

Rimaa hipoen selviää tilanteesta

Yleisten kielitutkintojen suomen kielen arvioijien käsityksiä kielitaidon arvioinnista ja suullisesta kielitaidosta

SARI AHOLA

Väitöksenalkajaisitelmä Jyväskylän yliopistossa 18. helmikuuta 2022

Tarkastelen väitöksenalkajaisitelmässäni kielitaidon arviointia ja arvioijana toimimista. Sana *arviointi* yhdistetään yleensä kouluissa suoritettavaan arviointiin ja ennen kaikkea todistuksessa oleviin numeroihin tai sanallisiin arvioihin. Väitöstutkimukseni keskiössä on kuitenkin testiarviointi, joka hieman eroaa opetukseen liittyvästä arvioinnista. Kouluarvioinnissa yksittäinen opettaja tekee päätöksen oppilaan osaamisesta nojautumalla opetuksen aikana suoritettuihin kokeisiin ja oppilaan tuntityöskentelyyn. Sen sijaan testiarvioinnissa arviointia tekevät eivät tunne arvioitavaa eivätkä tiedä hänen opiskeluhistoriaansa, vaan testiin osallistuvalla annetaan arvio osaamisestaan testitilanteessa suoritettujen näytteen perusteella. Jotta testattavan saama arvio on mahdollisimman oikea, testijärjestelmässä toimivien täytyy huolehtia siitä, että testattavan osaamisesta saadaan kattava näyttö ja että hänen saamansa arvio todella vastaa hänen taitojaan ja ennustaa osuvasti hänen selviytymistään erilaisissa kielenkäyttötilanteissa. Luotettava ja oikeudenmukainen testiarviointi on laadukasta. Laadukas arviointi tarkoittaa lyhyesti sitä, että arvioinnista vastaavat huolehtivat testissä käytettyjen monipuolisten tehtävien ja testikokonaisuuksien vaikeustasosta, toimivuudesta ja erottelukyvystä. Lisäksi on huomioitava, että arviointia ohjaavat arviointikriteerit kuvaavat kielitaitoa tarkasti ja että arvioijat noudattavat ja tulkitsevat arviointikriteereitä samalla tavoin. Keskeistä on, että arvioinnin joka vaiheessa varmistetaan, että osallistujat saavat yhtäläiset oikeudet osoittaa taitonsa ja että heitä kohdellaan arvioinnissa oikeudenmukaisesti ja yhdenvertaisesti. (Esim. Weir 2005; Bachman & Palmer 1996.)

Tutkimukseni keskiössä on puhumisen arviointi, joka on usein jäänyt kouluarvioinnissa vähemmällä huomiolla. Esimerkiksi ylioppilaskokeissa arviointi on kohdistunut kielen ymmärtämiseen ja kirjoittamiseen, mutta puhumisen taitoa niissä ei

toistaiseksi vielä arvioida. Tässä mielessä tutkimukseni konteksti, kansallinen kielitaidon arviointijärjestelmä Yleiset kielitutkinnot, on ollut edelläkävijä suomalaisessa kielikoulutuksessa. Tutkintojärjestelmässä puhumisen taitoa on arvioitu jo vuodesta 1994 lähtien. Puhumisen arvioinnista on siis kokemusta, mutta puhumisen arviointia on tutkittu melko vähän. Järjestelmässä toimii yli 160 arvioijaa, joista noin 100 arvioi suomen kielen tutkintosuorituksia kuusi kertaa vuodessa. Arvioijien toimintaa seurataan joka arviointikerta tilastollisten analyysien avulla, ja niiden perusteella tiedetään, millainen yksittäisen arvioijan arviointilinja on: onko hän ankara vai lempeä ja onko hän johdonmukainen arvioinnissaan. Matemaattisten mallien avulla ei kuitenkaan päästä käsiksi siihen, mitä arvioijat ajattelevat arvioinnista ja suullisesta kielitaidosta. Tämän vuoksi laadullisessa väitöskirjatutkimuksessani olenkin pyrkinyt selvittämään arvioijien käsityksiä puhumisesta ja yleensä arvioinnista.

Olen kouluttanut Yleisten kielitutkintojen suomen kielen arvioijia usean vuoden ajan, ja koulutuksissa olen usein pohtinut, miksi arvioijat päätyvät tiettyyn arvioon, miksi heidän lempeys–ankarus-linjansa vaihtelevat ja miksi he perustelevat samojen oppijoiden suorituksia joskus eri kielen piirteiden kautta. Näiden omasta työkokemuksestani kumpuavien kysymysten lisäksi halusin tutkia sitä, mikä merkitys tutkintojärjestelmän sisäisillä toiminnoilla, kuten koulutuksella ja arviointikriteereillä, on arvioijana toimimiselle ja miten ne tukevat arviointia. Lisäksi halusin tarkastella tutkintojärjestelmän ulkopuolisten tekijöiden merkitystä arvioinnissa. Tällaisia arviointikriteereiden ulkopuolisia tekijöitä ovat muun muassa arvioijan omat mieltymykset, kokemukset ja käsitykset eri oppijoista tai tavoista tuottaa suomea. Ei ole yhdentekevää, millä perusteella päätökset kielitaidon tasosta tehdään, koska tutkintoon osallistuneiden saamat arviot vaikuttavat heidän elämäänsä Suomessa. Suomen tutkintoihin osallistuu vuosittain noin 8 000 suomenoppijaa, ja suurin osa heistä tarvitsee tutkintotodistusta osoittaakseen kielitaitonsa Suomen kansalaisuutta, työelämää tai opiskelua varten.

Havainnollistan seuraavaksi puhumisen arviointiin liittyviä tekijöitä. Pyydänkin teitä kuvittelemaan mielessänne yhden arkisen tilanteen. Istutte bussissa, ja takananne istuu kaksi vieraalla korostuksella suomea puhuvaa henkilöä. Kuuntelette heidän keskusteluansa, mutta alussa ette ymmärrä kaikkea, mitä toinen heistä kertoo matkakumppanilleen. Jonkin aikaa kuunneltuanne ymmärrätte, että hän tarkoittaa sanalla *makkuttaminen* matkustamista ja että hänellä on vaikeuksia ääntää suomea: *tukkaan suomalaista luoasta* – tykkään suomalaisesta ruoasta. Toisen puhujan puhe kuulostaa taas melko suomalaiselta, vähän ehkä länsisuomalaiselta murteelta, mutta hän käyttää puheessaan sanoja, jotka eivät kuulosta suomalta: *asuttaan pienessä kaupunkissa, mutt ei mulla oo palju söpru siel*. Asiayhteydestä kuitenkin ymmärrätte, että hän puhuu ystävästä. Tällä lyhyellä esimerkillä halusin kuvata sitä, millaisia kielenkäyttötilanteita suomen kielen arvioijat kohtaavat ja mihin heidän huomionsa mahdollisesti kiinnittyy arvioinnin aikana. Arvioitaessa ymmärtäminen voi olla joskus vaikeaa esimerkiksi oudon aksentin vuoksi, ja joskus puheessa on niin paljon rakenteellisia ja sanastollisia puutteita, että arvioijan on vaikea seurata puhetta, vaikka hän on arvioinnin ja suomen kielen asiantuntija. Eräs arvioija totesikin, että joskus *oppijoiden puhetta olisi mahdotonta ymmärtää, jollei itsellä olisi ollut kyseisen ensikielen taustaisia opiskelijoita*.

Suullisen kielitaidon arviointi on tilanne, jossa tutkimukseni mukaan arvioija havainnoi puhujaa monesta eri kulmasta. Ensimmäisten sekuntien aikana hän tekee usein päätelmiä oppijan ensikielestä, ja sen jälkeen hän saattaa havainnoida oppijan tapaa puhua ja hänen puheensa sisältöä. Samanaikaisesti arvioija myös tarkkailee koko ajan kielitaitoa. Vaikka arviointikriteereissä arvioijia ohjataan tarkastelemaan puhetta usean eri analyttisen kriteerin kautta, vaikuttaa kuitenkin siltä, että vain osa kriteereistä on tärkeitä määriteltäessä oppijan puhumisen taitoa. Tällaisia kriteereitä olivat tutkimukseni mukaan muun muassa rakenteiden hallinta, sujuvuus, ääntäminen ja ymmärrettävyys. Vastaavasti oli taas kriteereitä, jotka jäivät arviointiprosessin aikana vähemmälle huomiolle, kuten joustavuus ja koherenssi ja sidosteisuus.

Rakenteiden hallinta on opetusmaailmassa perinteisesti ollut keskeinen kielellinen piirre arvioitaessa oppijoiden kielitaidon hyvyttä tai huonoutta (Saarinen, Kauppinen & Kangasvieri 2019; Honko 2017). Rakenteiden oikeellisuus vaikuttaa olevan myös tärkeä kriteeri suomen arvioijille. Tämä ei ole yllättävää, koska suomen kielessä väärän sijamuodon tai verbimuodon huomaa helposti, esimerkiksi *ma haluta mennä sun luona* tai *minä auttaa sinulle*. Lisäksi opettajina toimivat arvioijat ovat tottuneet tarkkailemaan rakenteiden puutteita ja korjaamaan niitä (mm. Tarnanen 2002). Tutkimukseni perusteella en kuitenkaan voi sanoa, että rakenteiden painottuminen arvioinnissa tarkoittaisi sitä, että arvioijien kielitaitokäsitykset olisivat formalistisia ja että he arvostaisivat kielitaidossa ainoastaan kielen oikeellisuutta. Mutta toki havainnot antavat viitteitä siitä, että rakenteilla on edelleenkin merkittävä rooli niin arvioinnissa kuin kielen opetuksessa, vaikka viime vuosina rakennekeskeisyydestä on pyritty siirtymään enemmän kielenkäytön ja viestinnällisyyden painottamiseen (myös Honko 2017).

Vaikka rakenteita pidettiin tärkeinä, arvioijat mainitsivat myös hyvin usein arvostavansa puhumisen taidossa sujuvuutta ja ymmärrettävyyttä. Näiden kahden piirteen toistuminen arviointipäätöksissä viestii minusta toiminnallisesta kielitaitokäsityksestä, koska viestinnällisyyttä korostavat opettajat pitävät tutkimusten mukaan (esim. Sato 2011; Iwashita & Grove 2003) sujuvuutta vastakohtana kieliopilliselle tarkkuudelle, ja ymmärrettävyyden korostaminen viestii taas holistisesta lähestymistavasta oppijoiden kielitaitoon (esim. Iwashita, Brown, McNamara & O'Hagan 2008). Puheen sujuvuus liitetään tavallisesti siihen, että puheessa ei ole taukoja ja että puhetta tuotetaan sopivalla, melko nopealla temmolla (esim. de Jong, Groenhout, Schoonen & Hulstijn 2015). Arvioijille sujuvuus oli kuitenkin laajempi merkitykseltään. Sujuvuuteen yhdistettiin oppijan taito käyttää tehtävässä annettu puheaika tuotoksella (myös Kormos & Dénes 2004) sekä aktiivisuus ja yritteliäisyys puhua (myös Ockey 2009), mutta sillä viitattiin myös yleisesti hyvään kielitaitoon (myös de Jong ym. 2015; Chambers 1997). Sujuvuus oli siis arvioijille eräänlainen yleiskriteeri, jota he käyttivät varsin laajasti kuvatessaan puhumisen taitoa oppijan kielitaidon tasosta riippumatta.

Puheen ymmärrettävyys oli sujuvuuden rinnalla varsin tyypillinen arviointiperuste. Lisäksi se usein liitettiin kielitaidon riittävyteen. Riittävä kielitaidon määritelmä ei kuitenkaan ole yksiselitteinen, koska se riippuu kunkin oppijan kielenkäyttötarpeista. Todennäköisesti sairaanhoitajana työskentelevälle suomenoppijalle riittävä kielitaito on erilainen kuin kaupungin lähiliikenteessä työskentelevälle bussikuskille. Hyvin vähäisenkin kielitaito voi olla riittävä esimerkiksi kaupassa asioimiseen, koska itsepalvelu-

kassalla ei suullista kielitaitoa yleensä tarvita. Arvioijien käsityksissä riittävä kielitaito tarkoitti oppijan kykyä selviytyä kielitaidollaan ymmärrettävästi arjessa ja täyttää yhteiskunnasta kumpuavat kielitaitovaatimukset, joissa Yleisten kielitutkintojen taitotaso 3 eli Eurooppalaisen viitekehyksen itsenäisen kielenkäyttäjän taso B1 on usein niin sanottu kynnystaso. Ymmärrettävyys ei ole erillinen analyttinen kriteeri Yleisten kielitutkintojen arviointikriteereissä, vaan se sisältyy kolmeen analyttiseen kriteeriin: sujuvuuteen, ilmaisun tarkkuuteen, laajuuteen ja idiomaattisuuteen sekä ääntämiseen ja fonologiseen hallintaan. Ymmärrettävyys oli kuitenkin yksi tavallisempia perusteluja arviointipäätökselle, mutta arvioijien perusteluissa oli ristiriitaisuuksia. Saman oppijan tuotosta saatettiin pitää ymmärrettävänä tai epäymmärrettävänä. Tästä syystä ymmärrettävyyttä voidaan pitää ehkä jopa subjektiivisena arviointikriteerinä: yhdelle ymmärrettävä ei ole välttämättä toiselle ymmärrettävä (esim. Isaacs, Trofimovich & Foote 2018; Derwing & Munro 1997). Jos ymmärrettävyyttä käytetään arvioinnin perusteena, on arviointia tekevän hyvä pohtia, kenelle puhe on ymmärrettävää: opettajana työskentelevälle arvioijalle vai arkielämän puhetilanteissa oppijan kanssa vuorovaikutuksessa olevalle. Tästä syystä puheen ymmärrettävyyttä ei pitäisi käyttää ainoana arviointiperusteena, vaan tarvitaan aina laajemmat perustelut oppijan saamalle arviolle.

Osin tutkimusaineistostani johtuen ääntämisen merkitys kielitaidon arvioinnissa korostui tutkimuksessani. Käyttämässäni *Rikkinäistä suomea* -hankkeen aineistossa arvioijia pyydettiin nimeämään oppijoiden ensikieli ja antamaan perustelunsa oletukselle. Ensikielen tunnistamista ei normaalissa arviointitilanteessa vaadita, mutta aikaisempien tutkimusten (esim. Kang 2008) perusteella on todettu, että arvioinnin aikana arvioijat tekevät havaintoja oppijan ensikielestä ja tämän perusteella päätelmiä jopa hänen taustastaan. Tutkimukseni mukaan ensikielen tunnistaminen loi odotuksia oppijoiden kielitaitoa kohtaan, koska esimerkiksi vironkielisiksi tunnistetuilta odotettiin lähdekielisyyden perusteella varsin hyvää suomen taitoa, kun taas thainkielisten ei odotettu saavuttavan keskitasoa parempaa puhumisen taitoa ensikielestä johtuvien ääntämiseen liittyvien vaikeuksien takia. Arvioijilla on paljon kokemuksia eritaustaisista oppijoista, minkä vuoksi he osasivat nimetä eri ensikielitaustaisille oppijoille tyypillisiä helppouksia ja vaikeuksia suomen kielessä. Ensikielen tunnistaminen herätti mielikuvia ja odotuksia myös silloin, kun oletettu ensikieli oli väärä. Oppijan suorituksesta kuultiin piirteitä, jotka olivat arvioijista tyypillisiä tämän vääräksi oletetun ensikielen puhujille. Ei ole ehkä mahdollista eikä järkevääkään pyrkiä häivyttämään puheesta ensikielen tai yleensä vieraan kielen piirteitä, mutta tutkimukseni osoittaa, että viestintätilanteissa oppijaa auttaa hyvä suomen ääntämisen taito. Tutkimustulosteni perusteella suosittelen, että opetuksessa kiinnitettäisiin aikaisempaa enemmän huomiota suomen ääntämiseen, koska se luo ensivaikutelman oppijan kielitaidosta, ja siten se voi vaikuttaa laajemminkin käsityksiin oppijan kielitaidon hyvydestä tai huonoudesta.

Arviointitilanne on prosessi, joka vaihtelee oppijakohtaisesti. Vaihtelua prosessiin tuo se, että tutkintoihin osallistuvien suomenoppijoiden taidot eivät ole samankaltaisia eivätkä he tuota suomea samalla tavoin. Myöskään yksittäisen oppijan suullisen kielitaidon eri piirteet eivät kehity samanaikaisesti. Esimerkiksi oppija voi hallita hyvin suomen sanat ja hänen ääntämisensä vaikuttaa ymmärrettävältä, mutta rakenteissa on paljon puutteita ja hän ei osaa yhdistää ilmauksia johdonmukaiseksi kokonaisu-

deksi. Osallistujien tuotoksen sisällöt myös vaihtelevat, vaikka tehtävöohjeistuksella niitä pyritään yhdenmukaistamaan ja näin helpottamaan arviointia. Arvioitavan suomenoppijan lisäksi myös arvioijat ovat erilaisia. Yhtäläisestä opettaja- ja arviointikoulutuksesta huolimatta arvioijilla on erilainen temperamentti, joka vaikuttaa lempeyteen ja ankaruuteen (esim. Zhu, Fung, Tse & Hsieh 2017). Heidän psyykkinen ja fyysinen olotilansa vaihtelee arviointitilanteissa, heillä on erilaisia kokemuksia arvioinnista ja eri kielitaustaisista oppijoista, ja he myös arvostavat suomen kielen suullisessa taidossa eri piirteitä. Arvioijien ja arvioitavien oppijoiden erilaisuus ja arviointitilanteiden vaihtelevuus ovat todellisuutta, joka tutkintojärjestelmän on huomioitava varmistakseen arvioinnin laadukkuus. Laadukkuus edellyttää jatkuvaa arviointikoulutusta, pyrkimyksiä edistää arviointikriteereiden yhdenmukaista tulkintaa sekä säännöllistä arvioijien toiminnan monitorointia ja palautteen antamista.

Yleisissä kielitutkinnoissa on pitkä kokemus arviointikoulutuksista. Arviointialan tutkimuksissa (esim. Lim 2011; Lumley 2005; Orr 2002) koulutuksen merkityksestä arviointikäyttäjymiselle esitetään ristiriitaisia tuloksia, mutta suomen kielen arvioijille koulutukset olivat tutkimukseni mukaan tärkeitä. Koulutusten myötä heidän käsityksensä kielitaidon kehittymisestä oli vahvistunut, ja koulutukset auttoivat heitä tunnistamaan paremmin eri taitotasolla olevat suoritukset. Koulutuksissa käytyjen keskustelujen ja tutkintojärjestelmästä saadun arvioijapalautteen kautta heille oli myös syntynyt käsitys omasta arviointilinjastaan. Arvioijaidentiteettiä vahvistivat myös arvioijajayhteisön tuki ja epäviralliset keskustelut kollegojen kanssa arviointitilaisuuksissa. *Arviointi on opetusmaailmassa usein yksinäistä puurtamista, vaikka siihen oikeasti tarvittaisiin tukea kollegoilta*, totesi eräs arvioijista haastattelussa kysyessäni häneltä arviointikoulutusten merkityksestä.

Arvioijien mukaan arviointikokemus helpotti päätösten tekoa, koska arviointikertojen ja suurien arviointimäärien myötä arviointikriteerit tulivat tutuiksi ja eri taitotasojen piirteet selkeytyivät. Lisäksi kokemuksen myötä arvioijat kokivat suvaitsevaisuuden kielitaidon puutteita kohtaan lisääntyneen ja arviointilinjan muuttuneen lempeämmäksi. Vastaava ilmiö on havaittu aikaisemmissa tutkimuksissa (esim. Weigle 1999), ja se on yhdistetty siihen, että kokeneet arvioijat tarkastelevat suorituksia holistisesti. Holistinen arviointitapa selittää tutkimukseni perusteella myös suomen arvioijien kokemuksia suvaitsevaisuuden lisääntymisestä ja arviointilinjan muuttumisesta lempeämmäksi. Tämä voi tuntua ristiriitaiselta päätelmästä, koska aiemmin mainitsin arvioijien havainnoivan arvioinnin aikana yksittäisiä kielen piirteitä, kuten rakenteita ja ääntämistä. Holistisesta arviointitavasta todistavat kuitenkin puheen ymmärrettävyyden ja sujuvuuden korostuminen arvioinnissa. Lisäksi arvioijat käyttivät holistista yleiskriteeriä hyväkseen etenkin palkitessaan oppijoita. Näin tapahtui muun muassa thainkielisten oppijoiden kohdalla, jotka saivat arvioijilta holistisesta yleiskriteeristä paremman arvion kuin yksittäisistä analyttisistä kriteereistä.

Tutkimukseni on tuonut esiin arvioijien erilaisia polkuja kulkea kohti lopullista arviointipäätöstä. Vaikka tilastollisten analyysien ja aikaisempien tutkimusten (esim. Lumley 2005; Orr 2002) perusteella arvioijat tavallisesti päätyvät samaan loppuarvioon polkujen erilaisuudesta huolimatta, tutkimukseni on kuitenkin tuonut lisää ymmärrystä arvioijana toimimisesta ja arviointikäyttäjymiseen vaikuttavista teki-

jöistä. Mutta tutkimuksella on merkitystä myös arvioinnin laadukkuudelle. Mainitsin lektioni alussa laadukkaaseen arviointiin liittyviä piirteitä: oikeudenmukaisuuden, oikeellisuuden, luotettavuuden, osuvuuden ja yhdenvertaisuuden. Tutkimukseni nostaa esiin vielä yhden keskeisen laadukkaan arvioinnin piirteen: arvioinnin läpinäkyvyyden. Ymmärrys siitä, mitä arvioijat ajattelevat, mihin heidän arvionsa perustuvat ja mitä he yleensä ajattelevat arvioinnista, on tutkintojärjestelmässä yksi tärkeä lisäaskel kohti arvioinnin läpinäkyvyyttä.

Lähteet

- BACHMAN, LYLE F. – PALMER, ADRIAN S. 1996: *Language testing in practice. Designing and developing useful language tests*. Oxford: Oxford University Press.
- CHAMBERS, FRANCINE 1997: What do we mean by fluency? – *System* 25 (4) s. 535–544. [https://doi.org/10.1016/S0346-251X\(97\)00046-8](https://doi.org/10.1016/S0346-251X(97)00046-8).
- DERWING, TRACEY M. – MUNRO, MURRAY J. 1997: Accent, intelligibility, and comprehensibility. Evidence from four Lis. – *Studies in Second Language Acquisition* 19 (1) s. 1–16. <https://doi.org/10.1017/S0272263197001010>.
- CHAMBERS, FRANCINE 1997: What do we mean by fluency? – *System* 25 (4) s. 535–544. [https://doi.org/10.1016/S0346-251X\(97\)00046-8](https://doi.org/10.1016/S0346-251X(97)00046-8).
- HONKO, MARI 2017: Kieli- ja kielitaitokäsitykset tutkivan opettajan kenttäpäiväkirjamerkinnöissä. – *Puhe ja kieli* 37 (4) s. 215–238. <https://doi.org/10.23997/pk.63203>.
- ISAACS, TALIA – TROFIMOVICH, PAVEL – FOOTE, JENNIFER ANN 2018: Developing a user-oriented second language comprehensibility scale for English-medium universities. – *Language Testing* 35 (2) s. 193–216. <https://doi.org/10.1177%2F0265532217703433>.
- IWASHITA, NORIKO – BROWN, ANNIE – MCNAMARA, TIM – O’HAGAN, SALLY 2008: Assessed levels of second language speaking proficiency: how distinct? – *Applied Linguistics* 29 (1) s. 24–49. doi:10.1093/applin/amm017.
- IWASHITA, NORIKO – GROVE, ELISABETH 2003: A comparison of analytic and holistic scales in the context of a specific-purpose speaking test. – *Prospect* 18 (3) s. 25–35.
- JONG, NIVJA DE – GROENHOUT, RACHEL – SCHOONEN, ROB – HULSTIJN, JAN H. 2015: Second language fluency. Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. – *Applied Psycholinguistics* 36 (2) s. 223–243. <https://doi.org/10.1017/S0142716413000210>.
- KANG, OKIM 2008: Ratings of L2 oral performance in English. Relative impact of rater characteristics and acoustic measures of accentedness. – *Spaan Fellow Working Papers in Second or Foreign Language Assessment* 6 s. 181–205.
- KORMOS, JUDIT – DÉNES, MARIANN 2004: Exploring measures and perceptions of fluency in the speech of second language learners. – *System* 32 (2) s. 145–164. <https://doi.org/10.1016/j.system.2004.01.001>.
- LIM, GAD S. 2011: The development and maintenance of rating quality in performance writing assessment. A longitudinal study of new and experienced raters. – *Language Testing* 28 (4) s. 543–560. <https://doi.org/10.1177%2F0265532211406422>.
- LUMLEY, TOM 2005: *Assessing second language writing. The rater’s perspective*. New York: Peter Lang.
- OCKEY, GARY J. 2009: The effects of group members’ personalities on a test taker’s L2 group

- oral discussion test scores. – *Language Testing* 26 (2) s. 161–186. <https://doi-org.ezproxy.jyu.fi/10.1177%2F0265532208101005>.
- ORR, MICHAEL 2002: The FCE Speaking test. Using rater reports to help interpret test scores. – *System* 30 (2) s. 143–154. [https://doi.org/10.1016/S0346-251X\(02\)00002-7](https://doi.org/10.1016/S0346-251X(02)00002-7).
- SAARINEN, TAINA – KAUPPINEN, MERJA – KANGASVIERI, TEIJA 2019: Kielikäsitukset ja oppimiskäsitykset koulutuspolitiikkaa linjaamassa. – Taina Saarinen, Pirkko Nuolijärvi, Sari Pöyhönen & Teija Kangasvieri (toim.), *Kieli, koulutus, politiikka. Monipaikkaisia käytänteitä ja tulkintoja* s. 121–148. Tampere: Vastapaino.
- SATO, TAKANORI 2011: The contribution of test takers' speech content to scores on an English oral proficiency test. – *Language Testing* 29 s. 223–241. <https://doi.org/10.1177-%2F0265532211421162>.
- TARNANEN, MIRJA 2002: *Arvioija valokeilassa. Suomi toisena kielenä -kirjoittamisen arviointia*. Jyväskylän yliopisto. <http://urn.fi/URN:ISBN:978-951-39-3419-4>.
- WEIGLE, SARA CUSHING 1999: Investigating rater/prompt interactions in writing assessment. Quantitative and qualitative approaches. – *Assessing Writing* 6 (2) s. 145–178. [https://doi.org/10.1016/S1075-2935\(00\)00010-6](https://doi.org/10.1016/S1075-2935(00)00010-6).
- WEIR, CYRIL J. 2005: *Language testing and validation. An evidence-based approach*. London: Palgrave Macmillan.
- ZHU, YU – FUNG, SHUI-LUNG – TSE, SHEK-KAM – HSIEH, CHI-YI 2017: Exploring the relationship between raters' personality traits and rating severity in writing assessment of Chinese as a second language. A pilot study. – Dongbo Zhang & Chin-Hsi Lin (toim.), *Chinese as a second language assessment* s. 203–223. Singapore: Springer. http://dx.doi.org/10.1007/978-981-10-4089-4_10.

Sari Ahola: Rimaa hipoen selviää tilanteesta. *Yleisten kielitutkintojen suomen kielen arvioijien käsityksiä kielitaidon arvioinnista ja suullisesta kielitaidosta*. JYU Dissertations 489. Jyväskylä: Jyväskylän yliopisto 2022. Väitöskirja on luettavissa osoitteessa <http://urn.fi/URN:ISBN:978-951-39-9005-3>.

Kirjoittajan yhteystiedot:
etunimi.sukunimi@jyu.fi