia, arviointikriteerit ja oma tausta osallistujien kaltaisten oppijoiden opettamisesta autoivat taitotason määrittelyssä.

Arviointikoulutukset olivat tärkeä osa arvioijaksi kasvamista, sillä ne muokkasivat arvioijina toimivien käsitystä itsestään arvioijana, mutta ne antoivat heille myös eväitä omaan päivätyöhön S2-opettajana. Koulutukset nähtiin eräänlaisina vertaisarviointitilaisuuksina, joissa arvioijat saivat välitöntä palautetta omasta arviointilinjastaan

ja pystyivät vertailemaan omaa tulkintaansa arviointikriteereistä muiden arvioijien tulkintaan. Koulutuksen avulla arvioijat kokivat saaneensa aiempaa selkeämmän ja laajemman käsityksen kielitaidon kehittymisestä ja eri taitotasojen piirteistä.

Koulutusten ja kokemuksen myötä arvioijat kertoivat sallivuuden virheille lisääntyneen arvioinnissaan. Lempyyden lisääntyminen oli kuitenkin tyypillisimmin yhteydessä siihen, että arviointikriteerit olivat tulleet tutummiksi ja niitä käytettiin laajemmin hyväksi eikä arvioinnissa enää turvauduttu niin voimakkaasti yksittäisiin analyttisiin kriteereihin. Arviointi oli holistisempaa, ja tästä syystä arvioijat kokivat myös kielitaitokäsityksensä muuttuneen holistisempaan suuntaan.

Arvioijan toimintaan vaikuttavat haastattelujen perusteella lisäksi monet arvioitavan henkilökohtaisiin ominaisuuksiin ja piirteisiin liittyvät tekijät. Nämä tekijät, kuten puhetyyli ja käsiala, eivät välttämättä liity suoraan kielitaitoon, mutta ne ovat kuitenkin läsnä arviointiprosessissa. Vaikka arviointiprosessissa pyritäänkin poistamaan arviointiin vaikuttavat ulkoiset tekijät luotettavuuden varmistamiseksi, arviointi on kuitenkin inhimillistä toimintaa, jossa monet arvioitavan henkilökohtaiset ominaisuudet, mielipiteet ja tavat sekä tiedot saattavat vaikuttaa arvioijan toimintaan. Jotta voidaan turvata arvioitavien yhtäläinen kohtelu arviointitilanteessa, tulee testijärjestelmän huolehtia siitä, että arvioijat ovat tietoisia myös näiden tekijöiden vaikutuksesta arviointiprosessissa. Omien arviointipäätösten reflektointi ja arviointikriteereiden merkityksen korostaminen päätöksenteossa auttavat myös arvioijaa itseään pyrkimyksissä objektiivisuuteen arvioinnissa.

Kirjallisuus

- Ahola, S. & H. Tossavainen 2014. Yleisten kielitutkintojen arvioijakoulutus. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014: 16. Helsinki: Opetushallitus, 125–134.
- Ahola, S. & H. Tossavainen 2016. Helppoa vai vaikea arvioida? –Yleisten kielitutkintojen suomen kielen arvioijien käsityksiä vironkielisten arvioinnista. *Lähivõrdlusi. Lähivertailuja*, 27, 27–53. DOI:10.5128/LV26.01. Tallinna: Eesti Rakenduslingvistika Ühing.
- Anderson-Hsieh, J. & Koehler, K. 1988. The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, 38 (4), 561–613.
- Barkaoui, K. 2010. Variability in ESL essay rating process: the role of the rating scale and rater experience. *Language Assessment Quarterly*, 7 (1), 54–74.
- Barkaoui K. 2011. Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice*, 18 (3), 279–293.
- Bonk, W. & G. Ockey 2003. A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*, 20 (1), 89–110.
- Congdon, P. & J. McQueen 2000. The stability of rater severity in large-scale assessment programs. *Journal of Educational Measurement*, 37 (2), 163–178.
- Cumming, A. 1990. Expertise in evaluating second language compositions. *Language Testing*, 7 (1), 31–51.

- Davies, L. 2016. The influence of training and experience on rater performance in scoring spoken language. *Language Testing*, 33 (1), 117–135.
- Eckes, T. 2008. Rater types in writing performance assessments: a classification approach to rater variability. *Language Testing*, 25 (2), 155–185.
- Engelhard, G. 1992. The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education*, 5 (3), 171–191.
- Flège, J., M. Munro & I. MacKay 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97 (5), 3125–3134.
- Fulcher, G. 2003. *Testing second language speaking*. Lontoo: Longman/Pearson.
- Gamaroff, R. 2000. Rater reliability in language assessment: the bug of all bears. *System*, 28 (1), 31–53.
- Hammerschmidt, S. & P. Sudsawad 2004. Teacher's survey on problems with handwriting: referral, evaluation, and outcomes. *American Journal of Occupational Therapy*, 58 (2), 185–192.
- Hirsjärvi, S. & H. Hurme 1985. *Teemahaastattelu*. 3.painos. Helsinki: Gaudeamus.
- Huhta, A. 1993. Suullisen kielitaidon arviointi. Teoksessa S. Takala (toim.) *Suullinen kielitaito ja sen arviointi*. Kasvatustieteiden tutkimuslaitoksen julkaisusarja B. Teoriaa ja käytäntöä 77. Jyväskylä: Kasvatustieteiden tutkimuslaitos, 143–225.
- Huot, B. 1993. The influence of holistic scoring procedures on reading and rating student essays. Teoksessa M. Williamson & B. Huot (toim.) *Validating holistic scoring for writing assessment*. Cresskill, NJ: Hampton Press, 206–232.
- Johnson, J. & G. Lim 2009. The influence of rater language background on writing performance assessment. *Language Testing*, 26 (4), 485–505.
- Lane, S. & C. Stone 2006. Performance Assessments. Teoksessa B. Brennan (toim.) *Educational Measurement*. Westport, CT: American Council on Education & Praeger, 387–431.
- Leckie, G. & J-A. Baird 2011. Rater effects on essay scoring: a multilevel analysis of severity drift, central tendency and rater experience. *Journal of Educational Measurement*, 48 (4), 399 – 418.
- Lim, G. 2011. The development and maintenance of rating quality in performance writing assessment: a longitudinal study of new and experienced raters. *Language Testing*, 28 (4), 543–560.
- Lumley, T. 2002. Assessment criteria in a large-scale writing test: what do they really mean to the raters? *Language Testing*, 19 (3), 246–276
- Lumley, T. 2005. *Assessing second language writing: the rater's perspective*. Frankfurt am Main: Peter Lang.
- Lumley, T. & T. McNamara 1995. Rater characteristics and rater bias: implications for training. *Language Testing*, 12 (1), 54–71.
- Metsämuuronen, J. 2005. *Tutkimuksen tekemisen perusteet ihmistieteissä*. Jyväskylä: Gummerus.
- Orr, M. 2002. The FCE speaking test: using rater reports to help interpret test scores. *System*, 30 (2), 143–154.
- Papajohn, D. 2002. Concept mapping for rater training. *TESOL Quarterly*, 36 (2), 219–233.
- Piske, T., I. MacKay & J. Flège 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29 (2), 191–215.
- Pollitt, A. & N. Murray 1996. What raters really pay attention to. Teoksessa M. Milanovic & N. Saville (toim.) *Performance testing, cognition and assessment*. Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press, 74–91.
- Reed, D. & A. Cohen 2001. Revisiting raters and ratings in oral language assessment. Teoksessa C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley & T. McNamara

- (toim.) *Experimenting with uncertainty: language testing essays in honor of Alan Davies*. Cambridge: Cambridge University Press, 82–96.
- Salo-Lee, L. 1996. Minä, me ja muut. Teoksessa L. Salo-Lee, R. Malmberg & R. Halinoja (toim.) *Me ja muut. Kulttuurienvälinen viestintä*. Helsinki: Ylen opetusohjelmat, 88–98.
- Sato, T. 2012. The contribution of test-takers' speech content to scores on an English oral proficiency test. *Language Testing*, 29 (2), 223–241.
- Schoonen, R., M. Vergeer & M. Eiting 1997. The assessment of writing ability: expert readers versus lay readers. *Language Testing*, 14 (2), 157–184.
- Shohamy, E., C. Gordon & R. Kraemer 1992. The effect of raters' background and training on the reliability of direct writing tests. *The Modern Language Journal*, 76 (1), 27–33.
- Tarnanen, M. 2014. Arvioija taidon arvottajana. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014: 16. Opetushallitus, 115–124.
- Thompson, I. 1991. Foreign accents revisited: the English pronunciation of Russian immigrants. *Language Learning*, 41 (2), 177–204.
- Tuomi, J. & A. Sarajärvi 2004. *Laadullinen tutkimus ja sisällönanalyysi*. Jyväskylä: Gummerus.
- Törmäkangas, K. & T. Törmäkangas 2009. *Osioanalyysi testien arvioinnissa*. Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.
- Weigle, S. 1994. Effects of training on raters of ESL compositions. *Language Testing*, 11 (2), 197–223.
- Weigle, S. 1998. Using FACETS to model rater training effects. *Language Testing*, 15 (2), 263–287.
- Weigle, S. 1999. Investigating rater/prompt interactions in writing assessment. Quantitative and qualitative approaches. *Assessing Writing*, 6 (2), 145–178.
- Weigle, S. 2002. *Assessing Writing*. Cambridge: Cambridge University Press.
- Wolfe, E. W. 2006. Uncovering rater's cognitive processing and focus using think-aloud protocols. *Journal of Writing Assessment*, 2 (1), 37–56.
- Wright, B. & M. Stone 1979. *Best Test Design*. Chicago, Ill: MESA.