

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Halonen, Mia; Huhta, Ari; Ahola, Sari; Hirvelä, Tuija; Neittaanmäki, Reeta; Ohranen, Sari; Ullakonoja, Riikka

**Title:** Ensikielen tunnistamisen merkityksestä suullisen kielitaidon arvioinnissa Yleisissä kielitutkinnoissa

**Year:** 2020

**Version:** Published version

**Copyright:** © Kirjoittajat & Suomen soveltavan kielitieteen yhdistys ry, 2020

**Rights:** In Copyright

**Rights url:** <http://rightsstatements.org/page/InC/1.0/?language=en>

**Please cite the original version:**

Halonen, M., Huhta, A., Ahola, S., Hirvelä, T., Neittaanmäki, R., Ohranen, S., & Ullakonoja, R. (2020). Ensikielen tunnistamisen merkityksestä suullisen kielitaidon arvioinnissa Yleisissä kielitutkinnoissa. In S. Grasz, T. Keisanen, F. Oloff, M. Rauniomaa, I. Rautiainen, & M. Siromaa (Eds.), *Menetelmällisiä käännteitä soveltavassa kielentutkimuksessa - Methodological turns in applied language studies* (pp. 56-70). Suomen soveltavan kielitieteen yhdistys ry. AFinLA:n vuosikirja, 2020. <https://doi.org/10.30661/afinlavk.89453>

Grasz, S., T. Keisanen, F. Oloff, M. Rauniomaa, I. Rautiainen & M. Siromaa (toim.) 2020. Menetelmällisiä käännteitä soveltavassa kielentutkimuksessa – Methodological Turns in Applied Language Studies. *AFin-LAn vuosikirja 2020. Suomen soveltavan kielitieteen yhdistyksen julkaisu* n:o 78. Jyväskylä. s. 56–70.

**Mia Halonen, Ari Huhta, Sari Ahola, Tuija Hirvelä,  
Reeta Neittaanmäki, Sari Ohranen & Riikka Ullakonoja**  
Jyväskylän yliopisto

## **Ensikielen tunnistamisen merkityksestä suullisen kielitaidon arvioinnissa Yleisissä kielitutkinnoissa**

In this paper, we present a multidisciplinary study addressing fairness in the speaking test in a high-stakes language proficiency test in Finnish, National Certificates of Language Proficiency. The background of the research lies in studies on language assessment and (reversal) linguistic stereotyping and language attitudes. The focus L1 groups were Thai, Estonian, Finland Swedish, Arabic and Russian. Altogether 49 speech samples of test takers of these L1s were rated on a digital platform by 44 raters of the test system. The current paper reports on the sub-study that investigated whether the raters' recognition of the test takers' L1 affected their ratings and whether the effect differed across various assessment criteria.

**Keywords:** Proficiency rating, familiarity, L1 effect, L1 recognition

**Asiasanat:** Kielitaidon arviointi, tuttuus, ensikielen vaikutus, ensikielen tunnistaminen

## 1 Johdanto

Suomen kansalaisuutta hakevan on osoitettava kielitaitonsa läpäisemällä Yleisten kielitutkintojen (Yki) taitotaso 3 eli Eurooppalaisen viitekehyksen taso B1 suomesta tai ruotsista (Laki kansalaisuudesta 359/2003 ja Laki kansalaisuuslain muuttamisesta 579/2011 17 §). Tutkintojärjestelmä toimii siten merkittävänä yhteiskunnallisena portinvartijana, mistä syystä on tärkeä tutkia siinä tehtävän arvioinnin oikeudenmukaisuutta ja tasapuolisuutta. Testijärjestelmän tutkijat seuraavat arvioijien arviointikäyttäytymistä muun muassa tutkimalla tilastollisten analyysien avulla arvioijien lempeyttä tai ankaruutta ja johdonmukaisuutta. Vaikka arvioijia seurataan säännöllisesti, tarvitaan tutkinnon yhteiskunnallisen merkityksen vuoksi enemmän tietoa siitä, miten he käyttävät, esimerkiksi painottavat, analyttisiä kriteerejä antaessaan yleistaitosarviota puhumisesta tai mikä merkitys osallistujien taustamuuttujilla, esimerkiksi ensikielellä, on arviointikäyttäytymiseen. Tähän haasteeseen olemme pureutuneet hankkeessamme ”Rikkinäistä suomea: Aksenttien arviointi yhteiskunnallisena portinvartijana” (Jyväskylän yliopisto 2018). Tarkastelemme hankkeessa tutkinnon puhumisen osakokeen arviointikriteerejä ja niiden käyttöä suhteessa niin arvioijien kuin testin suorittajien taustatietoihin.

Kielitaidon arvioinnin tutkimuksessa arvioitavan aksentin tuttuus on ollut yksi keskeinen, moniselitteiseksi osoittautunut taustamuuttuja. Tutkimusten perusteella on selvää, että aksentin tuttuus luonnollisesti lisää puhujan tuotoksen todellista ymmärtämistä (‘intelligibility’), mutta samalla näyttää siltä, että aksentin tuttuus voi yhtä lailla johtaa arvioijan ymmärtämishalukkuuden, käsittämisen (‘comprehensibility’), vähenemiseen ja sitä kautta alempiin arvioihin kielitaidosta. (Derwing & Munro 1997; Carey ym. 2010.) Se kumpaan suuntaan tuttuus painottaa arviota, voi riippua siitä, millaiseksi tämä ”tuttu” yleisesti koetaan, liitetäänkö siihen positiivisia tai negatiivisia mielikuvia. Tällaisia kielteisesti arvioituja tuttuja ovat esimerkiksi espanjan- ja kiinanaksenttiset englanninpuhujat Yhdysvalloissa (Brennan & Brennan 1981; Cargile 1997; Hu & Lindemann 2009) ja ranskanaksenttiset englanninpuhujat niin Yhdysvalloissa (Lindemann 2005) kuin Kanadan englanninkielisillä alueilla (Reid ym. 2019) sekä venäjänaksenttiset suomenpuhujat Suomessa (esim. Toivola 2011).

Kielitaidon arviointi ei tutkimusten valossa olekaan vain kielitaidon vaan myös puhujan arviointia. Mielikuvat puhujista perustuvat usein stereotyyppiseen ja homogeeniseen kuvaan kokonaisesta puhujaryhmästä: etnisyydestä, kansalaisuudesta, maantieteellisestä taustasta tai muusta puhujille yhteiseksi mielletystä taustatekijästä. (Esim. Cameron 1995; Heller 2004; Halonen 2012.) Arvio puhujan taustasta tehdään ääntämisen, usein aksentiksi nimitetyn foneettisen poikkeaman (Flege 1988) eli syntyperäisen puheesta tavalla tai toisella eroavien piirteiden valossa. Syntyperäisen puhujan ääntämisnormi näyttää olevan ihmiselle erityisen merkittävä seikka. Sen osoittaa esimerkiksi se, että arvio tehdään yleensä nopeasti, jopa parin sekunnin näytteen perusteella. Lisäksi ei-syntyperäisen puhujan tuotos

voidaan tunnistaa sellaiseksi jopa kielestä, jota kuulija ei itse edes osaa (Major 2007). Myös lapset jo alle vuoden vanhoista vauvoista lähtien suosivat leikitovereinaan natiiviaksentilla puhuvia (Nesdale & Rooney 1996; Kinzler 2008). Toisaalta tämän herkkyyden vastapaino on, etteivät ihmiset useinkaan ole tottuneet tunnistamaan aksentin taustaa eli puhujan lähtökieltä. Edes ammattilaiset, esimerkiksi toisen tai vieraan kielen opettajat tai kielitaidon arvioijat, eivät välttämättä tunnista lähtökieltä maallikkoja paremmin. (Esim. Toivola 2011.)

Aiemman tutkimuksen valossa (esim. Toivola 2011: 107) kokeneidenkin arvioiden arviot saattoivat siis olla painottuneita tuttuuden, tunnistamisen ja tuttuuden taustalla vallitsevien mielikuvien ohjaamina, mikä olisi Kielitutkintojen oikeudenmukaisuuden näkökulmasta ongelmallista. Tässä artikkelissa tarkastelemme, miten ensikielen tunnistamista tutkittiin hankkeessamme. Tutkimuskysymyksemme on: Vaikuttaako oppijan ensikielen tunnistaminen hänelle annettuihin arvioihin ja jos vaikuttaa, vaihteleeeko vaikutus arviointikriteereittäin ja/tai kieliryhmittäin? Etenemme esittelemällä ensin lyhyesti Ykin tutkintojärjestelmänä ja suomen keskitason puhumisen osakokeen. Tämän jälkeen syvennymme yleiseen tutkimusasetelmaamme ja aineistonkeruu- ja tutkimusmenetelmiimme. Tulostemme esittelyn yhteydessä pohdimme niiden potentiaalisia selittäviä tekijöitä ja lopuksi arvioimme hiukan niiden implikaatioita.

## **2 Yleiset kielitutkinnot ja suomen kielen puhumisen osakoe**

Yleiset kielitutkinnot on aikuisille suunnattu kansallinen kielitutkintojärjestelmä, jonka hallinnoimisesta ja kehittämisestä vastaa Opetushallitus yhdessä Jyväskylän yliopiston kanssa. Tutkintojärjestelmässä testin suorittajalla on mahdollisuus osoittaa kielitaitonsa yhdeksässä eri tutkintokielessä. Tutkinnot testaavat sellaista jatkapäiväisessä elämässä tarvittavaa kielen hallintaa, joissa aikuinen kielenkäyttäjä joutuu puhumaan, kuuntelemaan, kirjoittamaan tai lukemaan toisella/vieraalla kielellä. Kielitaidon arviointia ohjaa funktionaalinen kielitaitokäsitys, jonka mukaan kieli nähdään sosiaalisena ilmiönä ja vuorovaikutuksen välineenä. Tutkinnot eivät perustu mihinkään tiettyyn oppimäärään tai opetussuunnitelmaan, vaan niitä ohjaavat tutkinnon perusteet (Opetushallitus 2011), joissa on määritelty tutkinnon sisällöt, arvioinnin perusteet, aihepiirit ja kielenkäyttötilanteet.

Yleisestä kielitutkinnosta on tarjolla kolme eri tasoa, perus- keski- ja ylin taso, joista kukin koostuu tekstin ja puheen ymmärtämisen, puhumisen ja kirjoittamisen osakokeesta. Testin suorittaja saa jokaisesta kokeesta erillisen taitotasoarvion, joka perustuu kuusiportaiseen asteikkoon (ks. taulukkoa 1). Arviointi tapahtuu kriteeriviitteisesti, mikä tarkoittaa sitä, että oppijan taitoa ei verrata toisiin oppijoihin, vaan arvio tehdään vertaamalla suoritusta sanallisiin kuvauksiin.

TAULUKKO 1. Yleisten kielitutkintojen rakenne.

Ykin ja Eurooppalaisen viitekehyksen taitotasot	Kolme tutkintotaso	Neljä taitoa/osakoetta
1 ~ A1	Perustaso	Tekstin ymmärtäminen Kirjoittaminen Puheen ymmärtäminen Puhuminen
2 ~ A2		
3 ~ B1	Keskitaso	
4 ~ B2		
5 ~ C1	Ylin taso	
6 ~ C2		

Vuosittain suomen keskitason tutkintoon (taitotasot 3 ja 4) osallistuu n. 8000 henkilöä, joista valtaosa (n. 80 %) käyttää tutkintotodistusta kielitaidon osoittamiseen Suomen kansalaisuutta hakiessaan. Tutkintotodistusta käytetään yleisesti myös työn tai opiskelupaikan haussa. Kansalaisuuteen vaadittava kielitaito edellyttää Eurooppalaisen viitekehyksen taitotasoa B1 (Yki 3) vähintään yhdessä suullisessa ja kirjallisessa taidossa (ks. migri.fi). Tutkintoon osallistuvien ensikielien vaihtelevat, ja heidän koulutustaustansa, kohdekielen opiskeluhistoriansa ja sosioekonomiset taustansa ovat erilaiset. Kymmenen suurinta tutkintoon osallistuvaa ensikieliryhmää tällä hetkellä ovat venäjä, arabia, kurdi, englantia, viro, somali, turkki, kiina, albania, espanja ja dari. Nämä ryhmittymät muodostavat myös suurimmat maahanmuuttajataustaiset ryhmät Suomessa (SVT 2018). Kaikkiaan tutkintoon on osallistunut yli 200:n kielen puhujia.

Hankkeen aineiston lähtökohtana oleva puhumisen osakoe koostuu neljästä tehtäväkokonaisuudesta, joissa testataan erilaisten kielenkäyttöfunktioiden hallintaa, esimerkiksi mielipiteiden ja tunteiden ilmaisemista, asioiden hoitamista ja tietojen antamista, selostamista ja kysymistä. Puhumisen osakoe suoritetaan kielistudiossa ja suoritukset tallennetaan arviointia varten. Arvioinnin suorittavat Opetushallituksen rekisteröimät arvioijat vertaamalla suoritusta puhumisen taitoa kuvaaviin kirjallisiin arviointikriteereihin. Kriteereissä puhumisen taito kuvataan kuuden analyyttisen kriteerin eli sujuvuuden, joustavuuden, koherenssin/sidosteisuuden, ilmaisun/sanaston tarkkuuden, ääntämisen ja fonologisen hallinnan sekä rakenteiden tarkkuuden ja holistisen yleiskriteerin avulla. Yleiskriteeri sisältää piirteitä analyyttisistä kriteereistä. Vaikka arvioijilla on käytössään arviointitilanteessa analyyttiset, jopa yksittäisiin kielellisiin piirteisiin perustuvat taidon kuvaukset, he antavat jokaisessa tehtävässä osallistujille suorituksesta vain yhden holistisen yleis-taitotasoarvion.

### 3 Aineisto, sen keruu ja tutkimusmenetelmät

Tuttuuden vaikutuksia tutkiaksemme meidän oli ensinnäkin keskityttävä arvioijillemme potentiaalisesti tuttuihin testisuorittajaryhmiin. Hankkeen aineisto kerättiin 2015 ja 2016. Suorittajien kieliä valitessa otimme huomioon osallistujien ensikielikohtaisen lukumäärän ja tutkitun tiedon asenteista eri kielten puhujaryhmiä kohtaan. Tutkimukseen valittiin neljä vuosina 2015–2016 suurimman maahanmuuttajakieliryhmän (SVT 2016) ja Yki-testin osallistujaryhmän kieltä: thai, viro, arabia ja venäjä. Lisäksi mukana on yksi suomalainen vähemmistökieli, Suomen toinen virallinen kieli, suomenruotsi. Tiesimme aiemman tutkimuksen (esim. McRae ym. 1998; Reuter & Kyntjä 2005; Sirkkilä 2005; Jaakkola 2007) perusteella, että kaikkien näiden viiden kieliryhmän puhujiin liitetään Suomessa vahvoja, usein kielteisiä asenteita ja stereotyyppioita. Lisäksi edellä jo esittelemämme tutkimuskirjallisuuden valossa tiesimme, että asenteet voivat vaikuttaa kielitaidon arviointiin.

Tavallisessa operationaalisessa Yki-arvioinnissa suorituksesta annetaan holistisen kokonaistaitotasoarvio, joka nojaa kuuteen analyttiseen arviointikriteeriin ja yleiskriteeriin. Se ei siis kerro yksittäisten kriteerien käytöstä, tai kriteerien käytön suhteesta toisiinsa. Se ei myöskään anna tietoa muiden taustatekijöiden kuten suorittajien ensikielen ja sen mahdollisen tunnistamisen suhteesta arviointikriteereihin. Koska pyrimme selvittämään tarkemmin, mihin suullisen kielitaidon arvioijat kiinnittävät huomiota tehdessään päätöksiä, keräsimme arviot myös analyttisten kriteerien kuvaamista taidoista yleistaitotason lisäksi. Muuten aineistonkeruussa imitoitiin normaalia arviointitilannetta mahdollisimman pitkälle.

Kieleen liittyvien mielikuvien tutkimiseksi Dennis Preston (1989: 3) ja Nancy Niedzielski ovat kehittäneet ns. kansanlingvistiikan, joka kuuluu havaintodialektologisiin tutkimusmenetelmiin. Heidän mukaansa (Niedzielski & Preston 2000: 45; myös Preston & Niedzielski 2013: 294), mikäli halutaan tutkia kieleen – ja siten sen puhujiin – kohdistuvien arvioiden painottuneisuutta, on kriittisen tärkeää tietää, mitä kuulija on oletanut arvioivansa. Tarvitsimme varsinaisten arviointitietojen ohella siis tiedon siitä, minkä arvioijat olettivat olevan suorittajan ensikieli ja millä perusteella. Tästä syystä osana aineistonkeruuta arvioijille esitettiin kysymys: ”Minkä oletat olevan puhujan ensikieli/lähtökieli?”. Oletukselle pyydettiin myös perustelu. Koska tavoitteena oli tuottaa suoraan Ykin laadunvarmistuksen ja kehittämisen kannalta käyttökelpoista tietoa, oli tärkeää tietää, mitä arvioijat todella olettavat todellisesta suorittajasta. Näin esimerkiksi suorittajien taustaa manipuloivat keinotekoiset koeasetelmat eivät olisi vastanneet tutkimuksen tavoitteisiin.<sup>1</sup>

1 Esimerkiksi asennetutkimuksissa usein käytetty niin kutsuttua valepukutekniikkaa (‘matched-guise’; Lambert 1967), jossa näytteitä alustetaan (‘priming’; ks. esim. Preston & Niedzielski 2013: 288, 289) väittämällä samalle näytteelle eri(taustaisia) puhujia, ei olisi ollut relevantti metodi tutkia arvioijien mahdollisia painotuksia.

Molempiin kysymyksiin pyydettiin vapaamuotoista avovastausta, jolla pyrittiin saamaan kiinni arvioijien mielikuvat ja oletukset virittämättä kysymyksenasetteluun esimerkiksi valmiiden kielivalikkojen muodossa mitään ennakko-oletuksia (Moyer 2013: 108–109). Vastauksen mittaa tai muotoa ei ollut rajoitettu. Arvioijat myös arvioivat oman oletuksensa varmuutta neliportaisella Likert-asteikolla: erittäin epävarma, melko epävarma, melko varma ja erittäin varma.

Tutkimistamme viidestä ensikieliryhmästä (thai, viro, suomenruotsi, arabia ja venäjä) oli kustakin mukana kymmenen suorittajaa, joista jokaiselta oli yksi näyte puhumisen osakokeen n. 1,5 minuutin kertomistehtävästä. Näytteet olivat autenttisia siinä mielessä, että ne olivat aitoja Yki-suorituksia. Ne oli arvioitu todellisessa Yki-arvioinnissa, joten käytössämme oli myös tieto siitä, mille taitotasolle arvioijat olivat suoritusarvioineet.

Näytteet arvioi 44 Ykin arvioijakoulutuksen käynnyttä arvioijaa verkkopohjaisella digitaalisella arviointialustalla verkkoselaimen kautta kolmessa arviointierässä. Suunnittelimme ja kehitimme arviointialustan tätä nimenomaista tutkimusta varten. Jokainen puhenäyte arvioitiin numeerisesti asteikolla 1–6 jokaisesta puhumisen kriteeristä sekä yleistaitotasoarvion. Arvioijia pyydettiin myös merkitsemään, mikäli jokin analyttisistä kriteereistä oli heidän mielestään erityisen merkittävä yleistaitotasoarvion suhteen. Lisäksi arvioijat esittivät oletuksia suorittajan ensikielestä, perustelivat oletuksensa ja esittivät muita vapaavalintaisia kommentteja suorittajasta.

KUVA 1. Kuvakaappaus alustasta valikkoineen.

Yleiskriteeri	Sujuvuus	Joustavuus	Koherenssi / sidosteisuus	Ilmaisun tarkkuus / laajuus / idiomatisuus	Ääntäminen / fonologinen hallinta	Rakenteiden tarkkuus
Valitse...	Valitse...	Valitse...	Valitse...	Valitse...	Valitse...	Valitse...
<input type="checkbox"/> Erityisen merkittävä	<input type="checkbox"/> Erityisen merkittävä	<input type="checkbox"/> Erityisen merkittävä	<input type="checkbox"/> Erityisen merkittävä	<input type="checkbox"/> Erityisen merkittävä	<input type="checkbox"/> Erityisen merkittävä	<input type="checkbox"/> Erityisen merkittävä

Yleistaitotasoarvio: Valitse...  
 1  
2  
3  
4  
5  
6

Kommentoi puhujaa / arvioitasi:

Minkä oletat olevan puhujan ensikieli / lihtökieli? Valitse...  
 Kuinka varma olet oletuksestasi: Valitse...  
 1 = Erittäin epävarma  
2 = Epävarma  
3 = Varma  
4 = Erittäin varma

Perustele oletuksesi ensikielestä / lihtökielestä:

Tallenna arvio ja siirry seuraavaan

Koko alusta oli jokaisen näytteen kohdalla täytettävissä missä tahansa järjestyksessä, ja alusta tallensi tiedon täyttäjärjestyksestä. Arvioijien itsensä tuottaman informaation lisäksi alusta tallensi automaattisesti arviointiprosessiin liittyviä ajankäyttötietoja: kuinka paljon arviointiin kaikkiaan käytettiin aikaa, kuinka paljon aikaa käytettiin

näytettä kohti, kuinka pian näytteen kuuntelun aloittamisesta ensimmäinen arvio annettiin ja missä järjestyksessä analyyttisiä kriteerejä käytettiin. Koska alusta tallensi täyttöjärjestyksen ja -ajan, pääsimme näkemään esimerkiksi, että ensikieli oli usein kysymys, johon vastattiin ensimmäisenä, vaikka se sijaitsi alustalla loppupäässä. Siihen vastattiin ensimmäisenä 23 prosentissa kaikista arvioinneista, eli se vaikutti olevan osalle arvioijista paitsi merkittävä myös helposti tunnistettava suorittajan ominaisuus.

Kaikkiaan aineistoa kertyi näin yli 100 000 datapistettä. Jotta saatua monimuotoista, laadullista ja määrällistä, tietoa arvioijista, osallistujista, kriteeristä, ensikielestä ja sen tunnistamisesta pystyttiin tutkimaan näin suuresta datamäärästä, tutkimuksessa tarvittiin laadullisten menetelmien lisäksi tilastollisia menetelmiä. Tilastollisten menetelmien valintaan vaikuttivat tutkimuskysymyksen lisäksi aineiston koko ja sen muuttujien, erityisesti arviointikriteerien ja yleistaitotason järjestysasteikollinen luonne. Lisäksi valintaa ohjasivat myös yleisesti arviointitutkimuksessa käytössä olevat menetelmät (Myford & Wolfe 2003, 2004; Eckes 2011). Näistä mainittakoon erityisesti arviointiaineiston tutkimiseen kehitetty MFRM-analyysi (Multi-Facet Rasch Measurement; Linacre 2007) sekä ordinaalinen regressioanalyysi (McCullagh 1980). Esimerkiksi MFRM-analyysin avulla tutkimme mm. esiintyykö arvioijien toiminnassa erilaisia painotuksia oppijoiden ensikielen ja sen tunnistamisen suhteen sekä miten eri arviointikriteerit toimivat. Ordinaalinen regressioanalyysi puolestaan mahdollisti tilastollisen mallin rakentamisen, jonka avulla tutkimme useiden eri tekijöiden, sellistävien muuttujien – arviointikriteerien, ensikielen tunnistamisen ja arvioijien sekä oppijoiden taustamuuttujien – ja niiden mahdollisten yhdysvaikutusten vaikutusta selitettävään muuttujaan, suorittajien saamiin arvioihin. Samanlaista menetelmää ovat käyttäneet esimerkiksi toisella kielellä kirjoittamisen arviointia tutkineet Hyun Jung, Junkye ja Hyun-Jo (2019).

Tilastollisia analyyseja varten laadullinen vastausdata taas oli luokiteltava sen mukaan, tunnistiko arvioija ensikielen oikein, ja jollei, mikä oli hänen oletuksensa ensikielestä. Ensin yksi tutkija luokitteli ne oletettujen kielten mukaan, minkä jälkeen hän luokitteli nämä kielet suhteessa tunnistamiseen: oliko oletettu kieli suorittajan kieli eli tunnistettu vai ei eli ei-tunnistettu. Jotta voitiin varmistaa näiden luokkien pitävyys ja yhdenmukaisuus, kaksi muuta tutkijaa tekivät saman luokittelun vielä erikseen.

Itse hanketta varten keräämämme aineiston lisäksi asetelmassamme käytetään puhumisen osakoetta koskevaa numeerista tietoa Yleisten kielitutkintojen laajasta pitkittäisaineistosta ajalta 2009–2019. Paitsi että tätä aineistoa tutkitaan myös sellaisenaan ja rinnan digitaalisesti keräämämme arviointiaineiston kanssa, pitkittäisaineisto myös tukee tätä tutkimusta varten kerätyn aineistoanalyysin tuloksia.



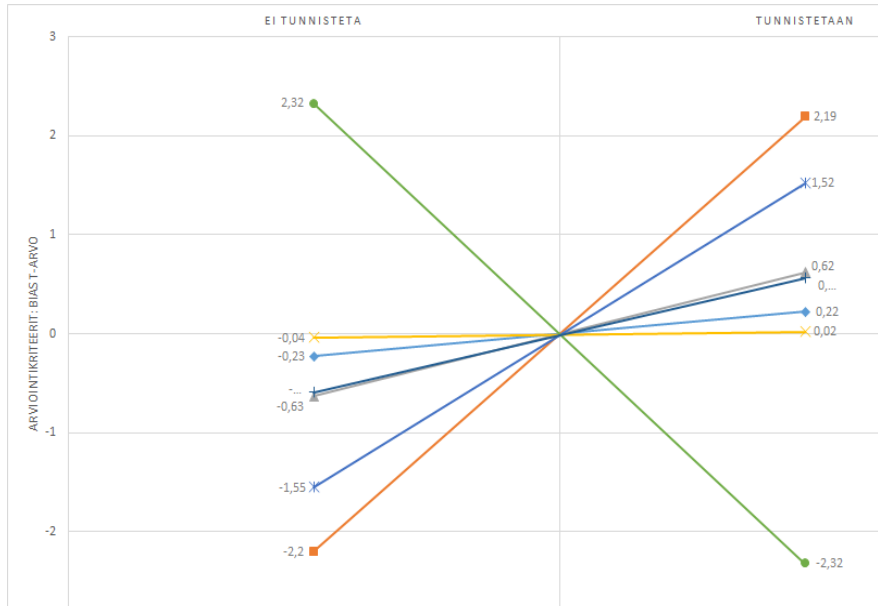
#### 4 Tuloksia: Ensikielen tunnistaminen laskee ääntämisen ja nostaa sujuvuuden arviota

Seuraavassa tarkastelemme ensikielen tunnistamisen vaikutusta koko tutkimusaineistossamme, jossa on siis mukana kaikkien arvioijien kriteerittaiset arviot kaikista suorittajista sen suhteen, ovatko he tunnistaneet suorittajan ensikielen vai eivät.

Kuviossa 1 näkyy koko aineistoamme koskevan ensikielen tunnistamisen ja arviointikriteerien MFRM-malliin pohjautuvan bias-analyysin tulokset. Bias-analyysissa lasketaan arviointikriteerien ja ensikielen tunnistamisen yhdysvaikutusparametreille estimaatit logit-skaalalle. Nollaa suuremmat bias-estimaatit tarkoittavat, että havaitut arvot ovat suurempia kuin mallin odottamat ja vastaavasti nollaa pienemmät arvot, että havaitut arvot ovat pienempiä kuin mallin odottamat arvot. Bias-estimaattien suuruuden tilastollista merkitsevyyttä testataan t-jakaumaan perustuvalla t-testisuureella eli yleisesti t-arvojen avulla (kuviossa *bias t-arvo*).

Kuva jakautuu keskeltä kahtia: sen vasemmalla puolella näkyy t-arvot niiden puhujien näytteistä, joiden ensikieltä ei tunnistettu (49 %), oikealla taas t-arvot oikein tunnistetuista (51 %) näytteistä. Arviointikriteerit – holistinen ns. yleiskriteeri, sujuvuus, joustavuus, koherenssi/sidosteisuus, ilmaisun/sanaston tarkkuus, ääntäminen ja fonologinen hallinta sekä rakenteiden tarkkuus – näkyvät kuvassa erivärisinä vektoreina (ks. kriteerejä vastaavat värit kuvan alareunasta).

Olennaista kuvion tulkitsemisessä on kriteereiden (vektoreiden) suunta ja t-arvojen merkitsevyys. Tilastollisesti merkitseviä (merkitsevyydellä 0,05) ovat t-arvot (kuvassa y-akselilla), jotka ylittävät arvon +2 tai alittavat arvon -2. Kuvasta voi heti nähdä, että ensikielen tunnistaminen vaikuttaa ääntämisen (vihreä vektori) ja sujuvuuden (oranssi vektori) arviointiin: ääntämisen arviot laskevat tunnistamisen myötä, ja sujuvuuden arvot puolestaan nousevat sen myötä. Ensikielen tunnistaminen ei vaikuta tilastollisesti merkitsevästi muista arviointikriteereistä saatuihin arviointeihin, vaikkakin niiden trendi näyttää nousua tunnistamisen myötä.



KUVIO 1. Ensikielen tunnistamisen ja arviointikriteerien bias-analyysin tulokset.

Vastauksena tutkimuskysymykseemme voimme todeta, että osallistujien ensikielen tunnistaminen vaikuttaa arviointiin. Kriteereistä kuitenkin vain kaksi, ääntäminen ja sujuvuus, näyttää reagoivan merkitsevästi sekä tilanteeseen, jossa suorittajan ensikieltä ei tunnisteta, että tilanteeseen, jossa se tunnistetaan. Aineiston kokonaiskuvassa ääntämistä siis arvioitiin lempeämmin, kun suorittajan ensikieltä ei tunnistettu. Tunnistamattomuus ja siitä seuraava ääntämisen arvion lempeys voi viitata ainakin kahteen tekijään. Ensinnäkin ääntäminen voi olla niin hyvää, kohdekielenomaista eli suomenmukaista, että siinä ei juuri erotu muiden kielten piirteitä. Toiseksi on mahdollista, että osalle arvioijista suorittajan ensikieli ja/tai sen mahdolliset siirtovaikutukset ovat niin vieraita, etteivät he kiinnitä niihin mitään huomiota eivätkä mahdollisesti edes tulkitse niitä poikkeamiksi ääntämisnormista. Ensikieli tunnistettaessa taas on selvää, että suoritus sisältää sellaisia ensikielen elementtejä, jotka ovat selviä poikkeamia suomen kielen ääntämisnormiin ja voivat näin laskea ääntämisen arviota. Ääntämisen heikompi hallinta voi siis paljastaa helpommin suorittajan ensikielen. Sama toimii myös päinvastaiseen suuntaan: ensikielen siirtovaikutus voi heikentää sekä itse ääntämistä että siitä annettavaa arviota.

Sujuvuuden kriteerin suhteen kuva on juuri päinvastainen: kun osallistujan ensikieltä ei tunnisteta, suorituksen sujuvuus arvioidaan alemmas kuin muut taidot, ja kun tunnistetaan, ylemmäs. On kiinnostavaa, että juuri ääntäminen ja sujuvuus

menevät tunnistamisen ja arvioiden suunnan suhteen keskenään ristiin. On mahdollista, että tämä johtuu siitä, että kun suorittajan ensikieli tunnistetaan usein heikommasta ääntämisestä, tuo heikompi ääntäminen sitten myös johtaa oletukseen yleisemmin heikosta kielitaidosta. Tämän valossa taas pienemmillekin merkeille sujuvuudesta, esimerkiksi puheen jatkuvalle virralle, annetaan enemmän painoa.

Molemmat tunnistamisen kannalta merkitseviksi osoittautuneet taidot, ääntäminen ja sujuvuus, ovat kielitaidon arvioinnin ytimessä, vaikkakin eri tavalla. Ääntäminen on sitä jo yksinkertaisesti siksi, ettei mitään puhetta voi tuottaa ääntämättä. Puhe ja siis ääntäminen paljastaa puhujastaan välittömästi esimerkiksi iän, sukupuolen ja mahdollisen vieraan aksentin ja muita seikkoja, jotka vaikuttavat kokemukseen puhujasta ja kielitaidon tasosta. Sujuvuus taas katsotaan kielitaidon ytimeksi, sillä termiä "sujuva" käytetään yleisesti kuvaamaan puhujaa, jolla on kokonaisuudessaan hyvä kohdekielen osaaminen (Davies ym. 1999: 64). Tätä tukee myös se, että muutamissa tutkimuksissa (esim. Iwashita ym. 2008; Duijm ym. 2018) sujuvuuden on todettu vaikuttavan paljon suullisen kielitaidon kokonaisarvioon. Sujuvuus on yleisesti käytetty arviointikriteeri, mutta kuten monen muunkin kriteerin kohdalla, sille ei ole vain yhtä ainoaa virallista määritelmää. Testijärjestelmissä toimivat tutkijat määrittelevät kriteerin sisällön kukin omista lähtökohdistaan, joten sisältö tulisi aina tarkistaa tapauskohtaisesti (Huhta ym. 2019: 129–130).

Ääntämisen osalta tunnistamisella näytti olevan suora suhde arvioituun taitotasoon: kun suorittajan ensikieli tunnistettiin, ääntäminen arvioitiin alemmaksi. Toisaalta tunnistaminen ja ääntäminen näyttävät olevan vuorovaikutuksessa niin, että yhtä lailla heikompi ääntäminen eli sellainen, jossa kuuluu ensikielen vaikutus, taas, luonnollisesti, helpottaa tunnistamista. Kielitaito ja etenkin ääntäminen ja ensikielen tunnistaminen kulkevat siis osittain kehänä. Esimerkiksi aineistomme ensikieleltään thainkieliset olivat kokonaisuudessaan aineiston heikoimmin suoriutuva ryhmä, ja heidät myös tunnistettiin 56 prosentissa arviointitapahtumista, kun taas ryhmänä vahvimmat puhujat, suomenruotsalaiset, tunnistettiin vain 18 %:ssa tapauksista. Väärin tunnistettaessa suomenruotsalaiset luokiteltiin useimmin virolaisiksi (22 % tapauksista), mikä selittyyneen muutamilla suomenruotsinkielisten ja vironkielisten samanlaisilla siirtovaikutuksilla suomessa, esimerkiksi ns. retrofleksisellä /-äänneellä (Kuronen 2016). Vironkielisiksi arvattaessa suomenruotsalaiset myös saivat alempia ääntämisen arvioita kuin oikein tunnistettaessa. Toisaalta taas hyvin suomea puhuneet vironkieliset tunnistettiin kuitenkin parhaiten, 78 %:ssa tapauksista. Vironkielisillä onkin puhumassaan suomessa piirteitä, jotka auttavat kuulijaa tunnistamaan heidän ensikielensä, erityisesti kun kyse on Yki-arvioijien kaltaisista kielen ammattilaisista, jotka ovat mahdollisesti opettaneet suomea virolaisille ja joista osa on myös itse opiskellut viroa. Ensikielen tunnistaminen oikein, ja toisaalta puhujan tunnistaminen väärin jonkin toisen puhujaryhmän edustajaksi on merkittävä tekijä arvioinnin ja ensikielen tunnistamisen suhteen tutkimuksessa. Siihen emme kuitenkaan syvenny tämän enempää tässä artikkelissa.

Vaikka tilastollisesti merkitseviä eroja havaittiin kahdenkin kriteerin arvioinnissa silloin, kun ensikieli tunnistettiin, niillä oli kuitenkin hyvin pieni vaikutus lopullisiin yleistaitotason arvioihin. Joissakin tapauksissa analyyttisten kriteerien numeerinen arvio on ennakoanut heikkoa arviota, mutta lopullinen taitotasoarvio poikkesi testin suorittajan hyväksi ylöspäin. (Thainkielisten osalta tästä, ks. Ahola, painossa.) Poikkeama saattaa kuvata arvioijien kykyä korjata mahdollisesti painottunutta arviointiaan, kykyä itsereflektioon ja kokonaiskielitaidon painottamiseen. Samalla se saattaa kertoa kriteerien luonteesta ja käyttökelpoisuudesta. Yksittäiset analyyttiset kriteerit eivät koskaan pysty kuvaamaan varsinaista kielitaidon monimutkaista kokonaisuutta.

## 5 Lopuksi

Olemme tässä artikkelissa tarkastelleet, vaikuttaako ensikielen tunnistaminen kielitaidon arviointiin Ykin suomen kielen puhumisen testissä ja miten olemme tätä hankkeessamme tutkineet. Tässä artikkelissa tarkastelimme tätä kysymystä koko aineistomme, kaikkien viiden fokuskielen ja näiden 49:n suorittajan ja kaikkien 44:n arvioijan arvioiden valossa. Analyysimme osoittaa, että tunnistaminen vaikutti merkittävästi ääntämisen ja sujuvuuden arviointiin. Yleistaitotaso ei kuitenkaan muuttunut minkään yksittäisen analyyttisen kriteerin ohjaamana tai muodostunut niiden keskiarvosta, vaan arvioijien voi kuvata korjanneen mahdollisen kumpaan tahansa suuntaan painottuneen arvionsa yleistaitotasoarviossaan. Tämä on luonnollisesti yleisten kielitutkintojen luotettavuuden ja oikeudenmukaisuuden kannalta olennaisinta. Yleistaitotasoarvionhan arvioijat antavat normaalissakin Yki-arvioinnissa, jossa suorituksia ei arvioida analyyttisten kriteerien tasolla. Tältä osin tutkimustulostemme voikin sanoa kertovan eniten analyyttisten kriteerien luonteesta ja käyttökelpoisuudesta arvioinnissa; arvioijat näyttävät käyttävän niitä kuin irrallaan lopullisesta tasoarviostaan. Tarkempi tekeillä olevan tutkimuksemme kriteerien käytöstä voi kuitenkin tuoda tähän uudenlaista valoa.

Vaikka itse Ykin tutkintojärjestelmän kannalta aineistomme tulokset eivät siis näytä tasapuolisuuden suhteen huolestuttavilta, on tunnistamisen vaikutus ääntämisen taidon matalampiin arvioihin olennainen tieto. Ensikielen tunnistaminen tarkoittaa aina sitä, että kielenpuhujaryhmä on arvioijalle tavalla tai toisella niin tuttu, että hän voi tunnistaa sen 1,5 minuutin mittaisesta puhenäytteestä. Tuttuus kantaa mukanaan monenlaista tietoa niin puhujan kielestä kuin tämän muustakin taustasta. Tieto voi liittyä niin ammattikokemuksiin kielitaidon arvioijana ja usein myös kielen opettajana kuin maallikkokokemuksiin kielenpuhujaryhmästä joko omina suorina kontakteina tai vaikkapa median välityksellä. Tutkimusasetelmamme suoritajaryhmiin liitetään Suomessa monenlaisia käsityksiä, jotka voivat osaltaan olla

vaikuttamassa myös kielitaidon kokemiseen ja sen arvioimiseen, myös Ykin ammat-  
tiarvioijilla.

Ääntämisen osalta tunnistamisella näytti olevan suora suhde arvioituun taitota-  
soon: kun suorittajan ensikieli tunnistettiin, ääntämisestä annettiin alempia arvioita.  
Tämä tunnistamisen vaikutus viittaa siihen, että tunnistamisella itsellään on oma,  
taidosta riippumaton vaikutuksensa arviointiin. Tätä tulkintaa tukee myös aiempi  
tutkimus esimerkiksi työhönottohaastatteluista, joissa on osoitettu tunnistettavan,  
etenkin kielteisesti stereotypisoidulle puhujaryhmälle kuuluvan aksentin, vaikut-  
tavan työnantajien halukkuuteen palkata hakija. (Esim. Kokkonen 2007; Hosoda &  
Stone-Romero 2010; Deprez & Sims 2013.) Tällaisten epäkohtien vähentämiseksi  
Hosoda ja Stone-Romero (2010) ehdottavat, että työnhakijoiden tausta esitettäisiin  
haastattelutilanteessa avoimesti ja haastattelihoita koulutettaisiin ymmärtämään ja  
sietämään erilaisia aksentteja. Deprez ja Sims (2013) puolestaan tarjoavat ratkai-  
suksi sitä, että haastattelihoiksi rekrytoitaisiin ihmisiä monista erilaisista taustoista.  
Kielitaidon arvioinnin puolella Winke, Gass ja Myford (2013) taas ovat tulok-  
seen, että arvioijan yhteinen kielitausta suorittajan kanssa tuottaa lempeämpiä ar-  
vioita.

Samalla on taas syytä muistaa, että tuttuus on tekijä, joka voi yhtä lailla johtaa  
suopeaan kuin ankaraankin arvioon. Tämä näkyy esimerkiksi siinä, kuinka ankarasti  
ihmiset usein arvioivat ”omiaan”, vaikkapa suomalaiset tunnistaessaan suomalaisen  
englanninaksentin (ks. Leppänen ym. 2010: 76–78). Myöskään painotuksen suun-  
nan muuttaminen kielteisestä ja siten usein ankarammasta arvioinnista myönteis-  
seen ja usein siis lempeään ei ole yhtään kestävämpi ratkaisu, kun tavoitteena on  
kaikenlaisten kielitaidon ulkopuolisten painotusten karsiminen. Ratkaisuehdotuksia  
varsinaisen kielitaidon ulkopuolelta tuleviin painotuksiin on siis useita, ja ne ovat  
keskenään ristiriitaisiakin. Tälläkin hetkellä Ykissä käytössä olevat arvioijapalautte,  
arvioijien koulutus ja johdonmukaisuuden ja tasapuolisuuden monitorointi ainakin  
ovat olennaisia järjestelmän luotettavuuden ja oikeudenmukaisuuden varmistami-  
sen menetelmiä.

## Kirjallisuus

- Ahola, S. painossa. Yleisten kielitutkintojen arvioijien käsityksiä thainkieliseksi tunnistettujen suomenoppijoiden suullisesta kielitaidosta. *Puhe ja kieli*.
- Brennan, E. & J. S. Brennan 1981. Measurements of accent and attitude toward Mexican-American speech. *Journal of Psycholinguistic Research*, 10 (5), 487–501. <https://doi.org/10.1007/BF01076735>.
- Cameron, D. 1995. *Verbal hygiene*. Florence, KY, USA: Routledge.
- Carey, M. D., R. H. Mannell & P. K. Dunn 2010. Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, 28 (2), 201–219. <https://doi.org/10.1177/0265532210393704>.
- Cargile, A. C. 1997. Attitudes toward Chinese-accented speech: an investigation in two contexts. *Journal of Language and Social Psychology*, 16, 434–443. <https://doi.org/10.1177/0261927X970164004>.
- Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley & T. McNamara 1999. *Dictionary of language testing*. Cambridge: Cambridge University Press.
- Deprez-Sims, A. & S.B. Morris 2013. The effect of non-native accents on the evaluation of applicants during an employment interview: the development of a path model. *International Journal of Selection and Assessment*, 21, 355–367. <https://doi.org/10.1111/ijsa.12045>.
- Derwing, T. & M. Munro 1997. Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition*, 20, 1–16. <https://www.jstor.org/stable/44488664>.
- Duijm, K., R. Schoonen & J. Hulstijn 2018. Professional and non-professional raters' responsiveness to fluency and accuracy in L2 speech: an experimental approach. *Language Testing*, 35 (4), 501–527. <https://doi.org/10.1177/0265532217712553>.
- Eckes, T. 2011. *Introduction to many-facet Rasch measurement. Analyzing and evaluating rater-mediated assessments*. Frankfurt am Main: Peter Lang.
- Flege, J. E. 1988. The production and perception of foreign language speech sounds. Teoksessa H. Winitz (toim.) *Human communication and its disorders*. Norwood, N. J.: Ablex, 224–401.
- Halonen, M. 2012. Rikkinäinen ja likainen: kieli-ideologiset prosessit kielentutkimuksen diskursseissa. *Virittäjä*, 116 (3), 443–462. Noudettu 8.6.2020 osoitteesta <https://journal.fi/virittaja/article/view/7163>.
- Hosoda, M. & E. Stone-Romero 2010. The effects of foreign accents on employment-related decisions. *Journal of Managerial Psychology*, 25, 113–132. <https://psycnet.apa.org/doi/10.1108/02683941011019339>.
- Hu, G. & S. Lindemann 2009. Stereotypes of Cantonese English, apparent native/non-native status, and their effect on non-native English speakers' perception. *Journal of Multilingual and Multicultural Development*, 30, 253–269. <https://doi.org/10.1080/01434630802651677>.
- Huhta A., H. Kallio, S. Ohranen & R. Ullakonoja. 2019. Fluency in language assessment. Teoksessa P. Lintunen, M. Mutta & P. Peltonen (toim.) *Fluency in L2 learning and use*. Blue Ridge Summit: Multilingual Matters, 129–145.
- Hyun Jung, K., L. Junkye & Y. Hyun-Jo 2019. Analysis of rater effect in the evaluation of second language grammatical knowledge in the context of writing: application of a generalized linear model. *Linguistic Research*, 36, 25–57. <https://doi.org/10.17250/khisli.36..201909.002>.
- Iwashita, N., A. Brown, T. McNamara & S. O'Hagan 2008. Assessed levels of second language speaking proficiency: how distinct? *Applied Linguistics*, 29 (1), 24–49. <https://academic.oup.com/applij/article-abstract/29/1/24/258796>.

- Jaakkola, M. 2009. *Maahanmuuttajat suomalaisten näkökulmasta. Asennemuutokset 1987–2007*. Helsinki: Helsingin kaupunki.  
[https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/09\\_02\\_19\\_Tutkimus\\_Jaakkola.pdf](https://www.hel.fi/hel2/Tietokeskus/julkaisut/pdf/09_02_19_Tutkimus_Jaakkola.pdf) [luettu 8.6.2020].
- Jyväskylän yliopisto 2018. *Rikkinäistä suomea: aksenttien arviointi yhteiskunnallisena portinvartijana -hanke* (Suomen Akatemia 2018–2022)  
<https://www.jyu.fi/hytk/fi/laitokset/solki/broken-finnish> [luettu 29.9.2020].
- Kinzler, K. 2008. *The native language of social cognition: developmental origins of social preferences based on language*. Julkaisematon väitöskirja. Cambridge, MA: Harvard University.
- Kokkonen, M. 2007. Vaatimuksena sujuva suomi. *Virittäjä*, 111, 253–261. Noudettu 8.6.2020 osoitteesta <https://journal.fi/virittaja/article/view/40570>.
- Kuronen, M. 2016. Uttal av S2-finska med fokus på svenskspråkiga talare. *Puhe ja kieli*, 36 (3), 147–174. Noudettu 8.6.2020 osoitteesta <https://journal.fi/pk/article/view/59007>.
- Laki kansalaisuudesta 359/2003. <https://finlex.fi/fi/laki/smur/2003/20030359> [luettu 29.9.2020].
- Laki kansalaisuuslain muuttamisesta 579/2011 17 §.  
<https://www.finlex.fi/fi/laki/alkup/2011/20110579> [luettu 29.9.2020].
- Lambert, W. E. 1967. A social psychology of bilingualism. *Journal of Social Issues*, 23 (2), 91–109. <https://doi.org/10.1111/j.1540-4560.1967.tb00578.x>.
- Leppänen, S., A. Pitkänen-Huhta, T. Nikula, S. Kytölä, T. Törmäkangas, K. Nissinen, L. Kääntä, T. Räisänen, M. Laitinen, H. Koskela, S. Lähdesmäki & H. Jousmäki 2010. *National survey on the English language in Finland: uses, meanings and attitudes*. Helsinki: Studies in variation, contacts and change in English, e-Series.  
<http://www.helsinki.fi/varieng/series/volumes/05/evarieng-vol5.pdf> [luettu 8.6.2020.]
- Linacre, J. M. 2007. *A user's guide to WINSTEPS Rasch-model computer programs*. Chicago, Illinois: winstep.com, MESA Press. <https://www.winsteps.com/winman/copyright.htm> [luettu 8.6.2020].
- Lindemann, S. 2005. Who speaks "broken English"? US undergraduates' perceptions of non-native English. *International Journal of Applied Linguistics*, 15, 187–212.  
<https://psycnet.apa.org/doi/10.1111/j.1473-4192.2005.00087.x>.
- Major, R. C. 2007. Identifying a foreign accent in unfamiliar language. *Studies in Second Language Acquisition*, 29, 539–556. <https://doi.org/10.1017/S0272263107070428>.
- McCullagh, P. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42, 109–142.  
<http://www.stat.uchicago.edu/~pmcc/pubs/paper2.pdf> [luettu 8.6.2020].
- McRae, K. D., S. E. Bennett & T. Miljan 1988. *Intergroup sympathies and language patterns in Finland: results from a survey*. Helsinki: Suomen Gallup.
- Moyer, A. 2013. *Foreign accent. The phenomenon of non-native speech*. Cambridge: Cambridge University Press.
- Myford, C. & E. W. Wolfe 2003. Detecting and measuring rater effects using many-facet Rasch measurement: part I. *Journal of Applied Measurement*, 4, 386–422.
- Myford, C. & E. W. Wolfe 2004. Detecting and measuring rater effects using many-facet Rasch measurement: part II. *Journal of Applied Measurement*, 5, 189–227.
- Nesdale, D. & R. Rooney 1996. Evaluations and stereotyping of accented speakers by pre-adolescent children. *Journal of Language & Social Psychology*, 15, 133–155.  
<https://doi.org/10.1177/0261927X960152002>.
- Niedzielski, N. A. & D. R. Preston 2000. *Folk linguistics*. Berlin: Mouton De Gruyter.
- Opetushallitus 2011. *Yleisten kielitutkintojen perusteet*. Helsinki: Opetushallitus.

- Preston, D. R. 1989. *Perceptual dialectology: nonlinguists' views of areal linguistics*. Dordrecht, Holland: De Gruyter Mouton.
- Preston, D. R. & N. Niedzielski 2013. Approaches to the study of language regard. Teoksessa T. Kristiansen & S. Grondelares (toim.) *Language (de)standardization in late modern Europe: experimental studies*. Oslo: Novum, 287–306. [https://lanchart.hum.ku.dk/research/slice/publications-and-news-letters/publications/language-destandardisation-in-late-modern-europe/p\\_287-306\\_Preston\\_and\\_Niedzielski.pdf](https://lanchart.hum.ku.dk/research/slice/publications-and-news-letters/publications/language-destandardisation-in-late-modern-europe/p_287-306_Preston_and_Niedzielski.pdf) [luettu 8.6.2020].
- Reid, K. T., P. Trofimovich & M. G. O'Brien 2019. Social attitudes and speech ratings: effects of positive and negative bias on multiage listeners' judgments of second language speech. *Studies in Second Language Acquisition*, 4 (2), 419–442. <https://doi.org/10.1017/S0272263118000244>.
- Reuter, A. & E. Kyntäjä 2005. Kansainvälinen avioliitto ja stigma. Teoksessa T. Martikainen (toim.) *Etnisyys Suomessa 2000-luvulla*. Helsinki: Suomalaisen Kirjallisuuden Seura, 104–125. [http://vaestoliitto-fi-bin.directo.fi/@Bin/80027205f047540d08d688bb5b4d8ffd/1591608061/application/pdf/386433/Maahanmuuttajanaiset\\_e.pdf](http://vaestoliitto-fi-bin.directo.fi/@Bin/80027205f047540d08d688bb5b4d8ffd/1591608061/application/pdf/386433/Maahanmuuttajanaiset_e.pdf) [luettu 8.6.2020].
- Sirkkilä, H. 2005. *Elättäjyyttä vai erotiikkaa. Miten suomalaiset miehet legitimoivat parisuhteensa thainmaalaisen naisen kanssa?* Jyväskylä: Jyväskylä Studies in Education, Psychology and Social Research 268. <http://urn.fi/URN:ISBN:951-39-2180-8>.
- SVT = Suomen virallinen tilasto (SVT) 2016. *Väestörakenne* [verkkójulkaisu]. ISSN=1797-5379. 2016, Liitekuvio 2. Suurimmat vieraskieliset ryhmät 2006 ja 2016. Helsinki: Tilastokeskus. [https://www.stat.fi/til/vaerak/2016/vaerak\\_2016\\_2017-03-29\\_kuv\\_002\\_fi.html](https://www.stat.fi/til/vaerak/2016/vaerak_2016_2017-03-29_kuv_002_fi.html) [luettu 8.6.2020].
- SVT=Suomen viralliset tilastot (SVT) 2018. *Vieraskieliset Suomessa* 2018. Helsinki: Tilastokeskus. <https://www.stat.fi/tup/maahanmuutto/maahanmuuttajat-vaestossa/vieraskieliset.html> [luettu 8.6.2020].
- Toivola, M. 2011. *Vieraan aksentin arviointi ja mittaaminen suomessa*. Helsinki: Helsingin yliopisto.
- Winke, P. S. Gass & C. Myford. 2013. Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30 (2), 231–252. <https://doi.org/10.1177/0265532212456968>.