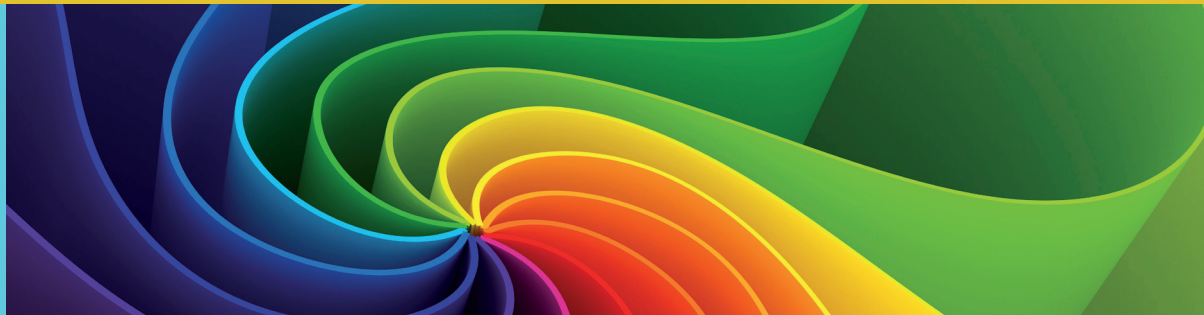


Toimittaneet
Ari Huhta & Raili Hildén



Kielitaidon arviointitutkimus 2000-luvun Suomessa

AFinLA-e

Soveltavan kielitieteen tutkimuksia 2016/n:o 9

ISBN 978-951-9388-64-9

ISSN 1798-7822

AFinLA-e
Soveltavan kielitieteen tutkimuksia 2016/n:o 9

Kielitaidon arviointitutkimus 2000-luvun Suomessa

Toimittaneet
Ari Huhta & Raili Hildén



Suomen soveltavan kielitieteen yhdistys AFinLA, Jyväskylä 2016
© kirjoittajat ja Suomen soveltavan kielitieteen yhdistys AFinLA

Sisältö

Esipuhe

Ari Huhta & Raili Hilden

- ☒ Kielitutkinnot ja muu laajamittainen kielitaidon arviointi Suomessa 3

Henna Tossavainen

- ☒ Kielitestien eettisyydestä ja oikeudenmukaisuudesta 27

Minna Martikainen

- ☒ Inkeriläisten paluumuuttajien kielitutkinnon synty keskustelussa 44

Sari Ahola, Reeta Neittaanmäki & Tuija Hirvelä

- ☒ Puheen ymmärtämisen tehtävien taitotasolle asettaminen Yleisissä kielitutkinnoissa 68

Sari Ahola

- ☒ Puhetta arvioinnista – Yleisten kielitutkintojen arvioijien käsityksiä arvioinnista 89

Marja Seilonen, Minna Suni, Marita Härmälä & Reeta Neittaanmäki

- ☒ Ammatillisen kielitaidon arviointikokeilu terveydenhuollon alalla 110

Marita Härmälä, Raili Hildén & Dmitri Leontjev

- ☒ Yhdeksäsluokkalaisten opiskelukäytänteet englannin, ruotsin ja ranskan oppitunneilla ja vapaa-ajalla 142

Heidi Harju-Luukkainen, Mirja Tarnanen & Kari Nissinen

- ☒ Monikieliset oppilaat koulussa: eri kieliryhmien sisäinen ja ulkoinen motivaatio sekä sen yhteys matematiikan osaamiseen PISA 2012 -arvioinnissa 167

Pirjo Pollari

- ☒ Daunting, reliable, important or "trivial nitpicking?" Upper secondary students' expectations and experiences of the English test in the Matriculation Examination 184

Sofia Stolt

- ☒ Kollegial bedömning i summativa bedömarkommentarer på studentexamensuppsatser i modersmål och litteratur 212

ESIPUHE

Tämä julkaisu sai alkunsa vuosina 2014 ja 2015 AFinLAN syysseminareissa järjestetyistä AFinLAN kielitaidon arvioinnin teemaverkoston kollokvioista, joiden esiintyjät raportoivat erilaisista meneillään olevista ja päättyneistä tutkimushankkeista Suomessa. Julkaisu on siis kyseisen teemaverkoston tähänastisen toiminnan merkittävin tuotos. Kun tarkastelimme noiden kahden kollokvion esitelmien aiheita, ne näyttivät kaikki liittyvän neljään arvioinnille hyvin tavalliseen tarkoitukseen: ns. portinvartiointiin, summatiiviseen päättöarviointiin, opetussuunnitelmien tehokkuuden arviointiin ja arvioinnin käyttöön tutkimusvälineenä. Päädyimme rajaamaan julkaisun näihin teemoihin, mutta avaamaan sen muillekin tämäntyyppisiä arviointeja tutkiville kuin niille, jotka olivat osallistuneet AFinLAN kollokvioihin. Nyt käsillä oleva julkaisu sisältääkin sekä kollokvion esiintyjien että muiden arviointitutkimusta Suomessa tekevien artikkeleita. Useimmat artikkelit käsittelevät toisen tai vieraan kielen arviointia, mutta joukossa on myös yksi äidinkielen arviointia koskeva tutkimus ja yksi käsitteellinen, kielitaidon arvioinnin yleiseen etiikkaan ja oikeudenmukaisuuteen liittyvä artikkeli.

Julkaisun tarkoitus ei siis ole antaa kattavaa kuvaa kaiken tyyppisistä arvioinneista, vaan keskittyä sellaisiin, joilla on huomattava vaikutus niihin osallistujien elämään tai suomalaiseen yhteiskuntaan laajemmin. Kattavuutta rajaa luonnollisesti myös se, millaisiin arviointeihin julkaisuun tarjotut ja hyväksytyt artikkelit liittyvät: kaikista Suomessa käytössä olevista tutkinnoista tai tärkeistä arvioinneista ei ole joko tehty tutkimusta tai niistä ei ainakaan tähän julkaisuun ollut tarjolla raportteja julkaistavaksi. Käsittelemme tätä aihetta hiukan tarkemmin julkaisun ensimmäisessä, toimittajien kirjoittamassa artikkelissa, vaikka sen pääpaino on muutaman suurimman tutkinto- tai arviointijärjestelmän kehityskulkujen kuvaamisessa ja joidenkin keskeisten arviointeihin liittyvien teemojen käsittelyssä.

Kiitämme lämpimästi kaikkia teemanumeron kirjoittajia, vertaisarvioijia ja AFinLAN julkaisutoimikunnan jäseniä. Ilman heidän panostaan tämä julkaisu ei olisi ollut mahdollinen.

Kiitämme Tieteellisten seurain valtuuskuntaa ja Koneen Säätiötä taloudellisesta tuesta teemanumeron julkaisemisessa ja taittaja Sinikka Lampista Jyväskylän yliopiston Soveltavan kielentutkimuksen keskukselta hyvästä yhteistyöstä.

Jyväskylässä ja Helsingissä 30.12.2016

Ari Huhta ja Raili Hildén

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 3–26.

Ari Huhta¹ & Raili Hildén²

¹Jyväskylän yliopisto, ²Helsingin yliopisto

Kielitutkinnot ja muu laajamittainen kielitaidon arviointi Suomessa

This paper describes the development, current state and expected future development of the assessment systems that most of the papers in the present volume cover. Our focus is on the systems serving general education, particularly the Matriculation Examination and national evaluations of educational achievement in languages, as well as the National Certificates of Proficiency aimed at adult language learners. We describe how changes in our understanding of what language proficiency is have affected these assessment systems and discuss the impact that some of them have been expected to have and what is known about that impact. We also discuss how the Common European Framework of Reference has been utilised in language assessment in Finland and how the introduction of technology is likely to affect language assessments.

Keywords: language testing, Finland, high-stakes assessment, impact, validity

Asiasanat: kielitaidon arviointi, suomi, kielitutkinnot, vaikuttavuus, validius

1 Johdanto

Tämä ensimmäinen artikkeli kuvaa koko julkaisun artikkeleissa käsiteltyjen suomalaisten kielitaidon arviointijärjestelmien ja tutkintojen kehittymistä, nykytilaa ja tulevaisuuden näkymiä. Päähuomio on yleissivistävän koulutuksen yhteydessä tapahtuvassa arvioinnissa, erityisesti ylioppilastutkinnossa ja oppimistulosten kansallisissa arvioinneissa, mutta käsittelemme myös yleisiä kielitutkintoja, jotka ovat merkittävien aikuisväestölle suunnattu kielitutkinto. Keskitymme kuvauksessamme muun muassa siihen, miten muuttuneet käsitykset kielitaidosta ovat näkyneet arvioinnissa ja millaisia vaikutuksia arvioinneilla on ajateltu olevan. Samalla kuvaus havainnollistaa sitä, miten kielitaidon arviointi on vähitellen ammattimaistunut maassamme huolimatta siitä, että se on edelleenkin varsin pieni erikoisala. Suomen koulutusjärjestelmä ei – toisin kuin monien maiden koulutusjärjestelmät – nojaa oppilaitosten ulkopuolella tuotettuihin testeihin ja tutkintoihin, joten vain harvat erikoistuvat testien ja muun tyyppisten arviointien suunnitteluun ja tutkimukseen. Lopuksi nostamme esille joitakin laajempia teemoja ja kehityskulkuja, jotka näkyvät vaikuttavuudeltaan tärkeissä arvioinneissa, kuten muuttuneet käsitykset kielitaidosta ja eurooppalaisen viitekehyksen hyödyntäminen. Lisäksi pohdimme, millaisia muutoksia tietotekniikka tuo arviointiin ja arvioitaviin taitoihin. Päätämme artikkelin analysoimalla, millaisia vaikutuksia kielitutkinnoilla on pyritty saamaan aikaan ja mitä niiden vaikutuksista itse asiassa tiedetään; käytämme tässä esimerkkinä yleisiä kielitutkintoja.

Käytämme artikkelissamme kielitaidon arviointiin viittaavia termejä seuraavalla tavalla. **Arviointia** käytämme kattavimpana terminä, jossa ei oteta kantaa siihen, miten ja miksi arviointia tehdään. **Testi** ja **koe** tarkoittavat tässä samaa asiaa: arviointia, joka tapahtuu kaikille osallistujille samanlaisilla välineillä ja tarkasti määritellyissä olosuhteissa. Teksti, josta on monivalintakysymyksiä, tai suullinen haastattelu ovat esimerkkejä testistä tai kokeesta. Joihinkin yhteyksiin on vakiintunut jommankumman nimityksen käyttö, joten noudatamme sitä artikkelissa: esimerkiksi ylioppilastutkinnon osiin viitataan yleensä kokeina, ei testeinä, joten käytämme tavallisesti termiä kielikoe, kun puhumme kyseisen tutkinnon osista, mutta muissa yhteyksissä yleensä testiä. **Tutkinto** puolestaan viittaa suurempaan kokonaisuuteen, joka koostuu useammasta osasta (testeistä tai kokeista). Termiin tutkinto liittyy yleensä ajatus tietystä arvioinnin tarkoituksesta, erityisesti portinvarioinnista tai summatiivisesta arvioinnista, joita käsitellään tarkemmin seuraavassa luvussa. Testi ja koe ovat puolestaan siinä mielessä neutraaleja termejä, ettei niihin suoraan liity mitään tiettyä arvioinnin käyttötarkoitusta.

2 Kielitaidon arvioinnin tarkoitukset

Käsitlemme aluksi kielitaidon arvioinnin tarkoituksia ja sitä, miten julkaisun artikkelit liittyvät niihin. Arvioinnin päämääriä ei ole helppo määritellä ja erotella toisistaan yksiselitteisesti. Koska arvioinnin tarkoitus on kuitenkin keskeisin tekijä, josta on syytä olla selvillä, kun arvioinnin laatua, vaikutuksia tai muita ominaisuuksia käsitellään, on välttämätöntä nojautua johonkin tapaan eritellä näitä tarkoituksia. Pohjaksi omaan analyysiimme olemme valinneet Brownin (2013) melko uuden, soveltavan kielitieteen kansainvälisessä hakuteoksessa julkaistun jaottelun. Yksi Brownin (2013) mainitsema keskeinen arvioinnin tarkoitus on portinvartiointi (e. *gatekeeping*), jossa selvitetään kielitaitoa jonkin pätevyuden tai tutkintotodistuksen myöntämiseksi (ks. myös Huhta & Takala 1999). Siinä halutaan varmistua siitä, että osallistuja on saavuttanut vaikkapa jossakin toimessa tai opiskelussa vaadittavan kielitaidon. Arviointi toimii siis porttina, joka mahdollistaa tai estää pääsyn tällaiseen ammattiin tai koulutukseen. Esimerkkejä suomalaisista kielitaidon arviointijärjestelmistä, joilla on tällainen funktio, ovat yleiset kielitutkinnot, joita Aholan ja Aholan, Neittaanmäen ja Hirvelän artikkelit käsittelevät sekä inkeriläisten paluumuuttajien kielitutkinto, joka on Martikaisen artikkelin aiheena. Myös Seilosen, Sunin, Härmälän ja Neittaanmäen artikkeli käsittelee osittain yleisiä kielitutkintoja, mutta se liittyy kiinteästi portinvartija-arviointiin terveydenhuollossa eli yhdellä työelämän alueella. Tossavaisen artikkeli kielitestiä eettisyydestä on puolestaan tematiikaltaan yleisempi, mutta keskittyy varsin pitkälti tutkintotyyppiseen portinvartiointiin.

Portinvartiointi on siis julkaisumme kannalta erittäin keskeinen arvioinnin päämäärä, ja artikkelit käsittelevät useita tämän päämäärän kannalta keskeisiä teemoja: arvioinnin eettisyyttä ja oikeudenmukaisuutta (Tossavainen), arvioitavan kielitaidon luonnetta (Seilonen ym.), puheen ja kirjoituksen arvioijien toimintaa (Ahola), ymmärtämisen arviointitehtävien vaikeustasojen asettamista (Ahola ym.) ja arvioinnin vaikutuksia (Martikainen).

Toinen julkaisun kannalta keskeinen arvioinnin tarkoitus on summatiivinen, tyypillisesti jonkin opetuskokonaisuuden lopussa tehtävä arviointi, jossa selvitetään, miten hyvin oppijat ovat saavuttaneet opetukselle ja oppimiselle asetetut tavoitteet (Brown 2013), ja jonka perusteella oppijoille yleensä annetaan arvosana. Opetuskokonaisuuksien laajuus voi vaihdella yksittäisistä kursseista kokonaisuun oppimääriin, kuten lukion päättökokeessa. Tyypillistä suomalaiselle arviointikulttuurille on ollut luottamus kunkin koulutusasteen opettajien antamiin päättöarvioihin, joten meiltä puuttuvat muutamia erikoisluokkia lukuun ottamatta esimerkiksi muualla maailmassa tavalliset valikoivat kokeet siirryttäessä alakoulusta yläkouluun. Myöskään tavanomaiseen lukiokoulutukseen ei ole ainekohtaisia valintakokeita, vaan ensimmäinen koko lukiokäluokkaa koskeva tärkeä tutkinto on lukiokoulutuksen päätteeksi suoritettava ylioppilas-

tutkinto. Se toimii lukion oppimäärän summatiivisena, valtakunnallisesti yhtenäisenä loppukokeena. Tässä julkaisussa ylioppilastutkintoa käsittelevät Pollarin ja Stoltin artikkelit. Pollarin tutkimuksessa tarkastellaan ylioppilastutkinnon englannin kielen kokeeseen valmistautuvien abiturienttien odotuksia ja käsityksiä tutkinnosta. Stoltin artikkeli puolestaan käsittelee abiturienttien suorituksia arvioivien sensoreiden kommentteja; aineistona hänellä oli ruotsinkielisten oppilaiden kirjoitelmia äidinkielen ja kirjallisuuden ylioppilaskokeessa.

Kolmas arvioinnin tarkoitus, jota julkaisun artikkelit käsittelevät, liittyy sellaiseen arviointiin, jossa huomio ei ole yksilöiden osaamisessa, vaan jossa ollaan kiinnostuneita suurten oppijajoukkojen keskimääräisestä osaamisesta ja siinä esiintyvistä vaihtelusta (Brown 2013; Huhta & Hildén 2013). Tällaisessa oppimistulosten arvioinnissa otetaan edustava otos tutkittavista oppilaitoksista ja oppilaista. Kaikkia oppilaita ei testata, koska huomio on yleistuloksissa ja korkeintaan halutaan selvittää, eroavatko tulokset maan eri osissa ja eri sukupuolten välillä tai paljonko eri koulujen tulokset vaihtelevat. Härmälän, Hildénin ja Leontjevin artikkeli perustuu Suomessa Opetushallituksen ja Kansallisen koulutuksen arviointikeskuksen Karvin tekemään 9.-luokkalaisten valtakunnalliseen kieliaineiden arviointitutkimukseen. Tutkimuksessa tarkastellaan englannin, ruotsin ja ranskan opiskelijoiden opiskelukäytänteitä, vapaa-ajan kielenkäyttöä ja kielitaitoa sekä tutkitaan niiden yhteyksiä. Harju-Luukkaisen, Nissisen ja Tarnasen artikkeli käsittelee puolestaan vuonna 2012 toteutettua kansainvälistä OECD:n PISA-tutkimusta. Heidän artikkelinsa vertailee maahanmuuttajataustaisten ja ensikieleltään suomenkielisten oppilaiden motivaatiota, oppimistuloksia ja niiden yhteyksiä.

Osassa tämän julkaisun artikkeleista on mukana myös arviointien käyttö tutkimusaineiston hankkimisen keinona, joka on neljäs Brownin (2013) kuvaama arvioinnin tarkoitus. Selvimmin tämä näkyy kummassakin julkaisun oppimistuloksia käsittelevässä artikkelissa (Hildén ym.; Harju-Luukkainen ym.). Jo tavoite selvittää oppimistuloksia kansallisesti tai kansainvälisesti on itsessään tutkimuksellinen. Jos tulosten raportointia syvennetään esimerkiksi tarkastelemalla, mitkä tekijät ovat yhteydessä saatuihin tuloksiin ja ehkä selittävätkin niitä, tutkimusote lähenee kielenoppimisen tutkimusta (e. *second language acquisition*). Myös Seilosen ym. tutkimuksessa (tässä julkaisussa) terveydenhuollon ammattilaisten kielitaidosta yhdistyvät arviointien käyttäminen tutkimusaineiston hankinnan välineenä ja tutkimus kielitaidon arvioinnista portinvartijana.

Monilla arvioinneilla ja testeillä on useampia kuin yksi päämäärä. Hyvä esimerkki tästä on ylioppilastutkinto. Vaikka sen päätarkoitus on toimia summatiivisena päätösarviointina, sillä on myös portinvartijafunktio silloin, kun sen tulokset huomioidaan korkeakoulujen opiskelijavalinnoissa. Hyvin yleistä on käyttää summatiivisia arviointeja oppilaitoksissa myös formatiivisiin, oppimista ja opetusta tukeviin tarkoituksiin. Osa yrityksistä käyttää arviointia eri tarkoituksiin on toimivia ja perusteltuja, mutta saman arvioinnin hyödyntäminen erilaisiin tarkoituksiin ei ole automaattisesti validia ja pe-

rusteltua, vaan aina, kun arviointia halutaan käyttää johonkin muuhun kuin mihin se on alun perin suunniteltu, tämä täytyy perustella (ks. esim. Fulcher & Davidson 2009). Mitä tärkeämmästä arvioinnista on kyse, sitä vakavampaa huomiota on syytä kiinnittää siihen, onko perusteita tehdä arvioinnin tuloksista muita kuin alkuperäiseen tarkoitukseen liittyviä johtopäätöksiä. Hyvä esimerkki ongelmista tässä asiassa on yritys käyttää ylioppilastutkintoa keinona arvioida opetuksen ja koulujen laatua, mihin lukioden ylioppilastutkintotuloksiin perustuvilla ranking-listoilla pyritään. Kyseinen yritys selvittää koulujen laatua yksilöiden tarpeisiin tarkoitetun summatiivisen tutkinnon perusteella edustaa epäonnistunutta tapaa yhdistää arvioinnin erilaisia tarkoituksia samaan testiin tai tutkintoon (ks. Kuusela 2003: 17).

Karkeasti ottaen kaikki arvioinnit voidaan jaotella sen mukaan, rakentuvatko ne jollekin oppimäärälle (esim. kurssi- tai opetussuunnitelmalle), yleisemmälle teorialle kielitaidosta vai näkemykselle jossakin toimessa tai ammatissa vaadittavasta kielitaidosta (perustuen tyypillisesti tarveanalyysiin). Oppilaitoksissa tapahtuva arviointi, josta summatiivinen päättöarviointi on yksi esimerkki, perustuu luonnollisesti kunkin kurssin, vuosiluokan tai kokonaisen oppimäärän opetussuunnitelmaan. Myös kansalliset oppimistulosten arvioinnit 9. vuosiluokan loppuvaiheessa nojaavat perusopetuksen opetussuunnitelmaan kyseisissä oppiaineissa. Portinvartija-arviointi perustuu puolestaan joko tarvekartoituksiin, kieliteorioihin tai niiden yhdistelmiin riippuen siitä, milaista porttia vartioidaan – johonkin ammattiin johtavaa vai vaikkapa vieraalla kielellä tapahtuviin korkeakouluopintoihin vievää. Myös kansainväliset oppimistutkimukset perustuvat teorioihin mitattavista osaamisalueista, koska ne eivät tietenkään voi perustua minkään yksittäisen maan opetussuunnitelmiin, jotteivat tulokset vääristyisi. Kuten dikotomiat ja jaottelut yleensä, ei arvioinnin teoreettisten ja käytännöllisten tavoitteiden välinen rajakaan ole yksiselitteinen. Esimerkiksi opetussuunnitelmien taustalta löytyy todennäköisesti kielitaitoon ja sen oppimiseen ja käyttöön liittyvää tutkimusta ja teoreettisia näkemyksiä.

Tähän julkaisuun kootut tutkimukset kielitaidon arvioinnista edustavat siis erilaisia arvioinnin tarkoituksia. Lisäksi ne täyttävät ainakin osan seuraavista tunnusmerkeistä: 1) niihin osallistuu huomattava osuus ikäluokasta tai tiettyyn kohderyhmään kuuluvista, 2) kokeesta suoriutumisen on suuri tai ratkaiseva merkitys suorittajan tulevaisuudelle, ja 3) toimeenpanossa noudatetaan säädöksiin ohjattua standardisoitua menettelyä, jolla pyritään yhdenmukaisiin suoritusoloihin. Mainitut tunnusmerkit painottuvat jonkin verran vaihdellen artikkeleissa, sillä kaikki kolme täyttyvät todella harvoin viime vuosikymmenien suomalaisessa koulutuksessa. Parhaiten ne täyttyvät ylioppilastutkinnon äidinkielen ja englannin kokeissa, joihin osallistuu noin puolet ikäluokasta, sekä inkerinsuomalaisten paluumuuttajien kielitutkinnossa, johon osallistui sen käytön aikana valtaosa Suomeen muuttaneista inkerinsuomalaisista (ks. Martikainen tässä julkaisussa). Tosin heidän lukumääränsä ei noussut lähellekään ylioppilastut-

kinnon suurien kieliaineiden osallistujamääriä. Myös yleisten kielitutkintojen suomen ja ruotsin kielen tutkinnot täyttävät varsin hyvin edellä mainitut kolme tunnusmerkkiä. Niihin osallistuvat käytännössä kaikki suomen kansalaisuutta hakevat aikuiset maahanmuuttajat, ja niiden tuloksia käytetään myös työ- ja opiskelupaikan haussa. Hyvin suuri osa maahanmuuttajista, joilla on jonkin verran alkeiskielitaitoa parempi suomen tai ruotsin taito, osallistunee yleisiin kielitutkintoihin.

Käsillä oleva julkaisu ei suinkaan kata kaikenlaista arviointia eikä kaikkia arvioinnin tarkoituksia, ei edes kaikkia suomalaisia kielitutkintoja. Esimerkiksi valtionhallinnon kielitutkinnot ja auktorisoitujen kääntäjien tutkinnot jäävät tässä käsittelemättä. Samoin julkaisun ulkopuolelle jää maahanmuuttajien kotoutumiskoulutukseen liittyvä arviointi. Kielitaitoa arvioidaan myös monissa koulutusohjelmissa, joissa tietyn tason kielitaidon osoittaminen on edellytyksenä akateemisen oppiarvon (kandidaatin- tai maisterintutkinnon) tai ammatillisen tutkinnon saavuttamiseen, mutta näiden arviointien toimeenpano on usein muodoiltaan vaihtelevaa ja korkeintaan paikallisesti sovittuja käytäntöjä noudattelevaa. Nämä arvioinnit jäävät monimuotoisuudessaan artikkelikokoelman ulkopuolelle, kuten muukin oppimista ja opettamista tukeva ns. formatiivinen arviointi ja sen tutkimus. Erityisesti formatiivinen arviointi on kuitenkin niin yleinen ja tärkeä kielitaidon arvioinnin tarkoitus, että sitä koskeva Suomessa tehtävä tutkimus ansaitsee oman julkaisunsa tulevaisuudessa.

3 Merkittävimmät kielitutkinnot ja arviointitutkimukset Suomessa

3.1 Yleissivistävän koulutuksen kielikokeet

Yleissivistävän koulutuksen laajojen kielikokeiden ja -tutkintojen tarkoitus on koulutuksen tavoitteiden saavuttamisen todentaminen; näitä ovat perusopetuksen oppimistulosten arvioinnit ja ylioppilastutkinto. Tämän lisäksi ylioppilastutkinnolla on portinvartijafunktio jatko-opintokelpoisuuden osoittajana.

Yleissivistävän koulutuksen ja etenkin lukiokoulutuksen päätteeksi suoritettavan ylioppilastutkinnon kielikokeiden kielitaitokäsitys on mukaillut kunkin ajanjakson viestinnällisiä ja pedagogisia painotuksia ja arvoja. Kirjallisen kieliperinnön ihannointi ja siihen nivoutunut muotopainotteinen pedagogiikka leimasivat kielikokeita aina 1960-luvulle saakka. Kielitaitokäsitystä hallitsivat yhden kulloinkin mitattavan kielen kieliopilliset taidot ja korkeintaan niiden soveltaminen kirjoitettuun tekstiin, jonka sisältö oli valmiiksi annettu. Sitä, mihin noita tekstejä tosielämässä tarvittaisiin, ei pohdittu,

eikä kokelaan edellytetty eikä sallittu ilmaista testattavalla kielellä omia ajatuksiaan. (ks. Kaarninen & Kaarninen 2002: 293–296.)

Luonnollisen kielenkäytön simulointi kielikokeissa käynnistyi samoihin aikoihin luonnollisen kielen käyttötarkoituksiin paneutuneen tutkimussuuntauksen kanssa. Syntyi käsitys viestinnällisestä kielitaidosta (Hymes 1972), jossa suullinen ja kirjallinen kieli ja sen ymmärtäminen ja tuottaminen tapahtuvat rinnakkain, peräkkäin tai toisiinsa sulautuen tilanteen vaatimalla tavalla. Tämä näkemys tuotiin esiin jo peruskoulun opetussuunnitelmassa (POPS 1970), ja sitä myötäilivät edelleen perusopetuksen ja lukion opetussuunnitelmien perusteet (POPS 1994 ja 2004; LOPS 1985, 1994 ja 2003). Kieliai- neet katsottiin näissä perusteissa tietoa, taitoa ja kulttuurista osallistumista edistäviksi. Uusimmissa yleissivistävän perusopetuksen (POPS 2014) ja lukiokoulutuksen (LOPS 2015) opetussuunnitelman perusteissa kielitaitokäsitys on avartunut entisestään. Perusopetuksessa kielen katsotaan nivoutuvan kaikkeen ajatteluun, oppimiseen ja vuoro- vaikutukseen, ja sen ilmenemismuodot kattavat paitsi verbaalin kielen myös modernin, monimediaisen toiminta- ja ilmaisuympäristön. Laaja-alaisen osaamisen komponentit kuvastavat tätä käsitystä kielestä ajattelun, kulttuurisen osaamisen ja vuorovaikutuksen erottamattomana osana. Monilukutaito ja tieto- ja viestintäteknologinen osaaminen tukevat ja laajentavat perinteisempää käsitystä kielestä viestintävälineenä. (POPS 2014: 20–24.) Kieliai- neiden tavoitteet jäsenyivät seuraaviin osa-alueisiin: kasvu kulttuuriseen moninaisuuteen ja kielitietoisuuteen, kielenopiskelutaidot, kielitaito (vuorovaikutuk- sessa ja taitona tulkita ja tuottaa tekstejä). Lisäksi määritellään näihin tavoitealueisiin liittyvät sisällöt ja laaja-alaisen osaamisen teemat. Hyvää osaamista ilmentävän arvosa- nan 8 tarkoittama taitotaso ilmoitetaan ohjeistamaan perusopetuksen päättöarvosa- nan antoa. Kuvausasteikko on sovellus edellisellä opetussuunnitelmakerroksella käy- tetystä eurooppalaiseen viitekehukseen (EVK 2003) pohjautuneesta asteikosta (POPS 2004; Hildén & Takala 2007) ja muista viitekehysasteikoista.

Perusopetukselle rakentuvan lukiokoulutuksen opetussuunnitelman perusteis- sa vallitsee jonkin verran perinteisempi kielitaitokäsitys. Siellä kielitaito liittyy kiinteim- min aihekokonaisuuksiin kulttuurien tuntemus ja kansainvälisyys sekä monilukutaito ja mediat. Kielitaitokäsityksen laajentunutta tulkintaa edustaa etenkin monilukutaidon kuvaus (lihavoinnit kirjoittajien lisäämät):

”Monilukutaidolla tarkoitetaan taitoja tulkita, tuottaa ja arvottaa tekstejä eri muodoissa ja konteksteissa. **Medialukutaito on osa monilukutaitoa.** Monilukutaito perustuu laaja- alaiseen tekstikäsitelyyn, jonka mukaan tekstit ovat **sanallisten, kuvallisten, auditiivisten, numeeristen tai kinesteettisten symbolijärjestelmien tai niiden yhdistelmien muodostamia kokonaisuuksia.** Monilukutaito tukee ajattelun ja oppimisen taitojen ke- hittymistä sekä syventää kriittistä lukutaitoa ja kielitietoisuutta. Erilaiset lukutaidot kehit- tyvät kaikessa opetuksessa ja kaikissa oppiaineissa.” (LOPS 2015: 30.)

Kielten ainekohtaisissa tavoitekuvauksissa kieltenopetuksen todetaan edistävän osaltaan monilukutaitoa. Kaikkien kielten yhteiset tavoitteet painottuvat rohkaisuun käyttää kieltä monin tavoin eri yhteyksissä ja ohjata oman kielitaitonsa kehittymistä sopivin strategioin ja itsearvioinnin avulla. Kielitaidon kuvausasteikon avulla on asetettu tavoitteet kielitaidolle eri oppimäärissä opiskelun päättyessä, ja ne ohjaavat myös kielitaidon arviointia. Käytössä on sama asteikko kuin perusopetuksessa, ja se jakautuu viiteen alueeseen: vuorovaikutus erilaisissa tilanteissa, viestintästrategioiden käyttö, viestinnän kulttuurinen sopivuus sekä tekstien tulkintataidot ja tuottamistaidot. Etenkin uusimpia opetussuunnitelmaperusteita hallitsevat formatiivinen arviointi ja arvioinnin oppimista tukevat tarkoitukset. (LOPS 2015: 87–128.)

Standardoitu summatiivinen arviointi tavoittaa kielen oppimisprossin ja siihen liittyvät taidot epäsuorasti niiltä osin kuin ne näkyvät suorituksessa. Varmaankin tästä syystä kielitaidon arviointi analyttisesti mitattavina neljänä osataitona (Lado 1961) on pitänyt summatiivisessa arvioinnissa pintansa meidän päiviimme saakka kaikkialla maailmassa. Suomessakin sekä oppimistulosten arvioinnit että ylioppilastutkinnon kielikokeet ovat rakentuneet pääasiassa osataidoittain.

Sekä kansallisten arviointien että ylioppilastutkinnon tehtävät pyritään laatimaan mahdollisimman hyvin kulloinkin voimassa olevaa opetussuunnitelmaa vastaaviksi. Ylioppilaskoe perustuu ajatukseen, että koetehtävät ovat otos mahdollisista kielenkäyttötehtävistä, joita koulussa on harjoiteltu ja joita kokelas voi kohdata. Osoittehtävistä suoriutumisen nojalla tehdään johtopäätöksiä kokelaan suoriutumisesta samankaltaisista tehtävistä koetilanteen ulkopuolella, eli kokeen tulos yleistetään sen ulkopuolelle (Bachman & Palmer 2010). Koetehtävien laadinnassa huolehditaan siitä, että tehtävissä on riittävästi olennaisia piirteitä yleistysavaruudesta eli niistä tehtävistä ja tilanteista, joihin tulos on tarkoitus yleistää. On kuitenkin selvää, että kertaluonteisen kokeen kyky mitata kokelaan arkielämässä pitkällä aikavälillä toteutuvaa strategista viestintätaitoa on rajallinen. Silti ylioppilastutkinnon kielikokeiden viestinnällisyyttä on tarvetta lisätä, ja sähköistyminen tarjoaa siihen aiempaa tuntuvasti paremmat välineet.

3.1.1 Ylioppilastutkinto

Alkuaan Kuninkaallisen Turun Akatemian ja sittemmin Keisarillisen Aleksanterin yliopiston pääsykuulusteluista kehittynyt ylioppilastutkinto kytkettiin lukion opetussuunnitelmaan vuonna 1852, minkä jälkeen opetussuunnitelmien muutokset ovat heijastuneet myös kielikokeisiin, tosin useimmiten huomattavalla viipeellä. (Ylioppilastutkintolautakunta 2016a; Kaarninen & Kaarninen 2002.)

Aluksi kokonaan suullisten ylioppilaskokeiden rinnalle säädetyt kirjalliset kokeet siirrettiin lukioihin 1874. Vuoden 1919 asetuksella pakollisiksi ylioppilaskoeaineiksi säädettiin viisi ainetta, joihin lukeutuivat toinen kotimainen ja yksi vieras kieli. Vieraan kielen kokeeseen varatun ajan kuluessa oli mahdollista suorittaa useammankin kielen koe.

Vieraan kielen koe oli mukana myös 1940-luvulla järjestetyissä ns. sotilasylioppilaskirjoituksissa yhtenä kolmesta pakollisesta aineesta. (Ylioppilastutkintolautakunta 2016a.) Vuonna 1947 pakollisten aineiden määrä vakiintui neljään, ja samalla sallittiin ylimääräisten kokeiden suorittaminen. Näihin lukeutuivat myös kielten lyhyet oppimäärät. Tutkinnon rakenneuudistus vahvistettiin vuonna 2005, jolloin toinen kotimainen kieli muuttui valinnaiseksi koeaineeksi.

Kielikokeiden tehtävätyypit ja suoritusmuoto pysyivät suurin piirtein ennallaan lähes sata ensimmäistä tutkintovuotta. Kieliopillisesti painottunutta klassisten kielten rakenneopetuksesta ammentavaa pedagogiikkaa myötäilevä kirjallinen käännöskoe säilytti asemansa aina 1960-luvulle saakka. Tällöin käynnistettiin kokeilu useampia osaitaitoja mittaavan kokeen kehittämiseksi. Vuonna 1974 ylioppilastutkinnon kielikokeet muuttuivat uuden asetuksen myötä kolmiosaisiksi siten, että niihin kuului kuullun ja luetun ymmärtämistä testaava osa sekä lyhyt kirjoitelma. (Ylioppilastutkintolautakunta 2016a; Kaarninen & Kaarninen 2002.)

Pääsääntöisesti kolme kielellistä osataittoa mittaava koe on säilynyt pitkässä, keskipitkässä ja lyhyessä oppimäärässä pienin muunnoksin meidän päiviimme saakka. Vaihteleva määrä sanaston ja rakenteiden hallintaa mittaavia tehtäviä on kuulunut eri kielten ja oppimäärien kokeisiin. Jo vuosikymmenien kiistakapula on puhumisen kokeen puuttuminen ylioppilastutkinnosta, vaikka jo Nykykielten opetussuunnitelmassa (1970) korostettiin kielen viestinnällisyyttä ja 1980-luvulla suullisen kielitaidon harjoitteluun annettiin hyvinkin modernia opastusta (Kurssimuotoisen lukion oppimäärät 1981: 44–45; LOPS 1985). Sittemmin opetussuunnitelma toisensa perään on tähdentänyt puhumisen tärkeyttä (LOPS 1985, 1994, 2004). Huolimatta merkittävistä kokeiluista ja ponnistuksista puhumisen arvioinnin sisällyttämiseksi ylioppilaskokeisiin käytännön toteuttamisen ongelmat ovat tähän saakka estäneet muutoksen. (Ks. esim. Lukiokoulutuksen suullisen kielitaidon arviointityöryhmän muistio 2006).

Opetussuunnitelmakauden 2003–2015 aikana ylioppilastutkinnon kielikokeiden tuloksia tutkittiin aiempaa enemmän tarkoituksena saada ne vastaamaan eurooppalaisen viitekehyksen taitotasoa, joita soveltaen eri kielten ja oppimäärien tavoitteet (ns. hyvän osaamisen taso) oli ilmaistu lukion opetussuunnitelmien vuonna 2003 annetuissa perusteissa (Hildén & Takala 2007). Tällaisella verrantamisella (e. *standard setting*) tarkoitetaan kielikokeissa saavutettujen tulosten ilmoittamista esimerkiksi EVK:ssa tai jossain muussa viitekehityksessä määriteltynä taitotasoina (ks. tarkemmin Ahola ym. tässä julkaisussa). Näin toteutettiin empiirisesti ylioppilastutkinnon lakisääteistä tehtävää todeta lukio-opiskelun aikana saavutettujen tietojen ja taitojen taso ja se, missä määrin kokelaiden suoriutuminen vastaa opetussuunnitelman perusteissa asetettuja tavoitteita. Samalla analysoitiin tutkintotehtävien ja opetussuunnitelman keskinäistä tasovastaavuutta ja kehitettiin tutkinnon kielikokeita.

Kevään 2012 kielikokeita koskevat verrannustulokset vaihtelivat kielittäin ja oppimäärittäin suhteessa opetussuunnitelman perusteissa määriteltyyn hyvän osaamisen tasoon. Jonkin verran eroja havaittiin myös saman kokeen eri osataitojen opetussuunnitelmaavastavuudessa. Pitkän englannin koe arvioitiin kokonaisuutena tavoitetasoa helpommaksi, ja sen vaativin osakoe mittasi sanastoa ja rakenteita. Helpoimmaksi osoittautui luetun ymmärtämisen osakoe. (Juurakko-Paavola & Takala 2013: 20.)

Toisen kotimaisen kielen ruotsin ja suomen keskipitkän oppimäärän kokeet vuonna 2012 osoittautuivat verrannusanalyysin perusteella myös tavoitetasoaan helpommiksi. Ranskan lyhyen oppimäärän tehtävät vaikuttivat suunnilleen tavoitetasonsa mukaisilta, kun taas saksan lyhyen oppimäärän tehtävät olivat jonkin verran vaativampia kuin niiden tavoitetasot. Kokelaiden suoriutumisesta havaittiin, että enemmistö saavutti tai ylitti hyvän osaamisen tason B2.1 englannin pitkässä oppimäärässä. Toisen kotimaisen kielen ruotsin ja suomen keskipitkissä oppimäärässä tuottamistaitojen (kirjoittamisen) tavoitetaso oli B1.1. ja ymmärtämistaitojen (kuullun ja luetun ymmärtäminen) B1.2. Kokelaiden enemmistön keskimääräinen suoritustaso jäi tasolle A2.2 eli yhden tasoaskelman alempaakin asetettua tavoitetasoa heikommaksi. Ranskan ja saksan lyhyiden oppimäärien suorittajista suurin osa ylsi tavoitetasoon tai ylitti sen. Tosin näiden oppimäärien kokeisiin osallistui myös pitkän tai yläkoulussa aloitetun B2-oppimäärän opiskelleita, joten kuva nimenomaan lyhyen oppimäärän tavoitteiden saavuttamisesta jää osin tulkinnanvaraiseksi. (Juurakko-Paavola & Takala 2013, 24–25.) Kyseiset tulokset huomioitiin 2015 opetussuunnitelman perusteissa siten, että ruotsin ja suomen keskipitkien oppimäärien tavoitteet yhtenäistettiin osataidoittain tasolle B1.1 (LOPS 2015, 88, 98, 108).

Ylioppilastutkinnon latinankieliset arvolausumat perustuvat normaalijakumaan, joka oli käyttökelpoinen silloin, kun tutkinto oli suppea ja kaikki kokelaat todella suorittivat kaikki kokeet. Näin ei ole tapahtunut enää vuosikymmeniin. Ylioppilastutkinnon sisältö vaihtelee eri kokelailla ja vaikeustason vakauttamispyrkimyksistä huolimatta myös saman oppimäärän eri koekerroilla. Kielten taitotasoverrannus sai pontta eurooppalaisen viitekehysten (EVK 2003) käyttöönotosta, mutta se liittyy rohkaisevana esikuvana myös yleisempään tarpeeseen viedä ylioppilastutkinnon tulosten raportointia kriteeriviitteisempään ja informatiivisempaan suuntaan. Latinankieliset arvosanat ovat sellaisinaan läpinäkymättömiä, eivätkä ne kerro mitään siitä, mitä tietoja ja taitoja kokelas on osoittanut kyseisellä koekerralla. Tieto jää puuttumaan niin kokelaalta itseltään kuin vastaanottavilta oppilaitoksiltakin. Ylioppilastutkinnolla on kuitenkin tarve palvella kansantaloudellisesti myös korkeamman asteen opiskelijavalintaa. On hyvä huomata, että myös muiden aineryhmien perinteiset arvolausumat approbaturista laudaturiin edustavat hyvin eritasoista osaamista, joskin piiloisemmin, koska niissä ei ole käytössä eurooppalaisen tasoasteikon kaltaista valmista absoluuttista kuvausasteikkoa. On kuitenkin selkeää näyttöä siitä, että joissakin aineissa korkeampien arvosanojen

saaminen on ollut paljon yleisempää kuin toisissa (Kupiainen, Marjanen & Hautamäki 2016): vaativina aineina ja oppimäärinä voidaan pitää esimerkiksi matematiikan ja fysiikan pitkää oppimäärää sekä saksan ja ranskan pitkää oppimäärää, kun taas terveystiedossa korkeimmat arvosanaluokat ovat keränneet väkeä enemmän kuin samojen kokelaiden yleismenestys tutkinnossa antaisi olettaa.

Epäkohdan korjaamiseksi ylioppilastutkinnon arvosanajakauman määrittelyssä on kevästä 2014 lähtien sovellettu standardoitujen yhteispisteiden keskiarvoon perustuvaa laskentatapaa (ns. SYK-jakauma), joka huomioi yksittäisen kokeen lisäksi myös kokelaan muut suoritukset ja vertaisryhmän suoritukset useammalta koekerralta. (Marjanen 2015; Ylioppilastutkintolautakunta 2016b.) SYK-malliin perustuvaan arvosanojen jakoon siirryttiin asteittain vuoteen 2016 mennessä, ja vaikutukset ovat nähtävissä esimerkiksi laudatur-arvosanojen määrän maltillisena nousuna eräissä aiemmin hyvin vaativina pidetyissä tutkintoaineissa. Loogisesti ajatellen kunkin aineen arvosanajakauman keskikohdan (cum laude approbatur) pitäisi sijaita hyvän osaamisen tasoon oikeuttavan pistemäärän tuntumassa. Vuonna 2012 tutkimusaineistossa odotus toteutui vain pitkässä englannissa. Ainekohtainen osuvuus ei ole ongelmaton myöskään SYK-määrittelyyn perustuvissa jakaumissa etenkään niissä aineissa, joiden kokeisiin osallistuu paljon esimerkiksi testattavaa kieltä äidinkielenään puhuvia kokelaita, joiden menestys muissa aineissa on heikompaa (Marjanen 2015). Jatkotutkimusta ja kehittämistyötä tarvitaan edelleen yhtäältä taitotasoverrannuksen validointiin ja toisaalta ainakin hyväksyttävän suorituksen vähimmäistason määrittelyyn kullekin oppiaineelle relevantilla tavalla. Ylioppilaskokeen kielikokeiden raportointi viitekehysten taitotasoina on pyrkimyksenä niin pian kuin se on luotettavasti toteutettavissa. (Tulossopimus 2016).

Puhumisen kokeen puuttuminen ylioppilastutkinnosta on askarruttanut kieltenopettajia ja arvioijia lähes kolme vuosikymmentä. Puhumisen arviointi kertasuorituksen perusteella koetaan usein muita osataitoja hankalammaksi, koska siihen liitetään muita tiiviimmin prosessuaalisuus ja tilannekohtaisuus. (Luoma 2004; Fulcher 2003). On selvää, ettei ylioppilastutkinnon kaltainen kertatutkinto tavoita minkään muunkaan osa-alueen toteutumista arkielämässä, mutta silti kirjalliset kokeet on hyväksyty edustamaan kunkin kielen taitoa tutkinnon toista vuosisataa kestäneen historian aikana. Kertanäyttö validiusrajoitteinen katsotaan käytännölliseksi tavaksi mitata osaa kaikkien kokelaiden kielitaidosta mahdollisimman samalla tavalla. Standardoidut suoritusolot ja ulkopuolinen arvioijakunta takaavat näin ollen muiden taitojen arvioinnin riittävän objektiivisuuden. Puhumisen kokeen puuttumisen suurimmat syyt ovat perinteisen arvioinnin työläys ja taloudelliset ehdot. Isoissa oppimäärissä tarvittaisiin kymmeniä, ellei satoja sensoreita arvioimaan jo muutaman minuutin kestoisen pariskeskustelu. Uusia mahdollisuuksia avaa kielikokeiden sähköistyminen, joka mahdollistaa sen, että ainakin osa tehtävistä arvioidaan koneellisesti puheentunnistukseen

perustuvan välineen avulla. Tällaisen älykkään tunnistimen työstäminen on meneillään DigiTala-hankkeessa (DigiTala 2016). Puhumisen mittaaminen ylioppilastutkinnon kielikokeissa on asetettu tavoitteeksi vuodelle 2020, jolloin arvioinnin suorittaisivat kone ja ihmisarvioijat yhdessä.

Suurin ylioppilastutkintoa koskeva uudistus vuosikymmeniin on juuri sähköistämisen. Kielikokeista ensimmäisenä sähköistettiin saksan kielten koe syksyllä 2016. Sen jälkeen sähköistyvät ranskan ja toisen kotimaisen kielen kokeet, ja syksyyn 2018 mennessä sähköinen koe järjestetään kaikissa kielissä ja oppimäärissä. Sähköistys mahdollistaa entistä autenttisemmat tehtävätyypit ja myös koneellisen arvioinnin siihen soveltuviin tehtäviin (ks. esim. Chapelle & Voss 2016).

Ylioppilastutkinnon kielikokeiden osallistujamäärät heijastavat yhteiskunnan muuttuvia kielitaitotarpeita. Lukion opetussuunnitelman perusteissa kaikki kielet ja oppimäärät ovat tasavertaisesti esillä, mutta tällä vuosituohannella kielten kirjoittajamäärissä on tapahtunut huomattavia muutoksia muiden aineiden hyväksi. Ainoastaan englannin kieli on säilyttänyt kiistattoman johtoaseman globaalistuvassa taloudessa, mediassa ja nuorisokulttuurissa – ja ylioppilastutkinnossa. Lähes kaikki kokelaat suorittavat jommankumman englannin oppimäärän.

3.1.2 Perusopetuksen kielten oppimistulosten arviointi

Perusopetuksen päättövaiheessa saavutettua kielitaitoa on tarkasteltu jokseenkin säännöllisesti peruskoulun alkutaipaleelta lähtien. Vieraisissa kielissä ensimmäiset oppimistulosten arvioinnit tehtiin ruotsin ja englannin kielten osaamisen selvittämiseksi (Takala 1971; Havola-Pitkänen 1974; Hakala & Kärkkäinen 1976). Sittemmin näiden meillä eniten opiskeltujen kielten tuloksia on tutkittu opetussuunnitelmien muutosten myötä (Kärkkäinen 1983; Havola & Saari 1993; Karppinen & Sarkkinen 1995; Tuokko 2000; Takala 2004). Nämä arvioinnit raportoitiin tyypillisesti ratkaisuosuuksina (kuinka monta prosenttia tehtävistä oppilaat osasivat ratkaista), joten niiden tuloksia on vaikea rinnastaa ja tulkita nykyisen taitotasojärjestelmän mukaan.

2000-luvun ensimmäisellä vuosikymmenellä Opetushallituksen arviointitoiminta keskittyi oppimistuloksiin (Jakku-Sihvonen 2014, 81–83). Vieraiden kielten oppimistuloksia on arvioitu ranskan, saksan ja venäjän pitkissä oppimäärissä (Tuokko 2002; Väisänen 2003) ja ruotsin keskipitkässä oppimäärässä (Tuokko 2009). Vuonna 2009 arvioitiin toisen kotimaisen kielen suomen pitkä ja äidinkielenomainen oppimäärä (Toropainen 2010). Tuoreimmissa arvioinneissa on mukana ollut myös puhetehtäviä, joita on suorittanut vähintään 10–20 % otos oppilaista. Englannin osaaminen on ollut keskimäärin tavoitteiden mukaista tai parempaa (Tuokko 2007), mutta ruotsin keskipitkän oppimäärän tavoitteiden saavuttamisessa on havaittu ongelmia (Tuokko 2009). Suomen pitkän oppimäärän arvioinnissa toisena kotimaisena kielenä hyvän osaamisen ta-

voitetaso ylittyi puhumisessa ja alittui kirjoittamisessa. Äidinkielenomaisessa suomessa tulokset olivat kaiken kaikkiaan paremmat. (Toropainen 2010.)

Viimeisin ja laajin vieraiden kielten oppimistulosten arviointi toteutettiin 2013, jolloin mukana oli myös toisen kotimaisen kielen ruotsin A-oppimäärä. Siinä kaikkien kielten oppimistulokset raportoitiin ensimmäisen kerran taitotasoinnain, mitä aiemmin oli tehty vain englannin (Tuokko 2009) ja toisen kotimaisen kielen suomen osalta (Toropainen 2010). Vuoden 2013 arvioinnin tulokset tiivistävät kuluvaan vuosituhannen ensimmäisen opetussuunnitelmakerroksen yleissivistävän peruskoulutuksen tuottaman kielitaitovarannon. Pitkistä oppimääristä englannin osaaminen ylitti asetetut tavoitteet puhumisessa ja kirjoittamisessa ja ruotsin kuullun ymmärtäminen ja puhuminen olivat erinomaista tasoa, kuten myös saksan ja venäjän luetun ymmärtäminen. Perusopetuksen yläluokilla alkaneista oppimääristä saksan tavoitteet ylittyivät kuullun ja luetun ymmärtämisessä ja puhumisessa sekä venäjän luetun ymmärtämisessä. Tavoitetaso alittui ranskan pitkän oppimäärän kuullun ymmärtämisessä ja kirjoittamisessa ja venäjän puhumisessa. Osaaminen vaihteli useimmissa oppimäärissä sukupuolen, vanhempien koulutustaustan, koulun opetuskielen ja jatko-opintosuunnitelmien mukaan sekä jonkin verran alueittain. Tytöt suoriutuivat tehtävistä poikia paremmin, ja ylioppilasvanhempien lapset osoittivat parempaa kielitaitoa kuin ei-ylioppilastaustaisten vanhempien lapset. Lukioon aikovien tulokset olivat paremmat kuin ammatilliseen koulutukseen aikovien, ja ruotsinkielisten koulujen oppilaat osasivat kaikkia kieliä keskimäärin paremmin kuin suomenkielisten koulujen oppilaat. Oppilaiden saamat arvosanat vastasivat arvioinnissa osoitettua taitotasoa parhaiten pitkissä oppimäärissä. (Hildén ym. 2014.)

Vuoden 2013 kansallisen arvioinnin pohjalta tehtyjä johtopäätöksiä hyödynnettiin laajasti vuoden 2015 opetussuunnitelman perusteissa, joissa englannin pitkän oppimäärän hyvän osaamisen tasoa korotettiin ja kielten opiskeluun annettiin suosituksia esimerkiksi tietotekniikan integroimiseksi opiskeluun entistä luontevammin. Formatii-visen arvioinnin ja palautteen muotojen kehittämiseksi muistutettiin eurooppalaisen kielisalkun mahdollisuuksista. (Hildén ym. 2014; POPS 2014.)

3.2 Yleiset kielitutkinnot

Yleiset kielitutkinnot (YKI) on osallistujamäärältään suurin koulutusjärjestelmän ulkopuolinen kielitaidon arviointijärjestelmä Suomessa. Tutkintojen kehittämistyöstä vastaavat Opetushallitus sekä Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus. Kyse on näyttötutkinnosta, johon voi osallistua kuka tahansa riippumatta siitä, miten on kielitaitonsa hankkinut. Tutkinto on tarkoitettu aikuisille toisen tai vieraan kielen käyttäjille, ja siinä testataan ns. yleiskielitaidon hallintaa puhumisessa, kirjoittamisessa, puheen ymmärtämisessä ja tekstin ymmärtämisessä. Tutkintokieliä on yhteensä yh-

deksän: englantia, espanjaa, italiaa, pohjoissaamea, ranskaa, ruotsia, saksaa, suomea ja venäjää. Useimmista kielistä on tarjolla kolme eritasoista tutkintoa, joista perustaso on suunnattu EVK:n taitotasolle A1 ja A2, keskitaso tasoille B1 ja B2 ja ylin taso C1 ja C2 -tasolle.

Pohja yleisille kielitutkinnoille luotiin Jyväskylän yliopistossa kielitaidon arvioinnin hyväksi tehdyllä työllä, josta varhaisimpia esimerkkejä olivat edellä kuvatut kielten oppimistulosten arviointitutkimukset silloisessa Kasvatustieteiden tutkimuslaitoksessa (KTL, nykyään Koulutuksen tutkimuslaitos; esim. Takala 1971). Kielitaidon arviointia kehitettiin jo varsin varhain eri laitosten välisenä yhteistyönä. 1970-luvulla perustettujen kielikeskusten yhteisiä tekstien ymmärtämisen tasotestejä suunniteltiin ja tutkittiin Korkeakoulujen kielikeskuksen, KTL:n ja englannin kielen laitoksen yhteistyönä jo 1980-luvun alusta lähtien. Vuosikymmenen lopulla nämä kolme laitosta kehittivät työelämän kielitutkinnon, jossa monella tapaa kehitettiin ja kokeiltiin käytännössä sellaisia toimintatapoja ja välineitä, joita sitten hyödynnettiin YKIn kehittämisessä (Huhta, Sajavaara & Takala 1993). Erityisesti suomen kielen YKI-testistön taustalla olivat myös tasotestistöt, joita 1990-luvun alussa kehitettiin ulkomailta Suomeen kielikursseille tuleville suomen kielen oppijoille. Ne laadittiin Jyväskylän yliopiston kielikeskuksen ja Korkeakoulujen kielikeskuksen yhteistyönä (Martin, Aalto, Huhta, Siitonen & Suni 1994).

Kun yleisiä kielitutkintoja ryhdyttiin toden teolla suunnittelemaan vuonna 1992 Opetushallituksen ja Jyväskylän yliopiston kesken, hankkeella oli siis jo vankka perusta. Jyväskylässä yhteistyö Korkeakoulujen kielikeskuksen, KTL:n, yliopiston kielikeskuksen ja englannin kielen laitoksen välillä jatkui, vaikka tutkinto sijoitettiin Korkeakoulujen kielikeskukseen, jonka nimi vuonna 1996 muutettiin Soveltavan kielitutkimuksen keskeiseksi (Solki).

Yleisten kielitutkintojen perustamiseen oli monta toisiinsa liittyvää syytä, joille yhteistä oli halu edistää aikuisten kieltenopiskelua. Voidaan siis sanoa, että pääsyy perustaa tutkinto oli saada aikaan myönteisiä vaikutuksia kielikoulutukseen. Keskeinen pyrkimys oli edistää mahdollisimman monien eri kielten opiskelua tarjoamalla kotimainen, koko maassa saatavilla oleva mahdollisuus osoittaa kielitaito erilaisia tarkoituksia varten. Tarkoitukset voivat vaihdella palautteen saamisesta omasta osaamisesta tutkintotodistuksen käyttöön työ- tai opiskelupaikan hankkimiseen tai työllä etenemiseen. Suomessa järjestetään kyllä joitakin kansainvälisiä kielitutkintoja, mutta vain joissakin suurimmissa kielissä ja vain muutamalla paikkakunnalla – joskus vain yhdessä paikassa koko maassa. Saatavilla olevaa kielivalikoimaa haluttiin laajentaa myös vähemmän opiskeltuihin kieliin ja pääsyä kokeisiin helpottaa. Heti alusta pitäen YKIn oli tarjolla seitsemän eri kielen testit.

YKIn perustamisella pyrittiin myös vaikuttamaan myönteisesti aikuisten kielenopetukseen Suomessa (ks. esim. Luoma & Tarnanen 1997). YKI edustaa toiminnallisen kielitaidon arviointia, eli se perustuu samanlaiseen käsitykseen kielitaidosta kuin edellä kuvattiin yleissivistävän koulutuksen opetussuunnitelmien yhteydessä. Siinä oppijat

joutuvat osoittamaan, miten hyvin he selviytyvät erilaisista luonnollisia viestintätilanteita ja tarkoituksia muistuttavista tehtävistä vieraalla tai toisella kielellä. Puhumisen testi oli alusta pitäen olennainen osa tutkintoa huolimatta sen järjestämiseen liittyvistä käytännön haasteista. Haluttiin, että kommunikatiivinen kielitutkinto vaikuttaisi viestinnällisyyden lisääntymiseen myös aikuisten kielinopetuksessa ja arvioinnista (ks. Harlen 2008).

Yksi YKIn epävirallisista tavoitteista oli kehittää kielitaidon arviointia tutkimusalueena Suomessa. Siksi tutkimusperusteinen kehittäminen on ollut alusta lähtien kiinteä osa YKIn toimintaa (Halvari & Tarnanen 1997; Huhta & Lammervo 2014), ja se on osaltaan vaikuttanut siihen, että kielitaidon arviointi on yksi Jyväskylän yliopiston tutkimuksen kärkialaksi nimetyn soveltavan kielitutkimuksen keskeisiä osa-alueita (ks. RECLAS 2017). Yleiset kielitutkinnot ja niiden taustalla ollut pitkäjänteinen kehittäminen ovat edesauttaneet myös suomalaisten arviointitutkijoiden osallistumista erilaisiin kansainvälisiin kielitaidon arviointihankkeisiin, joista merkittävin oli vuosina 1997–2004 kehitetty monikielinen verkon kautta toimiva diagnostinen DIALANG-arviointijärjestelmä (ks. esim. Alderson 2005). Yleiset kielitutkinnot – tai tarkemmin sanottuna niistä vastuussa olevat organisaatiot – ovat 1990-luvun puolivälistä lähtien lisäksi olleet eurooppalaisten kielitutkintojen yhteistyöjärjestön ALTE:n (Association of Language Testers in Europe) jäsenenä.

Yleiset kielitutkinnot ovat kansallisia ja virallisia kielitutkintoja, joita säätelevät yleisistä kielitutkinnoista annettu laki (964/2004) ja asetus (1163/2004). Tutkinnon alkuvuosina sen tulokset raportoitiin 9-portaisella asteikolla, joka muutettiin 6-portaiseksi vuonna 2002, jotta voitiin helpottaa vertailua Eurooppalaisen viitekehyksen 6-portaiseen asteikkoon (Leblay, Lammervo & Tarnanen 2014: 9). Tutkinnon osallistujamäärä on kasvanut huomasti sen olemassaolon aikana. Alkuvuosina osallistujia oli vain muutama tuhat vuodessa; ensimmäisten kolmen toimintavuoden aikana yhteensä 5 297 henkilöä (Salmela & Tarnanen 1997). Lähiaikoina osallistujamäärä rikkoo 10 000 vuotuisen osallistujan rajan. Yhteensä tutkintoon osallistui sen ensimmäisten 20 vuoden aikana (1994–2014) noin 80 000 henkilöä (Neittaanmäki & Hirvelä 2014). Osallistujamäärä on kasvanut voimakkaasti vuodesta 2003 lähtien, jolloin uusi kansalaisuuslaki (359/2003) tuli voimaan luoden kertaheitolla uuden, tärkeän syyn suorittaa YKI-tutkinto. Lain mukaan Suomen kansalaisuuden edellytyksenä oleva suomen tai ruotsin kielitaitovaatimus voidaan osoittaa yleisten kielitutkintojen keskitason tutkinnolla, mikä moninkertaisti suomen tutkinnon vuosittaisen osallistujamäärän yli 6 000:een vuoteen 2013 mennessä (Neittaanmäki & Hirvelä 2014: 46). Osallistujamäärä on vielä siitäkkin kasvanut.

Maahanmuuttajien lisäksi erityisiä YKIn suorittajaryhmiä on puolustusvoimien henkilökunta (ks. tarkemmin Saarelainen 2014). Puolustusvoimat on määritellyt YKIn suorittamisen yhdessä vieraassa kielessä edellytykseksi kansainvälisiin tehtäviin tai kou-

lutuskokonaisuuksiin hakeutumiselle. Useimmiten kyse on englannin tutkinnon suorittamisesta. Lisäksi tutkintoa käytetään nykyisessä tai tulevassa tehtävässä tarvittavien kielten hallinnan osoittamiseen. Toinen erityinen YKIn suorittajaryhmä on ulkomainen terveydenhuollon henkilökunta, joka tulee Suomeen työhön (Anttila 2014). Kielitaitovaatimuksista EU- ja ETA-maiden ulkopuolelta tulevalle henkilökunnalle päättää Valvira, joka on sosiaali- ja terveysministeriön hallinnonalan keskusvirasto, jonka tehtävänä on muun muassa potilasturvallisuuden varmistaminen. Osa potilasturvallisuutta on luonnollisesti riittävä suomen tai ruotsin kielen taito. Valvira käyttää hakijoiden kielitaidon riittävyden arvioimiseen YKIn ja Valtionhallinnon kielitutkintotodistuksia. Anttilan (2014) mukaan monet työnantajat vaativat YKI-tutkinnon suorittamista, vaikka se ei välttämättä yleiskielitaitoa mittaavana tutkintona olekaan paras mahdollinen tällaisen erityisalan kielitaidon arviointiin (ks. Seilonen ym. tässä julkaisussa).

Alun perin YKlissä oli viisi osakoetta: puhuminen, puheen ymmärtäminen, kirjoittaminen, tekstin ymmärtäminen ja sanaston ja rakenteiden hallintaa mittaava koe. Vuodesta 2011 lähtien tutkinnossa ei enää ole ollut erillistä sanastoa ja rakenteita testaavaa osaa, vaan niiden hallintaa arvioidaan osana muiden taitojen testejä. Samalla luovuttiin kielitaidon yleistason raportoinnista tutkintotodistuksessa, jossa nykyään kerrotaan osallistujan saavuttama taitotaso neljällä kielen hallinnan osa-alueella. Todistusten käyttäjät voivat päättää itse, miten he hyödyntävät tätä tarkempaa tietoa osaamisesta. Nämä muutokset ovat osa laajempaa ja pitempiaikaista kielitutkintotrendiä, jonka mukaan ei enää aina testata kielen peruselementtejä erillään laajemmista taitokokonaisuuksista vaan annetaan profilitodistuksia. Kokonaisarvosanoista luopuminen on tosin vielä harvinaista. Yleistasoarviosta luopumisesta huolimatta esimerkiksi Maa- ja metsätalouden tutkimuskeskus käyttää edelleen YKI-tuloksia kansalaisuuteen vaadittavan kielitaidon osoittamiseen. Virasto on päättänyt, että kansalaisuuden hakemista varten riittää, että kielitaito on osoitettu siten, että jokin seuraavista kolmesta testiyhdistelmästä on suoritettu taitotasolla 3 (eurooppalainen B1-taso): 1) kirjoittaminen ja puhuminen, 2) kirjoittaminen ja puheen ymmärtäminen tai 3) puhuminen ja tekstin ymmärtäminen (ks. Soramäki 2014).

4 Suomalaisten kielitaidon arviointijärjestelmien muutokset ja tulevaisuuden näkymät

Kielitaidon arvioinnin näkyvintä ja terävintä kärkeä edustavat erilaiset kielitutkinnot, joilla on merkittävä vaikutus niihin osallistuvien elämään. Onkin selvää, että juuri tutkintoja koskevat tiukimmat laatuvaatimukset: niiden tulee mm. rakentua perustellulle käsitykselle kielitaidosta ja mitata osaamista luotettavasti ja oikeudenmukaisesti (ks. esim. Bachman & Palmer 2010). Tutkintojen kaltaisten tärkeiden arviointien vaiku-

tuksista on jo varsin pitkään keskusteltu kielitaidon arviointikirjallisuudessa (ks. esim. Alderson & Wall 1993; Cheng & Watanabe 2004; Wall & Horák 2008). Tutkinnolla tulisi mielellään olla myönteisiä vaikutuksia sekä yksilöiden että organisaatioiden ja koko yhteiskunnankin tasolla – tai ainakaan sillä ei saisi olla merkittäviä kielteisiä vaikutuksia. Myös muun tyyppiset laajat arvioinnit, kuten kansalliset oppimistulosten arvioinnit, ovat tärkeitä ainakin koulutusjärjestelmälle ja siitä vastuussa oleville, elleivät yksittäisille osallistujilleen. Tällaistenkin arviointien laatuvaatimukset ovat korkeat, koska niihin uhrataan paljon resursseja ja tulosten perusteella saatetaan tehdä merkittäviä muutoksia tuleviin opetussuunnitelmiin.

Merkittävien arviointijärjestelmien laatua pyritään varmistamaan monin eri tavoin. Suurin osa tästä työstä tapahtuu kulissien takana osana uusien testitehtävien laadintaa, tulosten analyysiä, arvioijien koulutusta ja lukemattomia muita arviointiprosessin vaiheita. Osa laadunvarmistuksesta on kuitenkin niin yleisesti kiinnostavaa ja tärkeää, että se on julkaistavissa tieteellisissä julkaisuissa. Tämän julkaisun artikkelit ovat hyviä esimerkkejä tutkimuksista, jotka voivat paitsi parantaa kyseisten arviointijärjestelmien laatua myös edistää kielitaidon arviointia tieteenalana – ja tätä kautta osaltaan edistää arviointien myönteisiä vaikutuksia.

Kuten edellä olevista ylioppilastutkinnon, yleisten kielitutkintojen ja oppimistulosten arviointien kuvauksista käy ilmi, niillä on kaikilla varsin pitkä historia, ja ne ovat kokeneet monenlaisia muutoksia. Yksi keskeisimmistä kielitaidon arvioinnin ominaisuuksista on se näkemys kielestä ja kielitaidosta, jolle arviointi perustuu. Tällä ajatellaan olevan suuri osuus siinä, millainen vaikutus kielitutkinnoilla ja muulla kielitaidon arvioinnilla on (ks. esim. Cheng & Watanabe 2004). Selvimmin kielikäsitusten muuttuminen näkyy tarkastelluista arviointijärjestelmistä vanhimmassa eli ylioppilastutkinnossa, joka luotiin aikana, jolloin kielitaidon arviointi tai ylipäänsä testaaminen ei nykyaikaisessa muodossaan ollut edes syntynyt (ks. tarkemmin Spolsky 1995). Ylioppilastutkinnon uudistus, joka alkoi jo 1960-luvulla ja saatettiin päätökseen seuraavalla vuosikymmenellä, oli luonteeltaan erittäin radikaali, ja siinä siirryttiin varsin nopeasti 1800- ja 1900-lukujen kielikäsituksista uuteen viestinnälliseen ajatteluun ja arviointiin. Uudistuksen taustalla oli se kielenopetuksen kentällä alati vahvistunut käsitys, ettei vanhamuotoinen ylioppilastutkinto enää vastannut uusia käsityksiä kielitaidosta; tutkintoa painostettiin uudistamaan (Kaarninen & Kaarninen 2002: 293–295). Kielikoulutuksen ammattilaiset olivat selvästi sitä mieltä, että ylioppilastutkinto vaikutti lukioiden kielenopetukseen, tässä tapauksessa siis kielteisesti. Muutoksella haluttiin ainakin poistaa tuo haittavaikutus, mutta epäselväksi jää, missä määrin ylioppilastutkintolautakunta pyrki tietoisesti myös edistämään viestinnällistä kielenopetusta uudistuksen avulla (ks. yleisten kielitutkintojen tavoitteiden kuvaus tässä artikkelissa). Vain puhumistaidon arvioinnin puuttuminen aiheutti särön muuten hyvin ajan hermolla olleeseen uudistukseen. Tämän puutteen

vaikutuksista onkin sitten keskusteltu paljon, ja lukuisia selvityksiä puhumisen testin lisäämisestä tutkinnon kielikokeisiin on tehty, kuten edellä on kuvattu.

Oppimistulosten arvioinnit ja yleiset kielitutkinnot kehitettiin puolestaan aikana, jolloin kommunikatiivinen kielenopetus ja arviointi olivat jo yleistymässä, joten niiden perustana oleva käsitys kielitaidosta oli alusta pitäen suhteellisen moderni. Ainakin yleisten kielitutkintojen kohdalla sen pohjana oleva toiminnallinen kielikäsitys oli tärkeä peruste sille, että tutkinnolla toivottiin olevan myönteisiä vaikutuksia sellaisella kielikoulutuksen sektorilla, jolla toiminnallinen kielitaitokäsitys ei vielä ollut levinnyt kovin laajalle. Toki kukin tarkastelluista järjestelmistä on ottanut tai ottamassa lisäaskeleita, jotka tähtäävät niiden viestinnällisyyden korostumiseen. Ylioppilastutkintoon todennäköisesti lisäävät kielten puhumistestit ovat tästä selvin esimerkki, mutta myös yleisten kielitutkintojen luopuminen erillisestä sanaston ja rakenteiden hallintaa mitaavasta testistä voidaan nähdä askeleena kielen käyttötaidon painottumiseen tutkinnossa.

Yksi arvioinnin taustalla oleva kielitaitoon kiinteästi liittyvä ilmiö, joka näkyy vahvasti suomalaisessa arviointitoiminnassa, on Eurooppalainen viitekehys. Suomi on ylipäänsä ollut yksi edelläkävijöistä EVK:n hyödyntämisessä kielikoulutuksessa. Voidaan siis perustellusti väittää, että EVK on vaikuttanut vahvasti sekä kielitaidon arviointiin että yleensä kielikoulutukseen maassamme. Toisaalta vaikutus on ollut molemminpuolista: Suomi on myös antanut merkittävän panoksen viitekehukseen liittyvään kansainväliseen kehittämistyöhön; tästä voidaan mainita DIALANG-arviointijärjestelmä (ks. Alderson 2005), kielitestien EVK-linkitysoapas (ks. Figueras, North, Takala, Verhelst & Van Avermaet 2005) ja verkkoresurssi EVK:sta opettajankoulutuksen tueksi (CEFTain 2005). EVK:n taitotasot edustavat yhtä tapaa määritellä kielitaito, ja tätä varsin uutta lähestymistapaa on Suomessa näkyvimmin hyödynnetty opetussuunnitelmissa 2000-luvun alkupuolta lähtien, kuten toisaalla tässä artikkelissa kuvataan (ks. myös Hildén & Takala 2007). Onkin johdonmukaista, että opetussuunnitelmien taitotasokuvauksilla on ollut keskeinen osa uusimmissa kielten oppimistulosten arvioinneissa (esim. Tuokko 2007; Hildén ym. 2014). Tasokuvauksia on hyödynnetty sekä arviointitehtävien laadinnassa, suoritusten ja tehtävien tasojen arvioinnissa että tulosten raportoinnissa. Erityisesti ymmärtämistestien tulosten raportointi on tullut tämän ansiosta huomattavasti läpinäkyvämmäksi; ennen ne jouduttiin raportoimaan vain keskimääräisinä ratkaisuosuuksina, joita on vaikea tulkita. Myös yleisissä kielitutkinnoissa hyödynnetään EVK:n tasokuvauksia. Kuten aikaisemmin todettiin, YKIn tulokset raportoidaan asteikolla, joka rinnastuu EVK-asteikkoon. Lisäksi EVK:n tasokuvaimia hyödynnetään ymmärtämistehtävien verantamisessa, jota kuvataan tarkemmin Aholan ym. artikkelissa tässä julkaisussa.

Aivan kaikki uudemmat kielikäsitukset eivät ainakaan vielä näy suomalaisissa arviointijärjestelmissä. Esimerkiksi näkemys siitä, että kielitaito ei ole vain yksilön ominaisuus vaan kahden tai useamman keskustelukumppanin yhteinen luomus (ks. Roever

ja McNamara 2006) ei suoraan ilmene tässä kuvatuissa arvioinneissa, paitsi ehkä silloin, kun puhumistaidon arviointi on joko kokonaan tai osittain perustunut parikeskustelun kaltaisille testitehtäville. Monikielinen viestintätaito on toinen viime vuosina esille noussut näkemys kielitaidosta (esim. Shohamy 2011), jota ei Suomessa eikä juuri muuallakaan ole vielä yritetty soveltaa arviointiin.

Teknologian käyttö kielitaidon arvioinnissa on merkittävä uusi trendi, jolla on paitsi käytännön vaikutuksia myös periaatteellisia, kielikäsitteeseen liittyviä seurauksia (ks. Chapelle 2006; Chapelle & Voss 2016). Ylioppilastutkinto on jo sähköistymässä, ja myös kansalliset oppimistulosten arvoinnit ja yleiset kielitutkinnot ovat matkalla tähän suuntaan. Testien suorittaminen tietokoneella vaikuttaa väistämättä myös mitattavaan kielitaitoon varsinkin, kun arvioinnissa käytettäviä tehtävätyyppejä ollaan samalla monipuolistamassa. Ehkä selvin esimerkki tästä on puheen ymmärtämisen testaaminen niin, että syötökseenä käytetään äänen lisäksi myös kuvia tai jopa videonauhoitetta. Tällainen multimediaesityksen ymmärtäminen ei luonnollisesti ole täysin sama asia eikä vaadi täsmälleen samanlaista osaamista kuin pelkän äänitteen kuuntelu (ks. Baty 2015). On myös muistettava, että pelkästään kirjoittaminen näppäimistön avulla ei täysin vastaa käsin kirjoittamista eikä puhuminen tietokoneella olevan äänitallenteen perusteella vastaa kasvokkain toisen ihmisen kanssa tapahtuvaa puhumista. Se, miten ja kuinka paljon tietotekniikka muuttaa testattavaa kielitaitoa, on vielä varsin vähän tarkasteltu asia Suomessa, mutta sitä tullaan varmasti tutkimaan paljon tulevaisuudessa. Kielitaidon(kin) arvioinnilla on monenlaisia ennakoituja ja ennakoimattomia vaikutuksia, joista osa on toivottuja ja osa vähemmän toivottuja (ks. tarkemmin Tossavaisen ja Martikaisen artikkelit tässä julkaisussa; ks. myös Wall & Horák 2008). Kokonaisuutena katsoen kielitaidon arviointien vaikutuksia ei Suomessa ole kovin paljon tutkittu. Edellä kuvattujen arviointijärjestelmien ennakoitujen – ja osin myös toteutuneiden – vaikutukset on parhaiten dokumentoitu yleisistä kielitutkinnoista, joten käsittelemme niitä tässä lopussa hiukan tarkemmin.

Kuten edempänä todettiin, yleisillä kielitutkinnoilla pyrittiin jo alkuaan moniin päämääriin, joista tärkein oli edistää viestinnällisempää kielenopetusta ja arviointia aikuisten kielikoulutuksessa Suomessa. Todettakoon, että myös inkeriläisten paluumuuttajien kielitutkinnoilla oli osin samanlaisia, paluumuuttovalmennusta koskevia tavoitteita, kuten Martikainen toteaa artikkelissaan toisaalla tässä julkaisussa. Valitettavasti tämän tavoitteen toteutumista ei ole tutkittu järjestelmällisesti YKI-tutkinnon osalta, joten on mahdotonta sanoa, missä määrin se on saavutettu. Epäsuorasti voidaan kuitenkin päätellä, että YKI on ainakin osittain onnistunut levittämään kommunikatiivista arviointikäsitteistä ja etenkin taitotasoarvioinnin tuntemusta. Peruste tälle on se, että YKIssä tapahtuva arvioijien, haastattelijoiden ja tehtävänlaatijoiden säännöllinen koulutus on merkittävin maassamme järjestettävä säännöllinen arviointikoulutus aikuisten – erityisesti maahanmuuttajien – kielenopettajille. Vuosien mittaan tähän koulutukseen

on osallistunut useita satoja opettajia. Toinen YKIn yleistavoite, jonka toteutumista on vaikea arvioida, on kielikoulutuksen monipuolisuuden tukeminen tarjoamalla mahdollisuus suorittaa tutkinto monessa eri kielessä. Toki tutkinnon eri kieliin on osallistunut huomattava määrä oppijoita, mutta on mahdotonta arvioida, mikä YKIn vaikutus on ollut esimerkiksi haluun opiskella jotakin ns. harvinaisempaa kieltä.

Muita YKIlle asetettuja tavoitteita on jonkin verran helpompi arvioida. Tutkintoa voi suorittaa hyvin monella eri paikkakunnalla, mikä on tietysti helpottanut siihen pääsyä. YKI on myös osaltaan edistänyt kielitaidon arvioinnin kehittymistä omaksi – vaikkakin pienehköksi – tieteenalakseen maassamme ja avannut erilaisia kansainvälisiäkin yhteistyömahdollisuuksia (mm. DIALANG-hanke) ja jäsenyyksiä kansainvälisissä arviointijärjestöissä, kuten ALTEssa. YKI-tutkintoon liittyvää tutkimusta on myös esitelty lukuisissa konferensseissa, ja siitä on raportoitu kansallisissa ja kansainvälisissä julkaisuissa.

Lopuksi on kuitenkin todettava, miten vaikea kaikkia vaikutuksia on ennustaa ja miten todella suuriakin seurauksia tutkinnosta voi olla. Alun perin yleiset kielitutkinnot suunniteltiin suomalaisten kielienoppijoiden vaihtoehdoksi kansainvälisille kielitutkinnoille. Tutkinnon käyttö kansalaisuuteen vaadittavan kielitaidon osoittamiseen ei luultavasti edes tullut sen suunnittelijoiden mieleen muuten kuin ehkä yhtenä varsin marginaalisena syynä osallistua tutkintoon. Poliittisten päätösten seurauksena kielitaitoa ja kansalaisuutta koskeva lainsäädäntö kuitenkin muuttui, joten yleisistä kielitutkinnoista tuli tavallisin tapa osoittaa vaadittu suomen tai ruotsin kielen taito. Vuosina 2012–13 lähes 80 % suomen kielen YKI-tutkinnon suorittaneista kertoi tarvitsevansa todistusta kansalaisuuden hakemista varten (Neittaanmäki & Hirvelä 2014: 50). Vaikka moni ilmoitti käyttävänsä sitä myös muihin tarkoituksiin, tilasto osoittaa selvästi, että tärkein syy suorittaa tutkinto on siis muuttunut radikaalisti ja ennakoimattomasti siitä, mitä alun perin oli oletettu. Tämä havainnollistaa, miten vaikea on ennustaa arvioinnin vaikutuksia edes sellaisessa tutkinnossa, jonka suunnittelun keskeisenä lähtökohtana oli pyrkimys saada aikaan erilaisia myönteisiä vaikutuksia suomalaisen kielikoulutukseen ja arviointiin.

Kirjallisuus

Alderson, J.C. 2005. *Diagnosing foreign language proficiency: the interface between learning and assessment*. New York: Continuum.

Alderson, J.C. & D. Wall 1993. Does washback exist? *Applied Linguistics*, 14 (2), 115–29.

Anttila, I. 2014. Valviran käyttökokemukset YKI-tutkinnoista. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus, 44–45.

Bachman, L. & A. Palmer 2010. *Language assessment in practice*. Oxford: Oxford University Press.

- Batty, A. 2015. A comparison of video- and audio-mediated listening tests with many-facet Rasch modeling and differential distractor functioning. *Language Testing*, 31 (1), 3–20.
- Brown, A. 2013. Uses of language assessments. Teoksessa C. Chapelle (toim.) *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley.
- CEFTTrain 2005. <http://www.helsinki.fi/project/ceftrain/index.html> Luettu 20.12.2016.
- Chapelle, C. 2006. *Assessing language through computer technology*. Cambridge: Cambridge University Press.
- Chapelle, C. & E. Voss 2016. 20 years of technology and language assessment in Language Learning & Technology. *Language Learning & Technology*, 20 (2), 116–128.
- Cheng, L. & Y. Watanabe 2004. *Washback in language testing. Research contexts and methods*. New York: Routledge.
- DigiTala 2016. <http://blogs.helsinki.fi/digitala-projekti/tietoa-digitalasta/> Luettu 20.12.2016
- Euroopan neuvosto 2001. *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- EVK 2003 = Huttunen, I. & H. Jaakkola, H. (2003). *Eurooppalainen viitekehys: kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys (suomennos)*. Alkuteos: Common European framework of reference for languages: learning, teaching, assessment. Cambridge: Cambridge University Press. Helsinki: WSOY
- Figueras, N., B. North, S. Takala, N. Verhelst & P. Van Avermaet 2005. Relating examinations to the Common European Framework: a manual. *Language Testing*, 22 (3), 261–279.
- Fulcher, G. 2003. *Testing second language speaking*. New York: Routledge.
- Fulcher, G. & F. Davidson 2009. Test architecture. Test retrofit. *Language Testing*, 26 (1), 123–144.
- Hakala, P. & Kärkkäinen, K. 1976. *Ruotsin kielen kouluvaikutukset peruskoulun ala-asteen ja yläasteen päättyessä lukuvuonna 1972–1973*. Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita 74/1976. Jyväskylän yliopisto.
- Halvari, A. & M. Tarnanen 1997. Yleiset kielitutkinnot tutkimuksen tiellä. Teoksessa A. Huhta, A. Määttä, S. Takala & M. Tarnanen (toim.) *Yleiset kielitutkinnot ja kielenopetus*. Helsinki: Opetushallitus, 115–124.
- Harlen, W. 2008. *Assessment of learning*. Thousand Oaks, CA: Sage.
- Havola, L. & Saari, H. 1993. Kaksikymmentä vuotta englannin kuullun ymmärtämistä – onko taidoissa edistytty? Teoksessa P. Linnakylä & H. Saari (toim.) *Oppiiko oppilas peruskoulussa?* Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 157–173.
- Havola-Pitkänen, L. 1974. *Englannin kielen kouluvaikutukset peruskoulun ala-asteen ja yläasteen päättyessä*. Jyväskylän yliopisto, Kasvatustieteiden tutkimuslaitoksen julkaisuja 231.
- Hildén, R. & S. Takala 2007. Relating descriptors of the Finnish school scale to the CEF overall scales for communicative activities. Teoksessa Koskensalo, A., Smeds, J., Kaikkonen, P. & Kohonen, V. (toim.) *Foreign languages and multicultural perspective in the European context*. Münster: LIT Verlag, 291–300.
- Hildén, R., M. Härmälä, J. Rautopuro, M. Huhtanen & C. Silverström 2014. *Kielten oppimistulokset perusopetuksen päättövaiheessa*. Helsinki: Opetushallitus ja Kansallinen koulutuksen arviointikeskus.
- Huhta, A. & R. Hildén 2013. Kielitaidon arvioinnin metodologisia vaihtoehtoja. Teoksessa A. Räisänen (toim.) *Oppimisen arvioinnin kontekstit ja käytännöt*. Helsinki: Opetushallitus, 159–186.
- Huhta, A. & T. Lammervo 2014. Yleisten kielitutkintojen kehittäminen ja tutkimus. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus, 169–176.

- Huhta, A. & S. Takala 1999. Kielitaidon arviointi. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Kielenoppimisen kysymyksiä*. Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto, 179–228.
- Huhta, A., K. Sajavaara & S. Takala 1993. Recent developments in national examinations in Finland. Teoksessa A. Huhta, K. Sajavaara & S. Takala (toim.) *Language testing: new openings*. Kasvatustieteiden tutkimuslaitos: Jyväskylän yliopisto, 136–159.
- Hymes, D. H. 1972. On communicative competence. In: J. B. Pride & J. Holmes, J. (toim.) *Sociolinguistics: selected readings*. Harmondsworth: Penguin, 269–293.
- Jakku-Sihvonen, R. 2014. *Arviointia akanvirrassa*. Opetushallituksen arviointitoiminnan historiikki. Raportit ja selvitykset 2014: 14. Helsinki: Opetushallitus.
- Juurakko-Paavola, T. & S. Takala 2013. *Ylioppilastutkinnon kielikokeiden tulosten sijoittaminen Lukion opetussuunnitelman perusteiden taitotasolle*. Helsinki: Ylioppilastutkintolautakunta. https://www.ylioppilastutkinto.fi/images/sivuston_tiedostot/Raportit_tutkimukset/Fl_2013_kielikokeet_taitotasot.pdf.
- Kaarninen, M. & P. Kaarninen 2002. *Sivistyksen portti*. Ylioppilastutkinnon historia. Keuruu: Otava.
- Kansalaisuuslaki 359 / 2003. <http://www.finlex.fi/fi/laki/ajantasa/2003/20030359>
- Karppinen, M. & Sarkkinen, R. 1995. *Ruotsin kielen oppimistuloksia peruskoulun 9. luokan valtakunnallisessa kokeessa 1994*. Monistesarja 23/1995. Helsinki: Opetushallitus.
- Kupiainen, S., J. Marjanen & J. Hautamäki 2016. The problem posed by exam choice on the comparability of results in the Finnish matriculation examination / Das Problem der Wahlmöglichkeiten von Prüfungsfächern hinsichtlich der Vergleichbarkeit von Ergebnissen der zentralen Abschlussprüfung in Finnland. *Journal for Educational Research Online*, 8 (2), 87.
- Kurssimuotoisen lukion oppimäärät 1981*. Helsinki: Kouluhallitus.
- Kuusela, J. 2003. *Koulujen paremmuusjärjestyksestä*. Moniste 7/2003. Helsinki: Opetushallitus.
- Lado, R. 1961. *Language testing. The construction and use of foreign language tests. A teacher's book*. Longman.
- Kärkkäinen, K. 1983. *Peruskoulun tilannekartoitus 1 1979. Ruotsin kielen rakennekokeen osiokohtaiset tulokset yhdeksännellä luokalla*. Kasvatustieteiden tutkimuslaitos. Selosteita ja tiedotteita 258/1985. Jyväskylän yliopisto.
- Leblay, T. 2014. Yleiset kielitutkinnot 2010-luvulla: mietteitä lainsäädännöstä, haasteista ja kehittämiskohteista. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus, 26–31.
- Leblay, T., T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus.
- Luoma, S. 2004. *Assessing speaking*. Cambridge: Cambridge University Press.
- Luoma, S. & M. Tarnanen 1997. Opetus ja testaus kielitaidon oppimisen asialla. Teoksessa A. Huhta, A. Määttä, S. Takala & M. Tarnanen (toim.) *Yleiset kielitutkinnot ja kielenopetus*. Helsinki: Opetushallitus, 62–72.
- LOPS 1985 = Lukion opetussuunnitelman perusteet 1985*. Helsinki: Kouluhallitus.
- LOPS 2003 = Lukion opetussuunnitelman perusteet 2003*. Määräys 33/011/2003. Helsinki: Opetushallitus.
- LOPS 2015 = Lukion opetussuunnitelman perusteet 2015*. Määräykset ja ohjeet 2015: 48. Helsinki: Opetushallitus.
- Lukiokoulutuksen suullisen kielitaidon arviointityöryhmän muistio 2006*. Opetusministeriön työryhmämuistioita ja selvityksiä 2006: 26. Helsinki: Opetusministeriö.
- Marjanen, J. 2015. Ylioppilasarvosanojen vertailukelpoisuusongelma ja SYK-menettelyn edellytykset sen ratkaisemiseen. *Kasvatus*, 46 (4), 317–333.

- Martin, M., E. Aalto, A. Huhta, K. Siitonen & M. Suni 1994. Developing National Proficiency Tests of Finnish as a second language: the process and problems. Teoksessa A. Golden & A. Hvenekilde (toim.) *Rapport fra det andre forskersymposiet om Nordens språk som andrespråk*. Oslo: Institut for lingvistiske fag, Universitetet i Oslo, 294–302.
- Neittaanmäki, R. & T. Hirvelä 2014. Yleisten kielitutkintojen osallistujat taustatietojen valossa. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus, 46–60.
- Nykykielten opetussuunnitelma* 1970. Helsinki: Kouluhallitus.
- POPS 1970 = Peruskoulun opetussuunnitelma. Peruskoulun opetussuunnitelmakomitean mietintö 1, 2, Opetussuunnitelman perusteet: Oppiaineiden opetussuunnitelmat*. Helsinki: Opetushallitus.
- POPS 1994 = Perusopetuksen opetussuunnitelman perusteet 1994*. Helsinki: Opetushallitus.
- POPS 2004 = Perusopetuksen opetussuunnitelman perusteet 2004*. Määräykset 1–3/011/2004. Helsinki: Opetushallitus.
- POPS 2014 = Perusopetuksen opetussuunnitelman perusteet 2014*. Määräykset ja ohjeet 2014: 96. Helsinki: Opetushallitus.
- RECLAS 2017 = Research Collegium for Language in Changing Society. Jyväskylän yliopisto <https://www.jyu.fi/hum/laitokset/kielet/en/research/profile/Profiling-initiative>
- Roever, C. & T. McNamara 2006. *Language testing: the social dimension*. Oxford: Blackwell.
- Saarelainen, P. 2014. YKI-tutkintojen käyttö puolustusvoimissa. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus, 43.
- Salmela, T. & M. Tarnanen 1997. Yleiset kielitutkinnot – kielitaitotesti moneen tarpeeseen. Teoksessa A. Huhta, A. Määttä, S. Takala & M. Tarnanen (toim.). 1997. *Yleiset kielitutkinnot ja kielenopetus*. Helsinki: Opetushallitus, 9–20.
- Shohamy, E. 2011. Assessing multilingual competencies: adopting construct valid assessment policies. *The Modern Language Journal*, 95 (3), 418–329.
- Soramäki, A. 2014. Yleisten kielitutkintojen käytettävyys Maahanmuuttoviraston näkökulmasta. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014:16. Helsinki: Opetushallitus, 35–42.
- Spolsky, B. 1995. *Measured words*. Oxford: Oxford University Press.
- Takala, S. 1971. *Kokeiluperskoulujen yhteiset vieraiden kielten kokeet 1970–71*. Kouluhallituksen kokeilu- ja tutkimustoimiston toimeksiantotutkimuksia. Kasvatustieteiden tutkimuslaitoksen julkaisuja 115/71. Jyväskylän yliopisto.
- Takala, S. 2004. Englannin kielitaidon tasosta Suomessa. Teoksessa K. Leimu (toim.), *Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa*. Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos, 255–275.
- Toropainen, O. 2010. *Utvärdering av läroämnet finska i den grundläggande utbildningen. Inlärningsresultaten i finska enligt A-lärokursen och den modersmålsinriktade lärokursen i årskurs 9 våren 2009*. Uppföljningsrapporter 2010: 1. Helsingfors: Utbildningsstyrelsen.
- Tulosoppimus 2016 = Opetus- ja kulttuuriministeriön ja Ylioppilastutkintolautakunnan tulosoppimus vuodelle 2016*.
- Tuokko, E. 2000. *Peruskoulun 9. vuosiluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi 1999*. Opetushallitus.
- Tuokko, E. 2002. *Perusopetuksen päättövaiheen ruotsin kielen oppimistulosten kansallinen arviointi 2001*. Helsinki: Opetushallitus.
- Tuokko, E. 2007. *Mille tasolle perusopetuksen englannin opiskelussa päästään? Perusopetuksen päättövaiheen kansallisen arvioinnin 1999 Eurooppalaisen viitekehyksen taitotasoihin*

- linkitetyt tulokset*. Jyväskylä Studies in Humanities 69. <https://jyx.jyu.fi/dspace/handle/123456789/13426>.
- Tuokko, E. 2009. *Miten ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008*. Oppimistulosten arviointi 2/2009. Helsinki: Opetushallitus
- Wall, D. & T. Horák 2008. *The impact of changes in the TOEFL® examination on teaching and learning in Central and Eastern Europe: phase 2, coping with change*. TOEFLiBT Research Report 05-2008. Princeton, NJ: Educational Testing Service.
- Väisänen, T. 2003. *Perusopetuksen 9. luokan A-kielenä opetettavan ranskan, saksan ja venäjän oppimistulosten kansallinen arviointi*. Opetushallitus.
- Ylioppilastutkintolautakunta 2016a. <https://www.ylioppilastutkinto.fi/fi/yleistae/historia>
Luettu 15.12.2016.
- Ylioppilastutkintolautakunta 2016b. https://www.ylioppilastutkinto.fi/images/sivuston_tiedostot/Kehittaminen/fi_arvosanojen_vertailtavuus.pdf. Luettu 15.12.2016.

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

Henna Tossavainen

Jyväskylän yliopisto

Kielitestien eettisyydestä ja oikeudenmukaisuudesta

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

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

Asiasanat: eettisyys, oikeudenmukaisuus, asiantuntijuus, kielitaidon arviointi

1 Johdanto

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

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

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

2 Eettisyyden ja oikeudenmukaisuuden määrittelyä

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

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

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

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

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

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

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

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

3.1 Eettisyyden näkyminen kielitaidon arvioinnin ja kielentestauksen alalla

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

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

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

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

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

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

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

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

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

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

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

3.2 Ammattilaistuminen ja ammattieettiset ohjeet

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

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

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

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

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

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

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

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

4 Vastuulliseen ja oikeudenmukaiseen toimintaan pyrkimistä

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

5 Päätäntö

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

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

Kirjallisuus

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

- Bachman, L.F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Canale, M. 1987. The measurement of communicative competence. *Annual Review of Applied Linguistics*, 8, 67–84. DOI: 10.1017/S026719050001033.
- Davies, A. 2004. Introduction: Language testing and the golden rule. *Language Assessment Quarterly*, 1 (2 & 3), 97–107. DOI:10.1080/15434303.2004.9671778.
- Davies, A., A. Brown, C. Elder, K. Hill, T. Lumley & T. McNamara. 1999. *Dictionary of language testing*. Cambridge: Cambridge University Press.
- EVK 2003. *Eurooppalainen viitekehys: kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys*. Helsinki: WSOY.
- Fulcher, G. 2015. *Re-examining language testing: a philosophical and social inquiry*. London & New York: Routledge.
- Fulcher, G. & F. Davidson. 2010. *Language testing and assessment: an advanced resource book*. London & New York: Routledge.
- Hamp-Lyons, L. 2001. Fourth generation writing assessment. Teoksessa T. Silva & P. K. Matsuda (toim.) *On second language writing*. Mahwah, NJ: Lawrence Erlbaum Associates, 117–127.
- Hamp-Lyons, L. 2000. Social, professional and individual responsibility in language testing. *System* 28 (4), 579–591.
- Hamp-Lyons, L. 1997. Washback, impact and validity: ethical concerns. *Language Testing*, 14 (3), 295–303. DOI: 10.1177/026553229701400306.
- Huhta, A. 2014. Yleiset kielitutkinnot kansainvälisessä testauskentässä. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen. *Yleiset kielitutkinnot 20 vuotta*. Helsinki: Opetushallitus, 159–168.
- Kane, M. 2010. Validity and fairness. *Language Testing*, 27 (2), 177–82.
- Kunnan, A. J. 2004. Test fairness. Teoksessa M. Milanovic & C. Weir (toim.) *European Language Testing in a Global Context*. Cambridge: Cambridge University Press, 27–48.
- Kunnan, A.J. 2000. *Fairness and validation in language assessment*. Cambridge: Cambridge University Press.
- Linnankylä, P. & P. Atjonen. 2008. Arvioinnin eettisyys. Teoksessa E. Korkeakoski & H. Silvennoinen (toim.) *Avaimia koulutuksen arvioinnin kehittämiseen*. Koulutuksen arviointineuvoston julkaisuja 31. Jyväskylä, 47–64.
- Lynch, B. 2001. Rethinking assessment from a critical perspective. *Language Testing*, 18 (4), 351–372. DOI: 10.1177/026553220101800403
- McNamara, T. & C. Roever. 2006. *Language testing: the social dimension*. Malden, MA: Blackwell Publishing.
- McNamara, T. 2000. *Language testing*. Oxford: Oxford University Press.
- Messick, S. 1989a. Meaning and values in test validation: the science and ethics of assessment. *Educational Researcher*, 18 (2), 5–11.
- Messick, S. 1989b. Validity. Teoksessa R. Linn (toim.) *Educational measurement*. New York, NY: Macmillan, 13–103.
- Messick, S. 1981. Evidence and ethics in the evaluation of tests. *Educational Researcher*, 10 (9), 9–20.
- Messick, S. 1980. Test validity and the ethics of assessment. *American Psychologist*, 35 (11), 1012–1027. DOI: 10.1002/j.2333-8504.1981.tb01244.x.
- Salmenkivi, E. 2013. Ylioppilastutkinnon rakenne- ja reaaliokoeuudistusten vaikutuksia: miten lisääntynyt valinnaisuus ohjaa lukiolaisia. *Kasvatus & aika* 3/2013. Saatavilla osoitteessa: http://www.kasvatus-ja-aika.fi/site/?page_id=565. [luettu 13.11.2016].

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

Verkkolähteet

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

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

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

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 44–67.

Minna Martikainen

Jyväskylän yliopisto

Inkeriläisten paluumuuttajien kielitutkinnon synty keskustelussa

This is a study on impact, the meanings given to a language exam when the exam was launched. Ingrians are ethnic Finns who have been living in Russia for many generations. In order to immigrate to Finland on the basis of Finnish origin they have to prove Finnish language skills. The Ingrians' language exam was designed to test if the residence permit applicant is at A2 level in Finnish. This study is based on both oral and written data. In this article I analyze parliamentary protocols and media texts. In addition to these written materials I have interviewed persons who worked on the language exam when it was designed. The topics analyzed construct a continuum between the Finnish identity (past) and moving to Finland (future), both in a text and between texts written over several years. Finally, this article discusses some implications concerning language testing in the future.

Keywords: language exams, Ingrians, Finnish language, discourse analysis

Asiasanat: kielitutkinnot, inkeriläiset, suomen kieli, diskurssianalyysi

1 Johdanto

Kielitutkinto on voimakas keino asettaa kielitaitovaatimuksia ja valvoa vaatimusten täyttymistä, varsinkin kun lainsäätävä ottaa sen työkalukseensa. Tutkintoja pidetään usein objektiivisena totuutena, vaikka ne heijastavat luojiensa arvoja ja normeja (Van Avermaet & Pulinx 2013: 379). Kielitutkinnon seuraukset ovat merkittävä osa tutkinnon laadun ja täten sen oikeutusta (esimerkiksi Bachman & Palmer 1996; ALTE 2011: 17). Vaikka kielitutkintojen vaikutuksista on jo runsaasti tutkimusta, kaikkia vaikutuksia ei voida etukäteen ennakoida (Cheng 2004: 148). Tässä artikkelissa tarkastellaan niitä merkityksiä, joita erään tutkinnon valmisteluvaiheessa annettiin kielitaidon arvioinnille.

Kielitaitoa testataan niin epämuodollisesti opetuksen yhteydessä esimerkiksi luokahuoneessa kuin myös äärimmäisen muodollisilla, viranomaisten järjestämillä tutkinnoilla, joissa tutkintotilaisuus on irrallaan oppimistilanteesta. Tarkastelussa oleva tutkinto, inkeriläisille paluumuuttajille räätälöity suomen kielen tutkinto, on jatkumon jälkimmäisessä päässä: tutkinnon järjesti Suomen viranomainen ja suorituksia arvioivat asiantuntijat, jotka eivät olleet mukana tutkintoon valmistavan koulutuksen järjestämisessä tai opetuksessa. Tutkinnon painoarvoa lisäsi se, että sen suorittaminen oli yhtenä edellytyksenä oleskeluluvan saamiselle inkeriläisstatuksen perusteella. Testaamisen ja arvioinnin oikeudenmukaisuus luo raamit testaamisen hyvälle käytännölle (ALTE 2011: 10, 2016: 18). Kielitutkinnon asettaminen yhden maahanmuuttajaryhmän oleskeluluvan saannin ehdoksi sai osakseen kritiikkiä, mutta toisaalta sitä pidettiin myös perusteltuna ja tervetulleena muutoksena inkeriläisten maahanmuuttoprosessiin. Inkeriläisten maahanmuutossa 1990-luvulta tähän päivään saakka voi nähdä yhtymäkohdita nykyiseen turvapaikanhakijatilanteeseen, kun täysin erilaisista yhteiskunnallisista oloista saapuu ihmisiä Suomeen. Historiallinen ja poliittinen tilanne, olot lähtöalueella ja myös Suomessa ovat kuitenkin erilaisia. Kielitutkinnon käyttöönotosta osana nykyistä maahanmuuttoprosessia on ollut puhetta, mutta sitä ei ole toteutettu. On siis syytä tarkastella 1990-luvun lopulla vallinnutta tilannetta ja arvioida, mikä merkitys kielitutkinnon suhteen tehdyillä valinnoilla on ollut yhden väestöryhmän maahanmuutossa.

Tässä artikkelissa tavoitteena on 1) analysoida niitä merkityksiä, joita inkeriläisten paluumuuttajien kielitutkinnolle annettiin sen syntyvaiheessa ja jälkikäteen. Tavoitteena on myös 2) saada ymmärrystä siitä, millaisia tavoitteita tutkinnolla on haluttu saavuttaa. Kuvaan aluksi sitä historiallista tilannetta, jossa tämä kielitutkinto syntyi ja minkä aseman se on maahanmuuttoprosessissa saanut. Seuraavaksi kerron aineiston keräämisestä ja työtavoistani. Tämän jälkeen esittelen havaintojani aineistosta.

1.1 Dipaki-kielitutkinto

Inkeriläisten paluumuutto Suomeen perustui muuttajien suomalaisen syntyperään, ja suomen kielen taidon testauksella oli siinä osansa. Inkeriläisten paluumuuttajien kielitutkinto lpaki kehitettiin vuonna 2003 todentamaan inkerinsuomalaisten paluumuuttajien suomen kielen taitoa. Tutkinto digitalisoitiin vuonna 2005, ja siitä lähtien tutkinnosta on käytetty nimitystä Dipaki. Tämän kielitutkinnon toteutusta on kuvattu muutamissa artikkeleissa (Kärkkäinen & Lankinen 2008; Tarnanen, Juntila & Westinen 2009). Itse kieliaineistoa, jota tutkinnon kautta kertyi, ei ole analysoitu kielentutkimuksessa lainkaan.

Inkerinsuomalaisten paluumuutto käynnistyi vuonna 1990 varsin nopeasti presidentti Mauno Koiviston 11.4.1990 antaman lausunnon jälkeen, kun hän totesi, että inkeriläisiä voidaan pitää paluumuuttajina (Yleisradio 1990). Lausunnon jälkeen alettiin tehdä virkamiespäätöksiä, joiden pohjalta varsinainen paluumuutto käynnistyi. Paluumuuton säädöksiä muutettiin 1.10.2003, kun voimaan tuli ulkomaalaislain 18 § muutos (218/2003). Henkilön, joka tämän pykälän perusteella haki oleskelulupaa Suomeen, oli osoitettava riittävä suomen tai ruotsin kielen taito. Se, että tässä mainitaan myös ruotsin kieli, liittyy siihen, että suomenkielisten inkeriläisten lisäksi paluumuuttajiin kuuluvat Suomen kansalaiset ja heidän jälkeläisensä. Oleskeluluvan saaminen paluumuuton perusteella oli sikäli suotuisaa, että lupa oli tällöin jatkuva (ensin yhdeksi vuodeksi). Vaikka siihen liittyi myös asuntoedellytys, sen saaminen ei edellyttänyt työ- eikä opiskelupaikkaa. (Maahanmuuttovirasto 2015a.) Tästä syystä oleskelulupaa haettiin tällä ns. inkeriläisstatuksella, vaikka kielitaito olikin edellytyksenä.

Lain mukaan riittävä kielitaito vastasi eurooppalaisen viitekehyyksen mukaisen kielitaidon arviointiasteikon taitotasoa A2. Sitä tarkennettiin Inkerinsuomalaisten paluumuuttajien kielitaidon perusteissa (Työministeriö 2003a) tasoksi A2.1, joka on peruskielitaidon alkuvaihe. Jo joulukuussa 2003 järjestettiin ensimmäinen inkeriläisten kielitutkinto Pietarissa. Tämän jälkeen tutkintoja järjestettiin Pietarin lisäksi Petroskoissa, Moskovassa, Tartossa, Helsingissä ja Jyväskylässä (Maahanmuuttovirasto 2015c).

Suomen eduskunta on päättänyt, että 1.7.2011 jälkeen paluumuuttojonoon ei voi enää ilmoittautua vaan jono on suljettu. Paluumuuttojonoon ennen tätä ajankohdtaa ilmoittautuneilla oli mahdollisuus muuttaa Suomeen 1.7.2016 mennessä itse valitsemanaan ajankohtana. Kielitutkintoja järjestettiin siihen asti. (Maahanmuuttovirasto 2015d.) Tämän ajankohdan jälkeen ei siis enää ole ollut erillistä inkeriläisstatusta oleskelulupaa haettaessa, vaan oleskelulupaa haetaan muilla perusteilla.

Paluumuuttajat, jotka osallistuivat kielitutkintoon, olivat sekä suomen kielen taidoltaan että myös kielelliseltä identiteetiltään erilaisia. Toiset puhuivat suomea (lähes) äidinkielenään, toiset osasivat suomea ainoastaan paluumuuttovalmennuskursseilla tai kielikursseille oppimansa pohjalta. Tutkintoa suoritti 250–700 henkilöä vuo-

deksi. Tutkintoja suoritettiin aikavälillä 1.12.2003–30.6.2016 yhteensä 5476. Tutkinnon läpäisyprosentti oli noin 80. (Maahanmuuttovirasto 2016.)

Tutkinnon järjestämisestä vastasi Maahanmuuttovirasto. Tutkintoja järjestettiin pääasiassa keväällä, kerran kullakin paluumuuttovalmennuksen paikkakunnalla. Tutkintoon osallistujien määrästä riippuen tutkintoja järjestettiin usein kahtena peräkkäisenä päivänä. Tutkinnoissa opastivat ja valvoivat Maahanmuuttoviraston edustajat. Tutkintosuoritusten arviointi tapahtui keskitetysti Suomessa, viimeksi Helsingissä. Maahanmuuttovirasto järjesti myös arvioinnin kutsumalla tehtävään koulutettuja arvioijia eri puolilta Suomea.

Tutkinto on aikanaan kehitetty tilanteessa, jossa inkeriläisten paluumuuton ehtoja haluttiin tiukentaa (tiedote, Sisäministeriö 1999). Tällöin kehitettiin juuri tälle väestöryhmälle räätälöity tutkinto, joka vastasi poliittisesti määriteltyä kielitaidon tasoa. Työryhmässä, joka perustettiin muun muassa tutkinnon perusteiden laadintaa, toteutusta, seuranta ja kehittämistä varten, oli mukana myös kielitaidon arviointiin perehtyneitä kielitutkijoita (Työministeriö 2003b). Päävastuu tutkinnon järjestämisestä ja arvioinnista oli kuitenkin aluksi työministeriöllä, myöhemmin sisäministeriön alaisella Maahanmuuttovirastolla – samalla viranomaisella, joka teki oleskelulupapäätökset osin tutkinnosta saadun arvioinnin perusteella. Tutkinnon järjestämisen aikana kertynyt kielilaineisto on Maahanmuuttoviraston hallussa.

Dipaki-tutkinto asettuu kahden hyvin erityyppisen koulu- ja opiskelukulttuuriin välimaastoon. Tutkintoon valmistauduttiin paluumuuttovalmennuksessa pääsääntöisesti Venäjällä (aiemmin myös Virossa), ja opettajina toimivat paikalliset opettajat. Valmennuksessa pääpaino oli suomen kielen opetuksella. Valmennusta järjestettiin syksystä keväeseen kestäväna iltaopetuksena, pääsääntöisesti kahtena iltana viikossa neljä tuntia kerrallaan (Mielonen 2014, 2015). Valmennuskurssien koordinoijana ja opettajien ohjaajana toimi suomalainen oppilaitos, Etelä-Kymenlaakson ammattiopisto Ekami. Usein kritisoidaan pyrkimystä muuttaa koulutusjärjestelmää järjestämällä kielitutkinto (Alderson & Wall 1993; Cheng 2014: 2483). Dipaki-tutkinnon yhteydessä tutkinnon kehittämisen lisäksi panostettiin voimakkaasti myös opetuksen kehittämiseen (Mielonen 2014).

Tutkinnossa testattiin kielitaitoa neljällä osa-alueella: puhuminen, puheen ymmärtäminen, kirjoittaminen ja tekstin ymmärtäminen. Hyväksytyyn suoritukseen vaadittiin kaikkien neljän osakokeen suorittaminen ja puhumisen sekä jonkin muun osakokeen suorittaminen hyväksytysti. Testin arvioinnissa painotettiin siis aktiivista suullista kielitaitoa. (Työministeriö 2003a; Kärkkäinen & Lankinen 2008.)

Kielitutkinto kehitettiin yhteistyössä työministeriön, Jyväskylän yliopiston soveltavan kielitutkimuksen keskuksen ja Renet Oy Remote -yrityksen kesken. Tutkintoja järjestettiin Virossa, Venäjällä Pietarissa ja Petroskoissa sekä Suomessa. Aluksi puheosuudet toteutettiin haastatteluina, jotka nauhoitettiin C-kaseteille, mutta jo vuodesta

2005 alkaen koko tutkinto digitalisoitiin ja se voitiin suorittaa kokonaan tietokoneella. Dipaki-tutkinto oli offline-tutkinto. Se ei siis ollut saatavilla internetissä vaan oli tallennettuna tietokoneen kovalevyille. Testi suoritettiin valvotuissa testitilanteissa. Järjestelmässä oli eri ohjelmat testin laatijalle, testin suorittajalle ja arvioijalle. (Kärkkäinen & Lankinen 2008.)

Dipaki-testi oli testiin osallistujalle maksuton, mutta henkilö sai osallistua siihen vain kerran (Maahanmuuttovirasto 2015e). Jos hän epäonnistui testissä eikä saanut hyväksytyä tulosta, hän pystyi todistamaan riittävän kielitaidon Yleisen kielitutkinnon avulla, kun sai siitä arvosanaksi vähintään taitotason A2. (Kärkkäinen & Lankinen 2008; Maahanmuuttovirasto 2015e.) Venäjällä asuvalle Yleiseen kielitutkintoon osallistuminen oli kuitenkin käytännössä erittäin haastavaa.

1.2 Dipaki-tutkinnon asema maahanmuuttoprosessissa

Ihmisten liikkuminen ja tilapäinen tai pysyvä oleskelu muualla kuin synnyinmaassaan on lisääntynyt erityisesti 2000-luvun alusta alkaen. Joidenkin arvioiden mukaan lähes kolme prosenttia maailman väestöstä asuu muualla kuin synnyinmaassaan (United Nations 2015: 1–2). Euroopan unionin ja Schengen-alueen myötä puhutaan paljon vapaasta liikkuvuudesta. Toisaalta maahanmuuton edellytyksiä on tiukennettu ja esimerkiksi kielitaitovaatimuksia otettu käyttöön; selvitysten mukaan eritasoiset kielitaitovaatimukset ovat lisääntyneet huomattavasti 2000-luvun ensimmäisen vuosikymmenen aikana sekä maahantulovaiheessa että myös pysyvää oleskelulupaa haettaessa. (Saville 2013; Van Avermaet & Pulinx 2013: 380–381; ALTE 2016: 9–13).

Dipaki-tutkinnon rooli oli varsin merkittävä, kun katsotaan sen asemaa inkeriläisten Suomeen muuttamiseen liittyvissä velvoitteissa. Saville (2009) kuvaa maahanmuuttoprosessia byrokraattisten vaiheiden kautta. Kun maahanmuuttajien kielitaidon testaamista analysoidaan, on syytä tiedostaa, mihin maahanmuuton vaiheeseen testi sijoittuu. Ensimmäinen vaihe on ennen maahanmuuttoa. Dipaki-tutkinto suunniteltiin juuri tähän vaiheeseen, tehtäväksi lähtömaassa Venäjällä ennen Suomeen muuttoa. Aiemmin tutkintoa järjestettiin myös Virossa, mutta Viron liityttyä Euroopan unioniin Viron kansalaiset eivät tarvitse oleskelulupaa muuttaessaan Suomeen. Erityisesti työperäinen maahanmuutto Virosta Suomeen on käytännössä vapaata.

Dipaki-kielitutkintoon täytyi ilmoittautua ja suorittaa se hyväksytysti kahden vuoden kuluessa oleskelulupahakemuksen vireille tulosta (Laki ulkomaalaislain muuttamisesta 218/2003). Tutkintoa edelsi paluumuuttovalmennus, jonka saattoi käydä vain kerran. Valmennuksen päätyttyä hakija kutsuttiin kielitutkintoon. Mikäli henkilö ei osallistunut paluumuuttovalmennukseen, hän ilmoittautui tutkintoon suoraan Suomen edustustossa (Maahanmuuttovirasto 2015b, 2015e).

Tutkinnon asemaa tässä vaiheessa perustellaan käytännöllisillä syillä: muun muassa sillä, että paluumuuttajan kielitaito edistää hänen kotoutumistaan ja työllistymistään Suomessa (esimerkiksi Hallituksen esitys 165/2002). Kielitaidon testaamisessa ennen maahanmuuttoa tai maahanmuuton yhteydessä voidaan nähdä myös piirteitä identiteettikysymyksestä (McNamara 2005). Suomen kielen taidon osoittaminen voidaan nähdä osoituksena suomalaisesta alkuperästä. Inkeriläisten paluumuuttohan käynnistyi presidentti Koiviston lausunnosta, jossa hän totesi inkeriläisten olevan suomalaisia (Yleisradio 1990). Tämän maahanmuuttajaryhmän Suomeen muutolla on siis identiteettiin liittyvä tausta, vaikka maahanmuuton ensisijaiset syyt voivat olla muita. Saksassa vastaavan väestöryhmän, Venäjällä asuvan ja Saksaan muuttavan väestön kohdalla vuonna 1996 käyttöön tulleen kielikokeen tarkoituksena on ollut selvittää, onko saksalainen kulttuuriperintö säilynyt. Kuten Dipakiin, siihenkin voi osallistua vain kerran. (Struck-Soboleva 2006; Työministeriö 1998.)

Inkeriläisten Suomeen-muuton ”byrokraattiset vaiheet” (Saville 2009) jatkuvat hyväksytyin kielitutkinnon jälkeen oleskelulupahakemuksen täyttämällä ja jättämisellä. Jos henkilö sai myönteisen päätöksen, hän pääsi muuttamaan Suomeen. Oleskelulupahakemuksen jättämisestä henkilöllä oli kuusi kuukautta aikaa muuttaa Suomeen. Ensimmäinen oleskelulupa on aina määräaikainen, mutta neljän vuoden jatkuvan oleskelun jälkeen voi saada pysyvän oleskeluluvan. Tämän jälkeen on mahdollista hakea Suomen kansalaisuutta. (Maahanmuuttovirasto 2015a.)

2 Aiemmistä tutkimuksista

Kansainvälisissä toisen kielen arviointia koskevilla tutkimuksilla käytetään termejä *impact* ja *washback* (Wall 1997; sittemmin Cheng 1997; Saville & Hawkey 2004). *Impact*-termillä viitataan niihin (yleisiin) vaikutuksiin, joita testaamisella on yksilöihin tai käytäntöihin luokahuoneessa, koulussa tai laajemmin yhteiskunnassa. *Washback*-termi viittaa siihen, kuinka testaaminen vaikuttaa opetukseen ja oppimiseen. Tässä artikkelissa tavoitteena on analysoida kielitutkinnon valmisteluvaiheessa esiin nousseita tavoitteita, tarkoituksia ja perusteluja eli millä tutkinnon tarpeellisuutta on perusteltu ja mitä tutkinnolla on tavoiteltu. Koska tarkasteltava keskustelu ajoittuu pääosin tutkinnon toteutusta edeltävään aikaan, ei luokahuonevaikutuksia vielä ole, vaikka tutkintoa valmisteltaessa asetetaan odotuksia myös kielen oppimiselle. Artikkelissa haetaan siis tietoa *impact*-vaikutuksesta.

Monet kansainväliset *impact*- ja asennetutkimukset asettuvat kielitutkinnon elinkaareissa ajankohtaan, jossa tutkinto on jo olemassa. Alderson ja Wall (1993) tutkivat kielitutkinnon vaikutuksia siinä vaiheessa, kun englannin kielen tutkinto otettiin käyttöön Sri Lankassa. Tuolloin osoittautui, että tutkinnon käyttöönotto vaikutti opetuk-

sen sisältöön, muttei niinkään opetusmenetelmiin. Myöhemmin muun muassa Cheng (1997, 1998) tutki kielitutkinnon vaikutuksia opetukseen, Saville ja Hawkey (2004) vaikutuksia oppimateriaaleihin ja Watanabe (2004) opettajakohtaisia vaikutuksia. Erityisesti englannin kielen tutkintojen suorituspäämäärät ovat suuria ja henkilöitä, joita tutkinto koskettaa, on paljon. Tällöin on usein päädytty sekä laajoille joukoille suunnattujen kyselyiden ja yksilöllisempien haastattelujen ja havainnoinnin avulla koottuun aineistoon. Aineiston analyysi on usein sisällöllistä (mm. Cheng 1997; Saville & Hawkey 2004; Moore, Stroupe & Mahony 2009; Loumbourdi 2014). Diskurssianalyysitekniikat kielen arviointitutkimuksissa ovat lisääntyneet (Huhta, Kalaja & Pitkänen-Huhta 2006).

Kielitaidon ja -tutkintojen merkitystä on aiemminkin lähestytty niistä käydyn keskustelun kautta. Kun Suomessa tutkittiin ylioppilastutkinnon suorittajien käsityksiä tutkinnosta, päiväkirja-aineistosta esiin nousi rooleja, joita tutkintoon valmistautuvat ottivat puhuessaan tutkinnosta (Huhta ym. 2006). Virtasen (2011) media- ja haastatteleuaineistossa käsitykset kansainvälisten hoitajien ammatillisesta kielitaidosta näyttäytyivät tehtäväkohtaisina, mutta keskustelussa tuli ilmi hyvin vastakkaisia näkökulmia: joissakin puheenvuoroissa korostui ammattisanaston ja termien tuntemisen merkitys, toisissa nähtiin kieli laajana verbaalisen ja myös nonverbaalisena ilmiönä (Virtanen 2011: 167). Myös tässä artikkelissa aihetta lähestytään siitä käydyn keskustelun kautta. Tarkoituksena on tavoittaa niitä kielellisiä ilmaisuja, joita tarkastelussa olevaan kielitutkintoon on liitetty ja tarkastella niitä niissä konteksteissa, joissa ne aikanaan on esitetty.

Kielitaidon keskeisyys identiteetin osana ja rakentajana nousee esille myös Karjalaisen (2012: 195) tutkimuksessa amerikansuomalaisten kieli-identiteeteistä. Iskanuksen tutkimuksessa Suomessa asuville venäjänkielisisille nuorille kysymys kansallisuudesta oli varsin ongelmallinen ja kansallisuudella tarkoitettiin eri konteksteissa eri asioita: yhtäältä suomalaisuus on paluumuuttajille se status, jonka perusteella he ovat tulleet Suomeen, toisaalta se on sitä suomalaisuutta, jonka keskellä he Suomessa elävät (Iskanus 2006: 172, myös Davydova 2002: 156–159).

3 Tutkimusaineisto ja -menetelmät

Tutkimukseni hyödyntää sekä kirjallista että suullista aineistoa. Lainsäädäntötyössä on tallentunut kielitutkinnon valmisteluun liittyvää aineistoa eduskunnan täysistuntojen pöytäkirjoihin, hallituksen esityksiin ja valiokuntien lausuntoihin. Lisäksi olen haastatellut kielitutkintotyöryhmän jäseniä.

Aineistolajit:

1. Työhallinnon julkaisu 215/1998: Inkerinsuomalaisten maahanmuutto Suomeen 1990-luvulla. Hallituksen selonteko eduskunnan ulkoasiainvaliokunnalle. Työministeriö.

2. Eduskunta-aineistoa vuosilta 2002–2010.
3. Media-aineistoa: tiedotteita ja uutisia vuosilta 1999–2003.
4. Työministeriö 2003: Inkeriläisten kielitutkinnon perusteet.
5. Kolmen kielitutkintotyöryhmän jäsenen haastattelut vuonna 2015.

Aineistolajeista lajit 1–4 ovat olemassa jo ennen tämän aiheen valintaa ja aineistonkeruuvaihetta, mutta työryhmän jäsenten haastattelut (laji 5) olen tehnyt tätä tutkimusta varten. Lisäksi lajit 1–4 ajoittuvat tutkinnon syntyvaiheeseen, mutta haastatteluissa ilmiötä tarkastellaan ja sille annetaan merkityksiä jälkikäteen. Nämä seikat täytyy huomioida tuloksia analysoidessa.

Työhallinnon julkaisu vuodelta 1998 on julkaistu ennen varsinaista lain valmistelua, mutta se on toiminut lainsäätäjien taustamateriaalina, kun he ovat perehtyneet asiaan. Tämän vuoksi valitsin sen mukaan aineistoon. Eduskunnan aineistosta hakusana haku ”inkeriläis” tuotti 182 tulosta. Näistä 49 oli kirjallisia kysymyksiä, 34 pöytäkirjoja, muut hallituksen esityksiä, aloitteita ja erilaisia lausuntoja ja mietintöjä, ja vain osa niistä käsitteli kielitaitoasiaa. Asiakirjoissa oli päällekkäisyyttä: sama asiakirja oli joissain tapauksissa tallennettu kahteen kertaan eri nimellä. Joissakin tekstilajeissa, esimerkiksi tilintarkastajien kertomuksissa, ei ole asioita perustelevaa sisältöä, joten rajasin ne tarkastelun ulkopuolelle. Jotkut tekstit toistuivat useissa hakutuloksissa. Toistuneet tekstit karsin myös pois. Lopulliseen tarkasteluun eduskunta-asiakirjoja jäi 11. Ne olivat pääasiassa täysistuntojen pöytäkirjoja, jonka lisäksi kolme kirjallista kysymystä, valiokuntamietintö ja kaksi hallituksen esitystä. Media-aineisto koostuu Helsingin Sanomien ja Yleisradion uutisista ja viranomaisten tiedotteista.

Inkeriläisten kielitutkinnon valmistelua, toteutusta ja seuranta varten perustettiin työministeriön alainen työryhmä ajanjaksolle 1.4.2003–31.3.2004. Työryhmässä oli yhteensä kymmenen jäsentä, joista haastattelin kolmea – niitä joiden tiesin työtehtäviensä puolesta olleen kaikkein eniten tekemisissä kielitutkinnon suunnittelun ja toteutuksen kanssa. Olen itse toiminut tässä kielitutkinnossa tutkintosuoritusten arvioijana alusta alkaen. Tätä kautta osa tutkinnon taustasta ja toimintatavoista oli minulla jo tiedossa.

Aineistoa tarkastelin useiden lukukertojen aikana. Se on myös laajentunut ensimmäisen lukukerran jälkeen. Alkuvaiheessa kävin tekstit ja litteroidut haastattelut läpi ja poimin niistä osiot, jotka liittyvät sekä inkeriläisiin (tai paluumuuttajiin) että heidän kielitaitoonsa, sen puutteeseen tai kielitaidon testaamiseen. Tällä tavalla hahmottui aineiston koko laajuus.

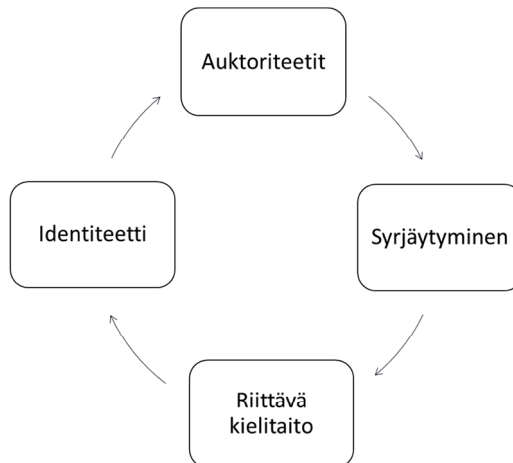
Toisella lukukerralla merkitsin ne kohdat, joissa tarkastelussa olevat aiheet nousivat esille. Säilytin aineiston taustatiedot: mistä aineistolajista on kyse ja milloin se on julkaistu tai syntynyt. Tässä vaiheessa otin käyttöön diskurssianalyysin välineitä hake-

malla vastausta siihen, mitä ilmiöstä sanotaan ja miten, millaisia merkityksiä ilmiö saa (Potter & Wetherell 1987: 160; Pietikäinen & Mäntynen 2009: 11–15).

Tarkasteltavaa keskustelua kielitutkinnosta käydään eri foorumeilla ja eri tilanteissa, ja puheenvuoroja käytetään eri tarkoituksiin. Oletettavaa on, että tämä viestijöiden, tilanteiden ja puhefunktioiden erilaisuus näkyy myös puheen ja tekstin tasolla. Diskursseja ei kuitenkaan tarkastella heijastumina diskurssin osapuolten ajatuksista vaan puheena tai kirjoituksena kielitutkinnoista (Huhta ym. 2006: 331). Tarkastelussa olevat ilmaisut näyttäytyvät historiallisten ja myös sosiaalisten prosessien, menneiden ja käsillä olevien tilanteiden, erilaisten toimijoiden, käytänteiden ja toiminnan risteymänä (Scollon & Wong Scollon 2004). Sosiaalista toimintaa tarkastellaan neksusanalyysin tapaan tutkimuksen kannalta olennaisissa solmukohdissa, neksuksissa, jotka puolestaan muokkaavat tätä toimintaa kuljettaessaan mukanaan aiempia käyttökonteksteja ja merkityksiä (Pietikäinen 2012: 418). Tutkimuksessani tämä tarkoittaa sitä, että tarkastelen kielitutkinnosta käytyä keskustelua niissä konteksteissa, joissa sitä käydään.

4 Neljä inkeriläisdiskurssia

Keskityn tarkastelussani siihen, millaisissa konteksteissa ja millaisilla sananvalinnoilla kielitaitovaatimus esitetään. Tarkastelussa aineistosta nousi aihepiirejä, jotka asettuvat seuraavien neljän laajemman otsikon alle: identiteettdiskurssi, auktoriteettdiskurssi, syrjäytymisdiskurssi ja riittävän kielitaidon diskurssi (kuvio 1).



KUVIO 1. Perustelussa käytetyt diskurssit.

Nämä neljä teemaa näyttäytyvät aineistossa diskurssia luovina ja rakentavina: niillä todetaan ja määritellään. Ne rakentuvat myös menneiden diskurssien varaan ja kommentoivat aiemmin esitettyä. Niissä hyvin usein toistetaan aiempia tekstejä – joko sellaisenaan tai muunnellen. Näin uusi diskurssi rakentuu aiemman päälle. Valtaosa tarkasteltavasta keskustelusta käydään julkisuudessa, mikä osaltaan vaikuttaa kielenkäyttäjien valintoihin.

4.1 Identiteetidiskurssi

Identiteetidiskurssia käydään suomalaisuuden määrittelyn ympärillä. Se liittyy etniseen taustaan ja syntyperään sekä kielen ja identiteetin väliseen suhteeseen. Identiteetidiskurssin voidaan katsoa ensimmäisen kerran tavoittaneen laajan mediayleisön presidentti Koiviston lausunnolla vuonna 1990, kun hän linjasi, että inkeriläiset ovat suomalaisia:

- (1) Haastattelija: Oliko kansalaismielipiteellä vaikutusta siihen, että Suomi on nyt päättänyt ottaa maahan inkeriläisiä?
 Mauno Koivisto: En uskoisi. Tässähän on kysymys verrattain vaatimattomista mittasuhteista ainakin toistaiseksi.
 > Ehkä sillä on ollut joku vaikutus. < (nopeasti)
 Joka tapauksessa kysymys on siitä,
 että nämä ovat suomalaisia
 jotka Ruotsin vallan toimesta on sille alueelle siirretty. (Yleisradio 1990)

Tätä siteerataan mediassa hyvin usein, muun muassa Helsingin Sanomissa vielä vuonna 2002: - - *presidentti Mauno Koivisto totesi inkeriläisten olevan suomalaisia* (13.2.2002). Identiteetin määrittelyä koskeva keskustelu jatkui. Joissain puheenvuoroissa identiteettiin viitataan melko epämääräisesti *yhteys-* ja *side-*sanojen avulla.

- (2) Kielivaatimus ehkä olisi se, jota tässä alun perin on haettu, että heillä on jonkin näköinen **yhteys Suomeen**. (Eduskunta 4/2002, 7.2.2002)
- (3) Jollei syntyperää pystytä asiakirjoilla osoittamaan, tulee selvityksessä ottaa huomioon kattavasti henkilön muu **yhteenkuuluvuutta osoittava side Suomeen**, suomalaisuuteen ja suomen- taikka ruotsinkielen taitoon. (Tiedote, Työvoimaministeriö ja Ulkoministeriö 27.2.2002)
- (4) Ilman riittävää kielitaitoa henkilöllä ei voida katsoa olevan sellaista suomalaista identiteettiä, joka tekisi mahdolliseksi paluumuutosta puhumisen, jos henkilö ei kuuluisi Inkerin siirtoväkeen tai Suomen armeijassa palvelleisiin. (Hallituksen esitys 165/2002, 27.9.2002)

Esimerkeissä 2 ja 3 suomen kieli ja Suomi on yhdistetty toisiinsa: esimerkissä 2 kielitaidon osoittaminen on tapa todentaa yhteys Suomeen, esimerkissä 3 suomen kieli, suomalaisuus ja Suomi (maana) kuvataan tasavertaisina tapoina osoittaa suomalainen syntyperä. Hallituksen esityksessä ulkomaalaislain muutokseksi (esimerkki 4) kielitaito yhdistetään jo varsin eksplisiittisesti identiteettiin: *Ilman riittävää kielitaitoa henkilöllä ei voida katsoa olevan sellaista suomalaista identiteettiä - -*.

Kielen ja identiteetin suhdetta ei nähty kaikissa puheenvuoroissa yhtä selkeinä. Seuraava eduskunnan täysistunnossa esitetty näkökulma (esimerkki 5) tekee eroa kielitaidon ja vahvemman siteen välillä. Siinä tuodaan esille myös kielen opittavuus ja yhteiskuntaan sopeutuminen, joita käsittelem tarkemmin alaluvuissa 3.3 ja 3.4.

- (5) Onko kielitaito juuri se oikea kriteeri, jolla arvioidaan, mikä maahanmuuttajan tausta on? Itse en usko, että tämä on välttämättä kaikkein oikein kriteeri. Moni oppii kielen nopeasti, jos hänellä on kykyä sopeutua yhteiskuntaan, ja sitten voi kysyä myöskin näiden iäkkäämpien henkilöiden osalta, onko perusteltua vaatia kielitaitoa. Heillä voi olla kuitenkin vahvempi side, todellinen side, vaikka kieli on kadonnut. Kieli kyllä katoaa yhden sukupolven aikana, ja silloin ei välttämättä voi syyttää sitä henkilöä, jolta tämä kieli on kadonnut. (Eduskunta 173/2002, 7.1.2003)

Kielitaitovaatimusta linjattiin laissa siten, että se ei koske niitä inkeriläisiä, jotka olivat asuneet Suomessa sota-aikana mutta jotka palautettiin Neuvostoliittoon sotien jälkeen. Eduskunnan täysistunnossa eräs puhuja nostaa esille sen näkökulman, että juuri tämä ryhmä osaa suomea parhaiten, mutta kielitaitoa testataan muilta, nuoremmilta ikäryhmiltä (esimerkki 6). Toinen puhuja samassa täysistunnossa näkee kuitenkin ristiriidan siinä, että kielitaitovaatimus koskee vain yhtä väestöryhmää, inkeriläisiä, mutta ei muita maahanmuuttajaryhmiä (esimerkki 7).

- (6) Kielitaitovaatimus on erittäin tervetullut ja välttämätön. Tietysti pieni paradoksi on se, että nehan, joita se ei siis koske, 43–44 siirtoväkeen kuuluvia inkeriläisiä, osaavat parhaiten suomea, eivät suinkaan nuoremmat, heidän jälkeläisensä, jotka ovat kasvaneet puhtaasti Neuvostoliitossa tai venäläisessä kulttuurissa. (Eduskunta 107/2002, 1.10.2002)
- (7) En pidä kovin hyvänä enkä järkevänä, että **yksiltä vaaditaan erityinen kielitaito** ja sitten meille kuitenkin muuttaa sadoittain toisiin kansallisuuksiin kuuluvia, joilta **ei edellytetä minkäänlaista kotoutumista**. (Eduskunta 107/2002, 1.10.2002)

Esimerkissä 7 liitetään inkeriläiskeskustelu osaksi maahanmuuttokeskustelua yleensä (sadoittain toisiin kansallisuuksiin kuuluvia), mutta myös yhdistetään kielitaito ja kotoutuminen: *yksiltä vaaditaan - - kielitaito, kun taas - - (toisilta) ei edellytetä - kotoutumista*.

Hallituksen selonteossa vuonna 1998 suomalaisuutta määritellään varsin yksityiskohdaisesti sekä taidon (suomen tai ruotsin kielen taito, myös tietoisuus) että toiminnan (vaaliminen, osallistuminen, ylläpitäminen, seuraaminen) kautta:

- (8) Suomalaisen identiteetin tunnusmerkistönä voidaan pitää tietoisuutta suomalaisesta syntyperästä ja sen julkituomista sekä esim. tapojen, perinteiden ja uskonnon vaalimista, osallistumista ulkosuomalaisten kulttuuri- ja järjestötoimintaan, yhteyksien ylläpitämistä Suomeen ja Suomessa asuviin, Suomen olosuhteiden ja tapahtumien seuraamista, suomen tai ruotsin kielen taidon ylläpitämistä ja siirtämistä seuraaville sukupolville jne. (Hallituksen selonteko 9.12.1998, 2–3)
- (9) Suomen kielen taito on olennainen osa suomalaista identiteettiä, mutta sitä ei kuitenkaan voida pitää yksinomaan ratkaisevana. (Hallituksen selonteko 9.12.1998, 3)
- (10) Suomalaiset odottavat, että suomalaisen syntyperän perusteella muuttoluvan saaneet osoittavat yhteenkuuluvuutta Suomeen ja suomalaisuuteen hankkimalla suomen kielen taidon. (Hallituksen selonteko 9.12.1998, 54)

Taitovaatimusta kuitenkin lievennetään seuraavassa virkkeessä: kielitaitoa *ei kuitenkaan voida pitää yksinomaan ratkaisevana* (esimerkki 9). Tätä määrittelyä tehtiin jo vuonna 1998, viisi vuotta ennen kuin kielitaitovaatimus kirjattiin lakiin. Esimerkki 10 on kiinnostava, sillä siinä identiteetin määrittelyä tehdään kahteen suuntaan. Ensimmäin kielitaidon hankkimisella osoitetaan yhteenkuuluvuutta Suomeen ja suomalaisuuteen. Toiseksi sanotaan, että tämä kielitaidon hankkiminen on suomalaisten odotus, siis sen ryhmän, johon kuulumista kielitaidolla pyritään osoittamaan.

Yksi tapa määritellä identiteettiä on vastakkainasettelu: henkilö on joko suomalainen tai jotakin muuta. Esimerkissä 11 venäläiset ja inkeriläiset niputetaan yhteen ja erotetaan heidät suomalaisista. Näissä puheenvuoroissa käytetään usein negaatiota: *ei ole kielitaitoa tai selvää suomalaista identiteettiä* (esimerkki 11), *ei ole enää yhteyksiä* (esimerkki 12), *puuttuu suomen kielen taito ja selvä suomalainen identiteetti* (esimerkki 13).

- (11) Viranomaiset eivät näe juuri eroja maahan muuttavien venäläisten ja inkeriläisten välillä, koska kummallakaan ryhmällä **ei ole kielitaitoa tai selvää suomalaista identiteettiä** (Helsingin Sanomat 13.2.2002).
- (12) Tänä päivänä jonossa on myös ihmisiä, joilla ei itse asiassa **ole enää yhteyksiä eikä perinteitä Suomen puolelle**. (Eduskunta 4/2002, 7.2.2002)
- (13) Joidenkin viranomaisten mukaan suurelta osalta paluumuuttajista **puuttuu suomen kielen taito ja selvä suomalainen identiteetti**, vaikka heiltä edellytetään kieliopin-toja ennen Suomeen muuttoa. (Helsingin Sanomat 31.3.2003)

- (14) Sen takia se ydinkysymys edelleenkin on - -, millä tavalla todella tämä näyttö suomalaisesta syntyperästä nykyistä paremmin voidaan antaa, jotta voidaan niin sanottu rikollinen aines sieltä kytkeä myöskin pois. (Eduskunta 107/2002, 1.10.2002)
- (15) Entisen Neuvostoliiton alueelta pääsee nykyään Suomeen ilman työpaikkaa niin sanottuna paluumuuttajana, jos pystyy asiakirjoilla todistamaan suomalaisen syntyperänsä. Tätä on kyetty käyttämään väärin, ja inkeriläisinä on tullut maahan myös rikollisia, sisäministeri Ville Itälä (kok) perusteli. (Helsingin Sanomat 2.10.2002)
- (16) Kielivaatimuksessa nähtiin kolmekin hyvää puolta: ei-suomalaisperäisiä karsiutuu, "oikeiden" paluumuuttajien 20 vuoden jono lyhenee ja tulijoiden kotoutuminen helpottuu. (Helsingin Sanomat 2.10.2002)

Mediassa vastakkainasettelua tehtiin jopa suomalaisen syntyperän ja rikollisuuden välillä, kuten esimerkiksi 15. Samoja asetteluja tehtiin myös eduskunnan täysistunnossa (esimerkki 14). Tuohon aikaan entisestä Neuvostoliitosta tulleiden rikollisuus oli voimakkaasti esillä tiedotusvälineissä, ja inkeriläisten oleskeluluvan saamiskriteereitä tiukentamalla haettiin ratkaisua tähän ongelmaan, kitkemään rikollinen aines ja estämään se, että suomalaista syntyperää käytetään väärin. Helsingin Sanomat summaa vuoden 2002 kielitutkintokeskustelua esimerkin 16 mukaisesti: suomalaisen identiteetin täyttyminen voidaan todentaa (kielitaito on näin osoitus suomalaisuudesta) ja Suomeen muuttaminen nopeutuu ja helpottuu.

Suomalaisuuden vaatimusta on laajennettu koskemaan myös kielitutkinnon suoritusten arvioijia (esimerkki 17).

- (17) Vuosittain järjestetään testitilaisuuksia Venäjällä ja Virossa. Kielitaidon arvioivat suomalaiset arvioijat. (Inkeriläisten kielitutkinnon perusteet 2003)

Mitä tällä suomalaisuudella tarkoitetaan, jää määrittelemättä.

Sen sijaan myöhemmin samassa asiakirjassa arvioijia määritellään hiukan laajemmin koulutuksen kautta eikä suomalaisuuteen enää viitata (esimerkki 18). Esimerkin 17 suomalaisuuden tulkintaa ohjaa edeltävä virke siitä, että testitilaisuuksia järjestetään Venäjällä ja Virossa. Suomalaisuuden voisi tässä kontekstissa tulkita siis venäläisyyttä ja virolaisuutta vasten.

- (18) Arvioinnin suorittavat tehtävänsä koulutetut arvioijat. Tarvittaessa käytetään kaksoisarviointia. Arvioijakoulutusta järjestetään jatkuvasti ja arviointia seurataan ja kehitetään saadun palautteen perusteella. (Inkeriläisten kielitutkinnon perusteet 2003)

Vuonna 2015 kielitutkintotyöryhmän jäsen nostaa haastattelussa esille aiemman, vuosia sitten olleen identiteetin, sen vahvistamisen ja uuden identiteetin luomisen.

- (19) Haastattelija: Mitä muita merkityksiä tai tarkoituksia siihen (= kielitutkintoon) nyt liittyy?

Haastateltava: Oisko sillä sitten jotakin merkitystä myöskin eräänlaisen uudenlaisen identiteetin luomisessa, että hallitsee tätä kieltä, kun tähän maahan muuttaa. Ehkä se sitten vahvistaa myös sitä identiteettiä, joka joillakin on ollut myös Venäjällä asuessaan. Hehän ovat kokeneet itsensä suomalaisiksi. (työryhmän 1. jäsenen haastattelu 2015)

Identiteettiä tuodaan siis esille monin tavoin. Joissain keskusteluissa suomalaisuus on toimintaa, kuten yhteyksien ylläpitämistä, Suomen tapahtumien seuraamista tai suomalaisen kulttuuriperinnön vaalimista. Suomalaisuutta määritellään myös vastakkainasettelun kautta: henkilö voi olla joko suomalainen (ja inkeriläinen) tai venäläinen, mutta ei molempia. Suomen kielen taito nousee tiheään keskusteluissa esille: suomalaisuus on suomen kielen taitoa, halua oppia ja käyttää suomea.

4.2 Auktoriteetidiskurssi

Auktoriteetidiskurssiksi nimitän sitä diskurssia, jota käydään oikeudesta, vallasta ja valtuutuksesta. Lainsäädäntötekstien luominen noudattaa omia käytäntöjään. En tarkastele tässä lakitekstin syntyä, vaan kuvaan lainsäädäntöön viittaamista sellaisena kuin se aineistossa on. Varsinaisina auktorisoituina toimijoina aineistossa nousevat esille Suomen viranomaisen tai edelleen Suomen viranomaisen hyväksymä toimija, ja lisäksi suomalaiset kouluttajat ja suomalaiset arvioijat.

Jo vuonna 1998 nostettiin esille se, että Neuvostoliiton alueelta Suomeen muuttavat eivät ole aivan samanlainen ryhmä kuin muista maista tulevat paluumuuttajat. Selonteossa todetaan, että tiedotus ja ohjaus eivät ole riittäviä keinoja tämän ryhmän paluumuuton ohjaukseen vaan tarvitaan voimakkaampia viranomaistoimenpiteitä (esimerkit 20 ja 21). Lainsäädännön tiukentaminen oli tämä voimakkaampi keino (esimerkki 22).

- (20) Tiedolla siitä, että Suomessa on korkea työttömyys ja työn saanti siten vaikeaa, ei ollut vaikutusta heidän muuttopäätökseensä. Siirtolaisuuden työntekijöillä eli heidän olosuhteillaan oli suurempi vaikutus muuttopäätökseen. Osoittautui, että **tiedotus ja ohjaus eivät yksin riittäneet heidän paluumuuttonsa ohjauskeinoiksi, vaan tarvittiin voimakkaampia viranomaistoimenpiteitä.** (Hallituksen selonteko 9.12.1998: 5)
- (21) Koska inkerinsuomalaiset historiallisen taustansa ja olosuhteidensa osalta eroavat muista ulkosuomalaisista - - , **on ollut tarpeen käynnistää viranomaistoimia heidän paluumuuttonsa ohjaamiseksi ja sääntelemiseksi** - - . (Hallituksen selonteko 9.12.1998: 52)

- (22) Perustuslakivaliokunnan lausuntoon viitaten hallintovaliokunta ehdottaa **sääntelyn täsmentämistä niin, että** perusteluissa mainittu **vaatimustaso ilmenee laista**. Muutetussa muodossa laissa säädetään myös siitä, että kysymys on **Suomen viranomaisten järjestämästä kielikokeesta**. (Hallintovaliokunnan mietintö 15/2002, 13.12.2002)

Lakitekstiin kielitaitovaatimus kirjattiin seuraavasti:

- (23) Sen lisäksi, mitä tässä laissa muutoin säädetään, myönnetään määräaikainen oleskelulupa entisen Neuvostoliiton alueelta peräisin olevalle henkilölle: - - jos hakija itse, toinen hänen vanhemmistaan tai ainakin kaksi hänen neljästä isovanhemmastaan on tai on ollut merkittynä asiakirjaan kansallisuudeltaan suomalaiseksi ja **hakijalla on riittävä suomen tai ruotsin kielen taito**. (Laki ulkomaalaislain muuttamisesta 218/2003, 18 a §)
- (24) - - hakija osallistuu lähtömaassa järjestettyyn paluumuuttovalmennukseen ja esittää **Suomen viranomaisen järjestämän kielikokeen suorittamisesta todistuksen, jonka perusteella hänellä on sellainen suomen tai ruotsin kielen taito, joka vastaa Euroopan neuvoston yleiseurooppalaisen viitekehityksen mukaisen kielitaidon arviointiasteikon taitotasoa A2** - - . (Laki ulkomaalaislain muuttamisesta 218/2003, 18 a §)

Myös lähtömaassa ennen kielitutkintoa ja Suomeen muuttota tapahtuvalle paluumuuttovalmennukselle määritellään vastuutaho: Suomen viranomaisten järjestämä tai hyväksymä maahanmuuttovalmennus (tiedote, Työvoimaministeriö ja Ulkoministeriö 2002). Määrittely ulottuu kouluttajiin ja kielitutkinnon arvioijiin asti (esimerkki 25 ja aiemmin esimerkit 17 ja 18).

- (25) Valmennuskurssien käytännön organisoinnista huolehtivat suomalaiset kouluttajat. (tiedote, Työvoimaministeriö ja Ulkoministeriö 27.2.2002)

Inkeriläisten kielitutkinnon lisäksi auktoriteettiasemaa laajennetaan Yleisiin kielitutkintoihin (esimerkki 26). Testausjärjestelmä ja sen auktoriteettiasema nousi esille sekä eduskunnan täysistunnon keskustelussa että kielitutkintotyöryhmän 3. jäsenen haastattelussa: *pitää taata, että näillä on suomen tai ruotsin kielen taito, ajatus laadun varmistamisesta* (jolla viitattiin paluumuuttovalmennuksen suomen kielen opetuksen laatuun) ja tarkistus siitä, *onko sitä taitoa riittävästi* (esimerkit 27–28). Virallisella kielitutkinnolla nähdään myös olevan kielen opiskeluun motivoiva merkitys (esimerkki 29).

- (26) Kielitaidon voi todistaa suorittamalla joko Maahanmuuttoviraston järjestämän kielikokeen tai yleisen kielitutkinnon (YKI) suomen tai ruotsin kielessä. (Tiedote, Maahanmuuttovirasto 4.6.2003)

- (27) Nyt tämän lain myötä, jonka tarkoitus on edistää kotoutumista kielitaidon avulla, luodaan tällainen kielitaidon testausjärjestelmä, koska jotenkinhan se **pitää taata, että näillä on suomen tai ruotsin kielen taito.** (Eduskunta 173/2002, 7.1.2003)
- (28) Haastattelija: Mitä ne perustelut tuolloin olivat?
 Miksi tällaista kielitutkintoa tarvittiin?
 Haastateltava: - - käsitys että ei voida olettaa, että nämä muodolliset kriteerit täyttävät hakijat de facto osaisivat suomea riittävästi niin pitäisi sitten jollakin tavalla tarkistaa että he osaa suomea riittävästi voidaksensa (.) tulla toimeen ja integroitua, kotoutua Suomeen. Ja siksi tämä tutkinto sitten tämän tietynlaisen laadun (.) kai sen voi sanoa että se oli tietynlainen **ajatus laadun varmistamisesta, tarkistuksesta että onko sitä taitoa riittävästi.** (työryhmän 3. jäsenen haastattelu 2015)
- (29) Haastateltava: Haluttiin että on tällainen mittari ja ihmisille motivaatio opiskella sitä kieltä että he sitten saavat todistuksen, virallisen arvion siitä kielitaidosta. (työryhmän 2. jäsenen haastattelu 2015)

Auktoriteettidiskurssissa nousi siis esille inkeriläisten erityinen asema paluumuuttajien joukossa. Toisin kuin muut paluumuuttajat, inkeriläisten paluumuuton ohjauskeinoiksi tarvittiin aineistoni mukaan voimakkaampia viranomaistoimia. Näitä toimia olivat lain-säädännön täsmentäminen, kielitutkinnon ja sen toimijoiden määrittäminen ja kielitutkinnon käyttö kielen opetuksen laadun varmistamiseksi. Käytännössä tämä muutos merkitsi kielitaidon arvioinnin asiantuntijoiden vallan kasvua paluumuuttoprosessissa: kielitutkinnon suorittaminen oli yhtenä edellytyksenä oleskeluluvan saamiselle. Esimerkiksi Saksassa paluumuuttoon liittyvän kokeen yhteydessä arviointivaltaa ovat käyttäneet lupaviranomaiset ja kokeessa on ollut piirteitä, joiden perusteella se on enemmän etnisyysskoe kuin kielikoe (Struck-Soboleva 2006: 64). Suomessa keskeiseksi viranomaistoimeksi on otettu kielitutkinto. Tutkinto on pyritty toteuttamaan kansainvälisiä ALTE:n kielitutkinnon laatukriteereitä noudattaen (työryhmän 3. jäsenen haastattelu 2015). Kielitaito, jota voidaan pitää tyypillisesti pehmeänä taitona (*soft skill*), on saanut paluumuuttoprosessissa merkittävän aseman.

4.3 Syrjäytymisdiskurssi

Syrjäytymisdiskurssia tarkasteltaessa on hyvä pitää mielessä tilanne, kun paluumuuton ehtoja 1990–2000-luvun vaihteessa ruvettiin kiristämään. Tuolloin Suomeen oli muutanut jo noin 20 000 inkeriläistä (tiedote, Sisäministeriö 1999), ja heidän tilanteestaan käytiin hyvin vilkasta keskustelua sekä eduskunnassa että mediassa. Inkeriläisten ja

venäläisten ongelmat olivat laajasti esillä mediassa, ja eduskunnalta vaadittiin toimia tilanteen korjaamiseksi.

Muutosta kuvattiin ensin termillä edellytysten tarkistaminen (tiedote, Sisäministeriö 1999), myöhemmin edellytysten kiristyminen, muuttoaallon ja muuton hillitseminen (vuodesta 2002 alkaen). Esimerkissä 30 yhdistetään kielitaito seuraukseksi halusta tulla Suomeen ja kotoutua:

- (30) Kun täällä puhutaan, että kielitaito mukautetaan lähtöalueella annettavaan opetukseen, toivon, että se ei tarkoita sen vaatimuksen vesittämistä vaan sitä, **että niitä menetelmiä parannetaan ja riittävät resurssit siihen varataan ja odotusaikaa hyödynnetään kielen oppimiseen, koska samallahan selviää se todellinen halu tulla Suomeen ja kotoutua tänne, työllistyä täällä, tulla tänne asumaan.** (Eduskunta 107/2002, 1.10.2002)

Eduskunnan täysistunnossa kielitaitovaatimusta argumentoitiin sekä Suomeen kotoutumista edistävänä (esimerkit 31 ja 32) että kohtuuttomana ja turhana (esimerkit 33 ja 34).

- (31) - - ne, jotka hakeutuvat ja pääsevät tänne sen jälkeen, kun edellytyksenä on kielitaito, niin kotoutuminen onnistuu Suomessa paljon paremmin. (Eduskunta 100/2002, 19.9.2002)
- (32) Paluumuuttajilta edellytettävän kielitaidon toivotaan edistävän paluumuuttajien kotoutumista ja työllistymistä Suomessa. (Eduskunta 107/2002, 1.10.2002)
- (33) Minusta on kohtuutonta vaatia näiltä uusilta nuorilta paluumuuttajilta, jotka ovat hyvin koulutettuja, kielivaatimusta siinä mielessä, että "työ tekijäänsä neuvoo" ja siinä se kielikin opiskellaan samalla työn ohessa. (Eduskunta 107/2002, 1.10.2002)
- (34) Kun henkilöllä on työpaikka täällä, hän on myöskin varsin motivoitunut varmasti suomen kieltä opiskelemaan. (Eduskunta 107/2002, 1.10.2002)
- (35) Paluumuuttajan kielitaito edistää hänen kotoutumistaan ja työllistymistään Suomessa. (Hallituksen esitys 165/2002, 27.9.2002)

Esimerkki 32 on myös osoitus siitä, kuinka diskurssit rakentuvat toinen toistensa päälle: hallituksen esitys ulkomaalaislain muutoksesta julkaistiin 27.9.2002 (esimerkki 35), ja siinä kielitaito yhdistettiin kotoutumiseen ja työllistymiseen. Kun asiasta käytiin keskustelua eduskunnassa 1.10.2002, asiaa esittelevä ministeri käyttää samoja sanoja: *kielitaito, kotoutuminen, työllistyminen*. Tätä kielitaidon, kotoutumisen ja työllistymisen yhdistelmää toistettiin vielä vuonna 2010, kun ulkomaalaislakia taas muutettiin tämän inkeriläispykälän osalta. Ilmaukset toistuivat edelleen samoina (esimerkki 36).

- (36) Kielitaitovaatimuksen lisäämistä lakiin perusteltiin sillä, että paluumuuttajan kielitaito edistää hänen kotoutumistaan ja työllistymistään Suomessa. (Hallituksen esitys 252/2010, 12.11.2010)

Usein tähän yhdistettiin myös kielitaidon puute ja syrjäytyminen:

- (37) - - kielitaidon puute aiheuttaa juurettomuutta, syrjäytymistä työelämästä ja vaikeuttaa koulunkäyntiä, mikä puolestaan helposti johtaa yhteiskunnan normien vastaiseen käyttäytymiseen. (tiedote, Sisäministeriö 24.5.1999)
- (38) Runsaan 10 vuoden aikana Suomeen muuttaneista 25 000 inkeriläisestä suuri osa on syrjäytynyt juuri kielitaidon puuttumisen takia. (Yleisradio 10.12.2003)

Keskustelua kielitaidon vaikutuksesta Suomeen muuton yhteydessä käytiin siis sekä puolesta että vastaan. Johdonmukaisesti toistui kielitaidon yhdistyminen kotoutumiseen ja työllistymiseen. Kaikkein voimakkaimpia väitteitä esitettiin negaation kautta: kielitaidon puute yhdistettiin juurettomuuteen, syrjäytymiseen työelämästä ja normien vastaiseen käyttäytymiseen (esimerkit 37 ja 38, aiemmin myös esimerkit 11–13). Suomen kielen taito esitetään niin kotoutumisen edellytyksenä kuin sen seurauksena.

4.4 Riittävän kielitaidon diskurssi

Kielitutkinnon valmistelun yhteydessä on määriteltävä, mikä on riittävä kielitaidon taso tutkinnon hyväksytyyn suorittamiseen. Inkeriläisten kielitutkinto on ns. läpäisy-tutkinto, jonka hyväksytyllä suorituksella henkilö saa kielitaitoarvion A2. Eduskunta- ja media-aineistosta löytyy melko vähän vihjeitä siihen, miten inkeriläisten kielitutkinnossa päädyttiin taitotasoon A2. Mediassa ja myös eduskunnan täysistunnoissa käyty keskustelu kielitaidosta on usein hyvin mustavalkoista, jolloin kielitaitoa joko on tai ei ole: *Työllistyminen edellyttää suomen kielen taitoa* (Hallituksen selonteko 1998). Siihen mikä tuo taitotaso on, ei oteta kantaa. Esimerkissä 39 kielitaitoa määritellään vuonna 1999 ilmauksella *tulee toimeen*:

- (39) Sisäasiainministeri - - pitää tärkeänä, että inkeriläisten maahanmuuttajien maahanpääsyn edellytyksiä tarkistetaan niin, että tulijoilla on Suomen kulttuurin tuntemus ja **sellainen kielitaito, että sillä tulee toimeen** suomalaisessa yhteiskunnassa. (tiedote, Sisäministeriö 24.5.1999)

Mediakeskustelussa tutkinnon taitotasoa määrittelevät ilmaukset vaihtelivat. Helsingin Sanomissa puhutaan *kakkostasosta, selviämisestä arkielämän tilanteissa, välttävästä kielitaidosta* (esimerkit 40–41). Eduskunta-aineistossa tason määrittelyssä taas turvaudu-

taan hallituksen esityksessä käytettyihin sanamuotoihin (esimerkki 42), joita toistetaan täysistunnoissa (esimerkki 43).

- (40) Sisäministeriö valmistelee erillislakia, jonka mukaan maahan saapuvilta inkeriläisiltä vaadittaisiin yleisen kielisuosituksen **kakkostasoa**. Tuo taso edellyttää **selviämistä suomen kielellä arkielämän tilanteissa**. Taso on **vaativampi kuin auttava suomen kielen taito**. (Helsingin Sanomat 28.2.2002)
- (41) Maahan saapuvilta inkeriläisiltä edellytetään **välttävää suomen kielen taitoa**. (Helsingin Sanomat 19.4.2002)
- (42) - - hakijalta edellytettäisiin sellaista suomen tai ruotsin kielen taitoa, joka vastaa **Euroopan neuvoston yleiseurooppalaisessa viitekehysasiakirjassa esitetyn 6 -portaisen kielitaidon arviointiasteikon taitotasoa A2** (niin sanottu **selviytymistaso**).
- - Arvioitaessa henkilön kielitaitoa yleiseurooppalaisessa viitekehyksessä sovitulla tavalla taitotasoa A2 vastaavaksi, henkilön tulee pystyä **selviytymään jokapäiväisissä puhetilanteissa ja kirjoittamaan yksinkertaisia tekstejä ja viestejä**. Taitotason mukaisesti henkilön tulee pystyä **ymmärtämään selkeää, jokapäiväisiä asioita käsittelevää puhetta tai yksinkertaista kirjoitettua tekstiä**. (Hallituksen esitys 165/2002, 27.9.2002)
- (43) Paluumuuttajilta edellytettäisiin sellaista kielitaitoa, että henkilö pystyisi **ymmärtämään selkeää, jokapäiväisiä asioita käsittelevää puhetta tai yksinkertaista kirjoitettua tekstiä**. Kyse on niin sanotusta selviytymistasosta. (Eduskunta 107/2002, 1.10.2002)
- (44) Inkeriläisten kielikoetta kiiteltiin eduskunnassa. Syntyperän todisteluun perättiin lisää pitävyyttä (Helsingin Sanomat 2.10.2002)

Helsingin Sanomissa yhtenä uutisen aiheena oli kielitutkintoasian käsittely eduskunnassa. Otsikko oli esimerkin 44 mukainen. Tekstissä referoitiin värikästä keskustelua. Kielitaidon tasoa ei kuitenkaan käsitelty lainkaan – ei eduskunnassa eikä myöskään ai-hetta käsittelevässä lehti uutisessa. Eduskuntakeskustelussa käsiteltiin tutkinnon järjes-telyjen luotettavuutta, mutta kielitaito nähtiin jonkinlaisena periaatteellisena kynnyk-senä. Osoitus siitä, että osaa jonkin verran suomea, on todiste motivaatiosta ja halusta kotoutua Suomeen. Kielitaidon tasolle ei puheessa asetettu vaatimuksia.

Kaikissa tekemissäni haastatteluissa nousi esille kielitaidon kehittymisnäkökulma, jonka mukaan on tärkeä päästä vaikuttamaan siihen, että Suomeen tulijat opiskele-vat tai muuten osaavat suomea jo ennen Suomeen muuttoa. Paluumuuttovalmennus-ta on järjestetty Suomen lähialueilla pitkään, jo vuosia ennen kielitutkinnon ottamista käyttöön. Haastatteluaineiston esimerkissä 45 tulee esille *suomen kielen opetuksen ryh-distäytyminen*, esimerkissä 46 *auttava kielitaito, jolla voi selviytyä Suomessa*. Huomiota kiinnitetään siis sekä kielen opetuksen laatuun (kuten myös aiemmassa esimerkissä 28) että oppimisen lopputulokseen eli kielitaitoon.

- (45) Haastateltava: No ainakin mitä nyt sitten tavoitettiin niin oli se että tämä (.) Venäjällä ja (.) pääasiassa Venäjällä annettavaa valmennus- valmennuskoulutus niin se kyllä on erittäin paljon ryhdistäytynyt - - Eli kyllä se on tehostanut kovasti paluumuuttovalmennusta, kielenopetuksen ja varmaan muu- muiltakin osin. (työryhmän 1. jäsenen haastattelu 2015)
- (46) Haastateltava: Joo, se tietysti oli kaikkien keskeisin oli se auttava puhuminen. - - Mut se oli se tärkein, haluttiin, että ihmisillä on auttava kielitaito, selviytymiskielitaito täällä Suomessa. (työryhmän 2. jäsenen haastattelu 2015)

Eräs haastateltava kertoi, että kielitutkintotyöryhmän jäsenet olivat tietoisia kansainvälisen kielitaidon testaamista koskevasta keskustelusta. Tuolloin harkittiin myös tasoa B1, mutta oltiin lopulta varsin yksimielisiä siitä, että A2.1-taso oli järkevä ja perusteltu (työryhmän 3. jäsenen haastattelu 2015). A2-taso vaikuttaa kohtuulliselta myös suhteessa sitä edeltävään kielikoulutukseen: inkeriläisten paluumuuttovalmennus kesti yhden lukuvuoden, pääsääntöisesti kahdeksan tuntia viikossa ja lisäksi joitakin intensiiviviikonloppuja. Vaikka kielenoppija olisi valmennuksen alussa aivan alkeistason suomenoppija, lukuvuoden aikana taitotason A2.1 saavuttaminen voi olla mahdollista.

Tämän aineiston perusteella inkeriläisten kielitutkinnon tasoon liittyviä määritelmät ovat *auttava*, *tydyttävä*, *jokapäiväisessä elämässä riittävä*, *selviytymistaso*. Taitotason perustelu jäi aineistossa kuitenkin varsin vähäiseksi.

5 Pohdinta

Olen tässä artikkelissa tarkastellut niitä merkityksiä, joita inkeriläisten paluumuuttajien kielitutkinnoille on annettu. Diskurssianalyysi osoittautui sopivaksi menetelmäksi: se salli aineiston kartuttamisen tutkimustyön edetessä ja sen avulla aineistosta nousi esille tutkimusaiheen kannalta merkittäviä ilmiöitä. Tarkasteltava ajanjakso aiheutti omat haasteensa aineiston tutkimiselle, sillä pääosa kirjallisesta aineistosta on peräisin vuosilta 2002–2003, kun taas haastattelut on tehty vuonna 2015. Tässä ajassa maahanmuuttotilanne ja maahanmuuttoa koskeva keskustelu ovat muuttuneet voimakkaasti. Jää kuitenkin myöhemmin arvioitavaksi, millä tavoin 2000-luvun alussa vallinnutta tilannetta ja kielitutkinnon käyttöönottoa voidaan verrata nykytilanteeseen ja siihen, mitä vaikutuksia kielitaidon nykyistä laajemmalla testaamisella olisi.

Aineiston analyysi osoittaa, että julkisuudessa käyty inkeriläisten kielitutkinnoon liittyvää keskustelua 2000-luvun alussa ammensi perusteluita tutkinnon tarpeellisuudelle monista lähteistä ja perustelut olivat sekä menneeseen että tulevaisuuteen

pohjautuvia. Tutkintoa valmisteltaessa käytiin keskustelua tutkinnon vaikutuksista kohderyhmän valmiuksiin, kielenopetukseen ja laajemminkin ympäröivään yhteiskuntaan, kuten ALTE:n ohjeessa päättäjille suositellaan (ALTE 2008: 2, 2016: 21). Vaikutukset, joita tavoiteltiin, olivat kuitenkin osin keskenään ristiriitaisia. Ristiriitaisuutta voi ymmärtää tarkastelemalla eri diskursseja, joita aiheesta käytiin.

Aiheen käsittelyä voi hahmotella neljän diskurssin avulla. Identiteettidiskurssi lähti liikkeelle presidentti Koiviston lausunnosta vuonna 1990 ja se toistui eduskunnassa, kun kielitutkinnosta keskusteltiin. Auktoriteettidiskurssi ammensi perustelunsa lakiteksteistä, ja yhdeksi teemaksi nousi inkeriläisten erityisasema muiden paluumuuttajien joukossa. Tarvittiin voimakkaampia viranomaistoimia kuin mitä muiden paluumuuttajaryhmien kohdalla on käytetty. Noita toimia olivat lainsäädännön tarkentaminen, kielitutkinnon ja sen toimijoiden määrittäminen ja kielitutkinnon käyttö opetuksen laadun varmistamiseksi. Kielitaidon osoittaminen virallisessa kielitutkinnossa sai merkittävän painoarvon paluumuuton hallinnollisessa prosessissa. Suomeen muuttamista koskeva keskustelu värittiin pääosin syrjäytymiskeskusteluksi, kuten luvussa 3.3 esitetyt esimerkit osoittavat. Syrjäytyminen yhdistettiin kielitaidon puuttumiseen. Kielitutkinnon avulla haluttiin testata sekä kielitaitoa että motivaatiota kielen opiskeluun. Julki-suudessa ei perusteltu kielitutkinnon taitotasovaatimusta oikeastaan lainkaan. Kielitaito nähtiin sellaisena ominaisuutena, että henkilöllä joko on tai ei ole suomen kielen taitoa. Aiemmissa tutkimuksissa (Alderson ja Wall 1993; Cheng 1998) on havaittu, että kielitutkinnoilla on usein vain vähän vaikutusta opetusmenetelmiin, ennemminkin opetuksen sisältöön. Dipaki-tutkinnon arvioinnissa puhumisen osakoikkeella oli keskeinen merkitys (Työministeriö 2003a; Kärkkäinen & Lankinen 2008). Tämä oli selkeä viesti myös siitä, mitä toivottiin opetettavan paluumuuttovalmennuksessa. Kyse ei siis ollut yksinomaan opetuksen sisällöstä, vaan tavoiteltiin myös opetusmenetelmällisiä muutoksia.

Edellä kuvatut neljä diskurssia (identiteettidiskurssi, auktoriteettidiskurssi, syrjäytymisdiskurssi ja riittävän kielitaidon diskurssi) muodostavat ajallisen jatkumon sekä yhden tekstin sisällä että eri vuosina syntyneiden tekstien välille. Tekstit aloitetaan identiteetistä, menneeseen suuntaavasta, ja teksteissä päädytään tulevaisuuteen suuntaavaan diskurssiin, jossa käsitellään Suomeen muuttoa, Suomessa pärjäämistä ja siihen riittävää kielitaitoa. Jatkumo hahmottuu myös tekstien välillä: tarkastelujakson alkuvaiheessa painottuu identiteettidiskurssi, loppua kohden Suomeen muutto ja riittävän kielitaidon diskurssi.

Dipaki-kielitutkinnon aikakausi on päättynyt, mutta keskustelu kielitaidosta, sen osoittamisesta ja arvioimisesta jatkuu vilkkaana maahanmuuttajien määrän noustua turvapaikanhakijoiden myötä. Saville (2013: 4) kuvasi kehityksen suuntaa vuonna 2013 siten, että siinä missä aiemmin perustason kielitaitovaatimuksia esittivät yliopistot ja

työnantajat, nykyään roolin ovat ottaneet viranomaiset viisumi- ja oleskelulupahakemusten yhteydessä. Tämä on ollut tilanne myös inkeriläisten kohdalla.

Kielitaidon arviointi on aikanaan saanut keskeisen roolin inkeriläisten oleskelulupaprosessissa. Kielitaidon arvioinnin asiantuntijat kehittivät tämän pienen mutta vaikutuksiltaan merkittävän tutkinnon täysin alusta alkaen varsin lyhyessä ajassa. Tässä artikkelissa tutkinnon vaikutuksia on tarkasteltu vain pieneltä osin. Jatkossa täytyy analysoida, ovatko tutkinnolle asetetut tavoitteet toteutuneet ja miten tutkinto on vaikuttanut niihin tahoihin, jotka sen kanssa ovat olleet tekemisissä.

Kirjallisuus

- Alderson, J. C. & D. Wall 1993. Examining washback: the Sri Lankan impact study. *Language Testing*, 10 (1), 41–69.
- ALTE 2008 = *Language tests for social cohesion and citizenship – an outline for policymakers* [online]. Strasbourg: Language Policy Division. Council of Europe [luettu 17.6.2016]. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?docu mentId=09000016802fc1b9>.
- ALTE 2011 = *Manual for language test development and examining 2011* [online]. Strasbourg: Language Policy Division, Council of Europe [luettu 17.6.2016]. http://www.coe.int/t/dg4/linguistic/ManualLanguageTest-Alte2011_EN.pdf.
- ALTE 2016 = *Kielitestien avulla osalliseksi, integroituneeksi ja kansalaiseksi*. Opas päätöksenteon tueksi. Cambridge: Association of Language Testers in Europe.
- Bachman L. & A. Palmer 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Cheng, L. 1997. How does washback influence teaching? Implications for Hong Kong. *Language and Education*, 11 (1), 38–54.
- Cheng L. 1998. Impact of a public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24 (3), 279–301.
- Cheng L. 2004. The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. Teoksessa C. Liying, Y. Watanabe & A. Curtis (toim.) *Washback in Language Testing*. Research Contexts and Methods. Mahwah, New Jersey: Lawrence Erlbaum, 147–170.
- Cheng L. 2014. Consequences, impact and washback. Teoksessa A. J. Kunnan (toim.) *The Companion to Language Assessment*. New York: John Wiley & Sons, Inc., 2479–2494.
- Davydova O. 2002. Interaktiivista identiteettipeliä – havaintoja paluumuutosta Suomeen. Teoksessa S. Laihiala-Kankainen, S. Pietikäinen & H. Dufva (toim.) *Moniääninen suomi. Kieli, kulttuuri ja identiteetti*. Jyväskylän yliopisto. Soveltavan kielentutkimuksen keskus, 154–168.
- Huhta, A., P. Kalaja & A. Pitkänen-Huhta 2006. Discursive construction of a high-stakes test: the many faces of a test-taker. *Language Testing*, 23 (3), 326–350.
- Iskanius S. 2006. *Venäjänkielisten maahanmuuttajaopiskelijoiden kieli-identiteetti*. Jyväskylä studies in humanities 51. Jyväskylä: Jyväskylän yliopisto.
- Karjalainen A. 2012. *Liikkuva ja muuttuva suomi. Diskursiivis-etnografinen tutkimus amerikansuomalaisen kielielämäkertoista*. Jyväskylä studies in humanities 186. Jyväskylä: Jyväskylän yliopisto.

- Kärkkäinen K. & J. Lankinen 2008. *Computer-delivered test of Finnish for Ingerian returnees* [online]. The 5th Annual EALTA Conference 8th–11th May 2008. Hellenic American Union, Athens, Greece [luettu 17.6.2016]. http://www.ealta.eu.org/conference/2008/docs/posters/Dipaki_handout.pdf.
- Loumbourdi L. 2014. *The power and impact of standardized tests: investigating washback on language exams in Greece*. Frankfurt/Main: Peter Lang.
- Maahanmuuttovirasto 2015a. *Oleskelulupalomakkeet* [online]. Maahanmuuttovirasto [luettu 19.6.2016]. http://www.migri.fi/asiointi/lomakkeet/oleskelulupalomakkeet#ole_en_fi.
- Maahanmuuttovirasto 2015b. *Paluumuuttovalmennus* [online]. Maahanmuuttovirasto [luettu 19.6.2016]. <http://www.migri.fi/paluumuuttovalmennus>.
- Maahanmuuttovirasto 2015c. *Raportti inkerinsuomalaisten paluumuuttajien kielitutkinnoista vuosilta 2003–2014*. Julkaisematon raportti.
- Maahanmuuttovirasto 2015d. *Viimeiset Maahanmuuttoviraston järjestämät kielitutkinnot*. [online]. Maahanmuuttovirasto [luettu 17.6.2016]. http://www.migri.fi/paluumuutto/paluumuuttaja_entisen_neuvostoliiton_alueelta/hakemuksen_jattamisen_edellytykset/kansallisuudeltaan-suomalainen/kielitutkinto/ilmoittautuminen.
- Maahanmuuttovirasto 2015e. *Kielitutkinto paluumuuttajille* [online]. Maahanmuuttovirasto [luettu 17.6.2016]. http://www.migri.fi/paluumuutto/paluumuuttaja_entisen_neuvostoliiton_alueelta/hakemuksen_jattamisen_edellytykset/kansallisuudeltaan-suomalainen/kielitutkinto.
- Maahanmuuttovirasto 2016. *Raportti inkerinsuomalaisten paluumuuttajien kielitutkinnoista vuosilta 2005–2016*. Julkaisematon raportti.
- McNamara T. 2005. 21st century shibboleth. Language tests, identity and intergroup conflict. *Language Policy*, 4 (4), 351–370.
- Mielonen R. 2014. *Inkerinsuomalaisten paluumuuttovalmennuskoulutus. Lukuvuosi 2013–2014 Pietarissa, Hatsinassa ja Petroskoissa*. Loppuraportti. Etelä-Kymenlaakson ammattiopisto. Julkaisematon raportti.
- Mielonen R. 2015. *Paluumuuttovalmennuksen toteutuksesta*. Haastattelu 15.7.2015.
- Moore S., R. Stroupe & P. Mahony 2009. Perceptions of IELTS in Cambodia: a case study of test impact in a small developing country. Teoksessa J. Osbourne (toim.) *IELTS report 6*, vol 13. Melbourne/Manchester: IELTS Australia and British Council, 293–401.
- Pietikäinen S. 2012. Kieli-ideologiat arjessa. Neksusanalyysi monikielisen inarinsaamenpuhujan kielielämäkerrassa. *Virittäjä*, 3, 410–442.
- Pietikäinen S. & A. Mäntynen 2009. *Kurssi kohti diskursssia*. Tampere: Vastapaino.
- Potter J. & M. Wetherell 1987. *Discourse and social psychology. Beyond attitudes and behaviour*. London: Sage.
- Saville N. 2009. Language assessment in the management of international migration: a framework for considering the issues. *Language Assessing Quarterly*, 6 (1), 17–29.
- Saville N. 2013. Language Testing and Immigration. Teoksessa C. A. Chapelle (toim.) *The Encyclopedia of Applied Linguistics*. Hoboken, New Jersey: Wiley.
- Saville N. & R. Hawkey 2004. The IELTS impact study: investigating washback on teaching materials. Teoksessa C. Liying, Y. Watanabe & A. Curtis (toim.) *Washback in language testing. Research contexts and methods*. Mahwah, New Jersey: Lawrence Erlbaum, 73–96.
- Scollon R. & S. Wong Scollon 2004. *Nexus analysis. Discourse and the emerging internet*. London: Routledge.
- Struck-Soboleva J. 2006. Controversies surrounding language policy and the integration process of Russian Germans in Germany. *Language and Intercultural Communication*, 6 (1), 57–96.
- Tarnanen M., J. Junttila & E. Westinen 2009. *Suunnitelma Yleisten kielitutkintojen keskitason testin digitalisoinnin toteuttamiseksi* [online]. Manner-Suomen ESR-hanke. Soveltavan

- kielentutkimuksen keskus. Jyväskylän yliopisto [luettu 17.6.2016]. Saatavissa: http://www.oph.fi/download/113827_KOKO_SUUNNITELMA_2402.pdf.
- Työministeriö 2003a. *Inkerinsuomalaisten paluumuuttajien kielitestin perusteet*. Helsinki: Työministeriö.
- Työministeriö 2003b. *Ulkomaalaislain 18 a §:n toimeenpanoa koordinoivan työryhmän asettaminen*. Asiakirja 1.4.2003. 635/021/2003TM. Helsinki: Työministeriö.
- United Nations 2015. *International migration report 2013*. Department of Economic and Social Affairs. Population Division. December 2013.
- Van Avermaet P. & R. Pulinx 2013. Language Testing for Immigration to Europe. *The Companion to Language Assessment*. New York: John Wiley & Sons, Inc., 376–389.
- Virtanen A. 2011. Käsitteitä kansainvälisesti rekrytoitujen hoitajien ammatillisesta kielitaidosta ja sen kehittymisestä mediakeskusteluiden ja asiantuntijan haastattelun analyysia. *Puhe ja kieli*, 31 (4), 153–172.
- Wall D. 1997. Impact and washback in language testing. Teoksessa C. Clapham & D. Corson (toim.) *Encyclopedia of language and education*. Dordrecht: Kluwer Academic, 291–302.
- Watanabe Y. 2004. Teacher factors mediating washback. Teoksessa C. Liying, Y. Watanabe & A. Curtis (toim.) *Washback in language testing. Research contexts and methods*. Mahwah, New Jersey: Lawrence Erlbaum, 129–146.

Aineistolähteet

- Täysistunnon pöytäkirjoja vuodelta 2002. Eduskunta.
- Hallintovaliokunnan mietintö 2002. Hallituksen esitys laiksi ulkomaalaislain 18 a §:n muuttamisesta [online]. Helsinki: Eduskunta [luettu 17.6.2016]. https://www.eduskunta.fi/FI/vaski/mietinto/Documents/havm_15+2002.pdf#search=Hallintovaliokunnan%20mietint%C3%B6%2015%2F2002.
- Hallituksen esitys 2002. Hallituksen esitys eduskunnalle laiksi ulkomaalaislain 18 § muuttamisesta [online]. Helsinki: Eduskunta [luettu 17.6.2016]. https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Documents/he_160+2002.pdf#search=he%20160%2F2002.
- Hallituksen esitys 2010. Hallituksen esitys eduskunnalle laiksi ulkomaalaislain 48 §:n muuttamisesta [online]. Helsinki: Eduskunta [luettu 20.6.2016]. https://www.eduskunta.fi/FI/vaski/HallituksenEsitys/Documents/he_252+2010.pdf#search=HE%20252%2F2010.
- Helsingin Sanomat. Artikkeleita vuosilta 2002–2003.
- Laki ulkomaalaislain muuttamisesta 218/2003 [online] [luettu 14.6.2016]. <http://www.finlex.fi/fi/laki/alkup/2003/20030218>.
- Tiedotteita. Viranomaistiedotteita vuosilta 1999–2003. Maahanmuuttovirasto, Sisäministeriö, Työvoimaministeriö ja Ulkoministeriö.
- Työministeriö 1998. Inkerinsuomalaisten maahanmuutto Suomeen 1990-luvulla. Hallituksen selonteko 1998. Hallituksen selonteko eduskunnan ulkoasiainvaliokunnalle. Työhallinnon julkaisu 215. Helsinki: Työministeriö.
- Työministeriö 2003a. Inkerinsuomalaisten paluumuuttajien kielitestin perusteet. Helsinki: Työministeriö.
- Työryhmän jäsenten haastattelut vuonna 2015. Työministeriön alainen 1.4.2003–31.3.2004 toiminut työryhmä.
- Yleisradio 1990. Tasavallan presidentti Mauno Koiviston haastattelu. Ajankohtainen kakkonen 10.4.1990.
- Yleisradio 2003. *Inkeriläisten kielitaito testiin* [online]. Uutinen [luettu 20.6.2016]. <http://yle.fi/uutiset/3-5151940>.

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 68–88.

Sari Ahola, Reeta Neittaanmäki & Tuija Hirvelä

Jyväskylän yliopisto

Puheen ymmärtämisen tehtävien taitotasoille asettaminen Yleisissä kielitutkinnoissa

The aim of this study is to explore rater effects, such as consistency and severity of the judgements of 12 expert judges, and to report the extent of assessment criteria used by them. The data of the study are based on the three different phases of the standardization procedure which were analyzed using the MFRM model. In addition, feedback gathered from the expert judges was used to find out their opinions about the procedure and to validate the results of the standardization. The results of this study indicate that general discussion on and statistical information about the items influenced judges' consistency and severity and their adaptation of the assessment criteria. The judges themselves also found that these factors affected their judgements. In addition, the nature of listening and the criteria caused difficulties during the procedure.

Keywords: standard setting, language testing, expert judges, proficiency levels, test items

Asiasanat: taitotasoille asettaminen, kielitaidon arviointi, asiantuntijat, taitotasot, tehtäväosiot

1 Johdanto

Yleisten kielitutkintojen osallistujilla on erilaisia käyttötarkoituksia tutkintotodistukselle. (Ks. lisää tutkinnon käyttötarkoituksesta ja osallistujista: Neittaanmäki & Hirvelä 2014.) Osallistujalle todistus on usein varsin tärkeä dokumentti, sillä sen perusteella hänen kyvyistään viestiä ja ymmärtää kieltä tehdään päätelmiä, joilla saattaa olla hyvinkin kauaskantoisia vaikutuksia hänen tulevaisuudelleen. Toisaalta taas tutkintotodistusta edellyttävien tahojen, kuten viranomaisten, työnantajien ja koulutusta tarjoavien, on myös voitava luottaa siihen, että todistuksessa olevat arviot vastaavat osallistujan taitoa ja että sen perusteella tehty tulkinta osallistujan kyvystä vastaa todellisuutta. Jotta tutkinnon osoittamia tuloksia voidaan pitää luotettavina, edellytetään tutkinnon järjestäjältä toimenpiteitä. On pyrittävä erilaisin keinoin varmistamaan, että arvioinnissa toteutuvat laadukkaan arvioinnin periaatteet, ts. arviointi on objektiivisuuteen pyrkivää, oikeudenmukaista, läpinäkyvää ja luotettavaa sekä eettisesti kestävä. (Ks. lisää validoinnista: Messick 1989, 1994; Weir 2005.)

Määriteltäessä tutkintoihin osallistuvien kielitaitoa arvioinnin tueksi tarvitaan tehtäviä, joiden avulla osallistujat voivat osoittaa kielitaitonsa. Vaikka tutkintojärjestelmässä saadaan jokaisen testikerran jälkeen tilastollisiin analyyseihin pohjautuvaa tietoa tehtävien toimivuudesta ja vaikeustasosta, tarvitaan näiden tietojen tueksi vielä tutkintokielen asiantuntijoiden arvioita tehtävien vaativuustasoista ja testin pisterajoista.

Standard setting eli taitotasolle asettaminen tarkoittaa yleensä testin pisterajojen määrittämistä suhteessa kielitaidon kehittymistä kuvaavaan asteikkoon (Reckase 2009). Käytännössä tällä tarkoitetaan sitä, että kielen asiantuntijat määrittelevät tehtävissä vaadittavan osaamisen tason kielitaidon kehittymistä kuvaavien taitotasokuvainten avulla. Tämä prosessi on tärkeä vaihe arvioinnin ja koko tutkintojärjestelmän validoinnissa, sillä jos kielitaidosta halutaan tehdä luotettavia päätelmiä testin avulla, on varmistettava, että testin sisältö sopii käyttötarkoitukseen ja että testissä käytettävät tehtävät mittaavat oikean tasoista kielitaitoa (Bejar 2008). Bejar (2008) ja Cizek ja Bunch (2007) pitävätkin tärkeänä, että *standard setting* liitetään olennaiseksi osaksi testin kehittämistä ja suunnittelua.

Yleisissä kielitutkinnoissa on viimeisen neljän vuoden aikana järjestetty tilaisuuksia, joissa suomen kielen asiantuntijat ovat asettaneet tekstin ymmärtämisen tekstejä ja niihin liittyviä tehtäväosiota¹ taitotasolle, mutta puheen ymmärtämisen tehtävien standardointiprosessi on tutkintojärjestelmässä vasta aluillaan. Standardointiprosessit näissä ymmärtämisen taidoissa ovat varsin samankaltaiset, koska tekstin

1 Tehtäväosiolla viitataan tekstikohtaan liittyvään yksittäiseen kysymykseen, jolle annetaan erillinen pistemäärä. Tehtävä puolestaan sisältää sekä tekstin että siihen liittyvät osiot.

ja puheen ymmärtäminen ovat myös taitoina monilta osin samanlaisia. Molemmissa taidoissa tarvitaan mm. lingvistisiä taitoja, kontekstin tuntemusta, maailmantietoa ja päättelytaitoa.

Puheen ymmärtämisen testaukseen liittyy kuitenkin piirteitä, jotka on hyvä huomioida asetettaessa tehtäviä taitotasolle. Puheen ymmärtäminen on luonteeltaan reaaliajassa tapahtuvaa, ja se edellyttää osakokeen suorittajalta auditiivisten ja lingvististen taitojen lisäksi semanttisia ja kognitiivisia taitoja sekä hyvää työmuistia. Testiin osallistujan on hallittava nauhalta tulevan tekstin kuunteleminen, kuullun tiedon tunnistaminen ja sen prosessointi sekä tulkinta. Tämän lisäksi on ymmärrettävä kuunneltavaan tekstiin liittyvät tehtäväosiot ja vastattava niihin. Yksittäiset sanastolliset tai syntaktiset vaikeudet kuuntelun aikana saattavat aiheuttaa sen, että testiin osallistuja ymmärtää kuulemastaan vain vähän tai ei mitään, eikä hän välttämättä pysty enää korjaamaan tilannetta, koska puheen ymmärtämiseen ei voi palata samoin kuin kirjoitettuun tekstiin. Lisäksi puheen ymmärtämiseen liittyy myös monia ymmärrettävyyteen ja vaikeustasoon vaikuttavia ulkoisia tekijöitä, kuten äänitteellä kuuluvien puhujien puhenopeus ja artikulaatio. (Lisää puheen ymmärtämisestä: Buck 2001; Rost 1990.)

2 Tutkimuskysymykset ja tutkimuksen tarkoitus

Tutkimuksemme on tapaustutkimus (*case-study*), jonka avulla pyrimme saamaan yksityiskohtaisempaa tietoa eri standardointimenetelmien soveltuvuudesta testijärjestelmään. Tapaustutkimusta käytetäänkin usein pyrittäessä ymmärtämään tutkittavaa asiaa kokonaisvaltaisemmin, jotta tutkimuksen avulla saatuja tuloksia voidaan siirtää käytäntöön. (Tapaustutkimuksesta: Saarela-Kinnunen ja Eskola 2001.) Tutkimuksemme nivoutuukin tiiviisti yhteen Yleisten kielitutkintojen kehittämisen kanssa. Tavoitteena on ennen kaikkea luoda tutkintojärjestelmän omaan testauskontekstiin sopivat käytännöt ja vahvistaa näin entisestään arviointiprosessin luotettavuutta.

Taustoitamme tutkimustamme ensin kuvaamalla artikkelissa lyhyesti ensimmäisen suomen kielen puheen ymmärtämisen standardointiprosessin. Tämän jälkeen analysoimme prosessia sekä kvantitatiivisen että kvalitatiivisen aineiston avulla. Tuloksissa esittelemme, kuinka yhdenmukaisia asiantuntijoiden näkemykset tehtävien taitotasosta olivat ja mitä muutoksia ankaruuseroissa ja arviointilinjan johdonmukaisuudessa havaittiin standardointiprosessin eri vaiheissa. Lisäksi tarkastelemme asiantuntijoilta saatujen kirjallisten kommenttien avulla, mitkä tekijät vaikuttivat taitotason määrittelyyn ja mitä vaikeuksia taitotasolle määrittelyyn liittyi.

Yleisissä kielitutkinnoissa taitotasolle asettamisen tavoitteena on arvioida yksittäisten testitehtäväosioiden vaativuutta sen sijaan, että määritellään tiettyyn taitotasoon vaadittava pistemäärä kokonaiselle osakokeelle. Koska Yleisissä kielitutkinnoissa

tehtävät eivät ole kertaluonteisia, tutkintojärjestelmässä on käytössä osiopankki, joka koostuu hyvin toimivista tehtäväosioista. Jokaiselle tehtäväosiolle on määritelty Raschin mallin avulla vaikeustasoa kuvaava lukuarvo, jonka perusteella tehtäväosiot sijoittuvat vaikeustasoa kuvaavalle skaalalle (Raschin mallista: Wright & Stone 1979; Törmäkangas & Törmäkangas 2009). Yleisten kielitutkintojen standardoinnin tavoitteena onkin etsiä tälle vaikeustasoskaalalle taitotasojen katkaisukohtat, jotta voidaan varmistaa se, että testikerrasta riippumatta testikokonaisuudet ovat vertailukelpoisia ja samanlaatuisia keskenään. Tavoitteen saavuttamisen ehtona on kuitenkin se, että standardointia tekevät päätyvät arvioissaan mahdollisemman yhdenmukaiseen käsitykseen testitehtävien ja -osioiden vaatimasta osaamisesta eli taitotasosta.

3 Tutkimusaineisto

3.1 Aineisto

Puheen ymmärtämisen standardointiprosessiin osallistuneet asiantuntijat asettivat taitotasoaasteikolle 35 suomen keskitason puheen ymmärtämisen tehtäväosiota, jotka edustivat sekä tehtävätyypeiltään että vaikeustasoiltaan kattavasti Yleisten kielitutkintojen osiopankkia. Valitut tehtäväosiot asetettiin taitotasolle standardointiprosessia varten koottujen taitotasokuvainten avulla. Kuvaimet koottiin eri dokumenteista: Eurooppalaisesta viitekehuksesta, Yleisten kielitutkintojen perusteista ja laadintaohjeista. Tällä pyrittiin varmistamaan se, että puheen ymmärtämisen taidosta saadaan mahdollisimman kattava kuva ja että kuvaus vastaa tutkintojärjestelmän dokumenteissa kuvattua taitoa.

Jotta tuloksia voidaan pitää luotettavina, myös prosessiin osallistuvien asiantuntijoiden on täytettävä tietyt vaatimukset. Cizekin ja Bunchin mukaan (2007) asiantuntijat ja heidän ominaisuutensa vaikuttavat standardoinnin tuloksiin vähintään yhtä paljon – tai jopa enemmän – kuin valittu *standard setting* -menetelmä. Asiantuntijoilta edellytetään ensisijaisesti vahvaa asiantuntemusta mitattavasta asiasta, mutta myös muut taustatekijät, kuten sukupuoli, ikä ja alueellisuus, on huomioitava valittaessa taroitukseseen sopivaa asiantuntijaryhmää. Standardoinnin tavoite ja asiantuntijoille asetetut edellytykset määrittelevät myös pitkälti asiantuntijoiden lukumäärän. Riittävänä määränä luotettavien tulosten saamiseksi voidaan kuitenkin pitää vähintään 10 asiantuntijaa (ks. lisää: Kaftandjieva 2010).

Yleisten kielitutkintojen standardointiprosessiin valittiin 12 suomen kielen asiantuntijaa. He olivat eri-ikäisiä, pääosin S2-opettajina eri puolilla Suomea toimivia henkilöitä, jotka ovat toimineet vuosia Yleisissä kielitutkinnoissa joko arvioijina ja/tai tehtävän laatijoina. He tunsivat hyvin tutkintojärjestelmän, suomen kielen tutkinnon

ja taitotasojen asettamisessa käytettävät taitotasokuvaimet sekä heillä oli päivätyönsä kautta kokemusta testiin osallistujien kaltaisista kielenoppijoista. Lisäksi heillä kaikilla oli aikaisempaa kokemusta Yleisten kielitutkintojen tekstin ymmärtämisen tehtäväosien taitotasoilta asettamisesta ja siihen liittyvästä prosessista.

Koska kyseessä on tutkintojärjestelmän kehittämiseen tähtäävä tutkimus, oli tärkeää myös huomioida asiantuntijoiden mielipiteet prosessista. Asiantuntijoilta kerättyä palautetta voidaan käyttää apuna kehitettäessä tutkintojärjestelmään puheen ymmärtämisen standardointikäytänteitä, mutta palautekyselyn vastauksia voidaan hyödyntää myös osana standardointiprosessin tulosten luotettavuuden tarkastelua. Tästä syystä käsittelemme myös artikkelimme loppuosassa asiantuntijoilta saatua palautetta, jonka olemme analysoineet sisällönanalyysimenetelmää käyttäen. (Sisällönanalyysista: Tuomi & Sarajärvi 2009.)

3.2 Standardointiprosessin kuvaus

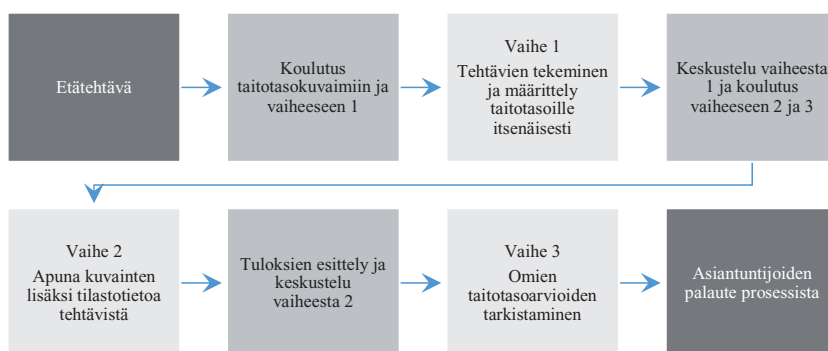
Erilaisia menetelmiä taitotasojen asettamiseksi on kymmeniä. Menetelmän valinta on tehtävä huolellisesti, sillä se vaikuttaa tuloksiin, ja eri menetelmillä on myös erilaiset vaatimukset aineistolle. (ks. menetelmistä: Cizek & Bunch 2007; Kaftandjieva 2010.) Reckasen (2009) mielestä ongelmallista standardoinnissa onkin juuri se, että yhtenäistä *standard setting* -teoriaa ei ole olemassa. Eri menetelmät tuottavat usein toisistaan poikkeavia tuloksia, ja tulokset ovat riippuvaisia monista prosessiin ja sen toteutukseen liittyvistä yksityiskohdista.

Yleisissä kielitutkinnoissa standardoinnissa käytettävän menetelmän valinta perustui standardoinnin tavoitteiden lisäksi olemassa olevaan IRT (*item response theory*) -pohjaiseen osiopankkiin, käytössä oleviin tehtävätyyppeihin ja taitotasosteikkoon sekä aika-, kustannus- ja henkilöresursseihin. Lisäksi puheen ymmärtämisen menetelmä haluttiin pitää mahdollisimman samankaltaisena kuin tekstin ymmärtämisen, jotta voitiin hyödyntää standardointiin osallistuvien asiantuntijoiden aiempaa kokemusta taitotasojen asettamisesta.

Standardoinnin tuloksiin vaikuttavat keskeisesti menetelmästä riippumatta siihen osallistuvat asiantuntijat ja se, kuinka heidät on perehdytetty koko prosessiin (Cizek & Bunch 2007; Council of Europe 2009). Perehdytyksen tavoitteena on, että asiantuntijat ymmärtävät mm. standardoinnin tavoitteet, keskeiset määritelmät ja taitotasokuvaimet ja että he osaavat käyttää valittua menetelmää prosessin eri vaiheissa. Yleisissä kielitutkinnoissa asiantuntijoiden perehdytys taitotasojen asettamiseen alkoi puheen ymmärtämisen 6-portaiseen taitotasokuvaimiin liittyvällä etätehtävällä jo ennen varsinaista standardointitilaisuutta. Etätehtävän avulla saatiin tietoa asiantuntijoiden taitotasokuvaimiin liittyvistä näkemyks- ja tulkintaeroista, joita voitiin hyödyntää varsinaisena standardointipäivänä annettavassa koulutuksessa. Koska asiantuntijat

tunsivat jo entuudestaan tutkintojärjestelmän, perehdytystä tutkintojärjestelmään ei tarvittu, vaan koulutuksessa voitiin keskittyä taitotasokuvainten sisältöihin. Koulutuksen tavoitteena oli, että asiantuntijoille syntyisi mahdollisimman yhdenmukainen kuva eri taitotasoilla vaadittavasta osaamisesta ja että he tulkitsisivat taitotasokuvaimia samankaltaisesti.

Taitotasoille asettaminen on monivaiheinen prosessi, joka koostuu useista eri vaiheista (ks. kuvio 1). Taitotasokuvaimien koulutuksen jälkeen alkoi varsinainen työskentely tehtävien parissa (kuvio 1, vaihe 1), joka aloitettiin suorittamalla tehtävät samoin kuin osallistujat testitilanteessa. Tällä pyrittiin siihen, että asiantuntijat pystyivät ottamaan huomioon tallenteelta kuuluvien puhujien vaikutuksen tehtävien vaikeustason määrittelyssä. Tämä myös auttoi asiantuntijoita asettautumaan testattavan rooliin, ja samalla he tutustuivat tehtäviin. Tehdessään tehtäviä he samanaikaisesti lisäksi arvioivat itsenäisesti ensimmäisen kerran tehtäväosioissa vaadittavan osaamisen tason. Nämä arviot perustuivat asiantuntijoiden omiin käsityksiin ja tulkintoihin taitotasoista.



KUVIO 1. Yleisten kielitutkintojen standardointiprosessin vaiheet.

Standardoitavien tehtäväosioiden toimivuudesta ja vaikeustasosta on tutkintojärjestelmässä saatavilla laajasti tilastotietoa, joka perustuu tehtäväosioista riippuen tuhansien osallistujien vastauksiin (vrt. Yleisten kielitutkintojen osiopankki, s. 72). Tämä tieto haluttiin tarjota päätöksenteon tueksi, ja vaiheessa 2 (kuvio 1, vaihe 2) asiantuntijoita pyydettiinkin huomioimaan taitotason määrittelyssä kuvainten lisäksi osallistujien suorituksiin perustuva tilastotieto tehtäväosioiden vaikeudesta. Asiantuntijoille jaettiin tehtävät uudestaan siten, että tehtäväosiot oli laitettu vihkoon niiden vaikeustason mukaiseen järjestykseen helpoimmasta vaikeimpaan. Asiantuntijoiden tehtävänä oli nyt uudelleen tarkastella tehtäväosioita ja arvioida, missä eri taitotasojen rajat ovat. Käytännössä asiantuntijat päättivät ensin, minkä taitotason osaaja pystyy mahdollisesti ratkaisemaan vihkon ensimmäisen eli helpoimman osion. Tämän jälkeen he etenivät osio osiolta saman taitotason puitteissa niin pitkälle, kunnes vastaan tuli tehtäväosio,

jonka ratkaisemiseksi heidän mielestään tarvittiin jo osaamista seuraavalta taitotasolta. Näin he etenivät viikkon viimeiseen eli vaikeimpaan tehtäväosioon saakka. Koko vaiheen ajan asiantuntijoilla oli mahdollisuus palata kuuntelemaan tehtäviä uudelleen, jos he kokivat sen taitotason määrittelyn kannalta tarpeelliseksi.

Ennen vaiheeseen 3 siirtymistä asiantuntijoille esitettiin vaiheen 2 tulokset. Vaiheesta 2 saatuja tuloksia pidettiin prosessin kannalta merkityksellisinä, koska asiantuntijoilla oli tässä vaiheessa käytössään tehtäväosioista paljon enemmän tietoa kuin vaiheessa 1. Lisäksi pyrittäessä yhdenmukaisuuteen asiantuntijoiden oli hyvä nähdä, miten muut olivat arvioineet ja kuinka kukin itse oli arvioinut tehtäväosion vaativuuden suhteessa muihin asiantuntijoihin. Tulosten esittämistä voidaan myös perustella sillä, että vaiheessa 3 asiantuntijaryhmän on jo päästävä koko prosessin onnistumisen kannalta hyvin samankaltaisiin tulkintoihin osioiden vaikeustasosta. Jotta tämä tavoite saavutetaan, keskityttiin ennen vaihetta 3 käydyssä keskustelussa ennen kaikkea niihin tehtäväosioihin, joissa asiantuntijoiden käsitykset tehtäväosion taitotasosta poikkesivat huomattavasti toisistaan tai tehtäväosiot olivat ratkaisevassa roolissa taitotasorajojen määriteltäessä. Tämän keskustelun pohjalta vaiheessa 3 asiantuntijat saattoivat vielä muuttaa arviointejaan lähemmäksi toisiaan (kuviossa 1 vaihe 3).

Edellä kuvatut vaiheet perustuvat pääosin Bookmark-menetelmään (ks. menetelmästä lisää: Cizek & Bunch 2007), vaikkakin joiltain osin menetelmää on muokattu Yleisten kielitutkintojen tavoitteisiin sopivammaksi. Teknisesti menetelmä ei aiheuttanut asiantuntijoille vaikeuksia, sillä se oli tekstin ymmärtämisen standardointilaisuudesta tuttu, mutta menetelmään sopivan oppijan määrittely oli koulutuksen ja myös asiantuntijoiden näkökulmasta vaikeaa. Koulutuksessa asiantuntijoita ohjeistettiin taitotasolle määriteltäessään (vaiheet 1, 2 ja 3) miettimään oppijaa, joka todennäköisesti juuri ja juuri kykenee vastaamaan tehtäväosioon oikein. Koska tuloksiin vaikuttaa se, miten yhdenmukainen käsitys ja tulkinta asiantuntijoilla on tällaisen kahden taitotason rajalla olevan oppijan osaamisesta, aiheesta käytiin keskusteluja prosessin eri vaiheissa.

Asiantuntijat saivat keskustella myös tehtävistä, taitotasosta ja standardointiprosessiin liittyvistä tunnelmistaan aina jokaisen vaiheen jälkeen ennen seuraavaan vaiheeseen siirtymistä. Avoin ja salliva keskustelu on tärkeä osa standardointiprosessia, mikä edellyttää standardointiprosessin vetäjältä taitoa ohjata keskustelua. Prosessin vetäjän keskeinen tehtävä on luoda asiantuntijoiden välille luottamuksellinen ja kannustava ilmapiiri, jossa voi vapaasti ilmaista omia mielipiteitään. Vetäjän on myös huolehdittava siitä, että ryhmän sisäinen dynamiikka toimii ja että keskustelun fokus säilyy tehtävien taitotasossa ja taitotasokuvaimissa. Keskustelulla tuetaan asiantuntijoiden arviointityötä ja pyritään siihen, että heidän käsityksensä tehtäväosioiden vaatimuksista saadaan mahdollisimman yhdenmukaiseksi.

4 Tulokset

4.1 Tilastolliset tulokset

Standardointiprosessin päätulokset ovat tehtävien osiokohtaiset taitotasot, mutta tässä artikkelissa emme raportoi niitä, vaan keskitymme tarkastelemaan asiantuntijoiden toimintaa, kuten ankaruutta, taitotasoasteikon käytön johdonmukaisuutta ja laajuutta sekä näihin liittyviä muutoksia prosessin eri vaiheissa.

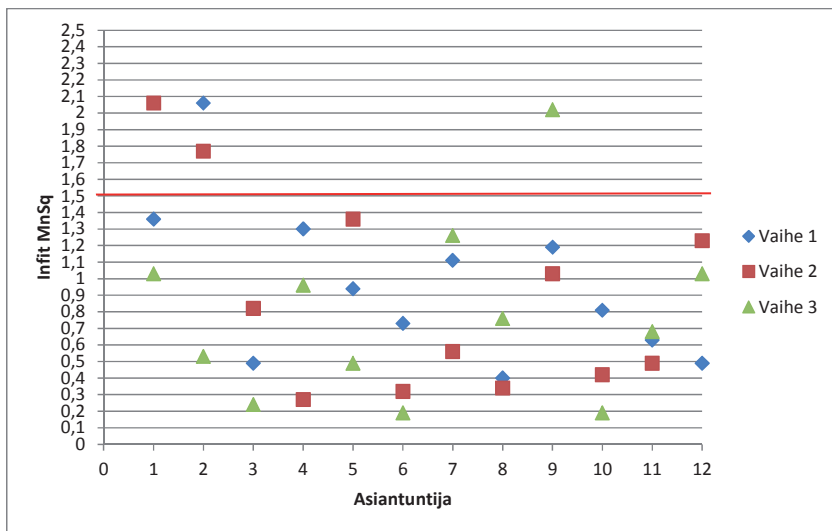
Asiantuntijoita tarkasteltaessa on kuitenkin hyvä muistaa, että tehtäväosioiden taitotasoille arviointi on aina subjektiivista toimintaa, johon vaikuttavat myös useat ulkopuoliset tekijät, kuten asiantuntijan henkilökohtaiset kokemukset ja ominaisuudet. Näiden tekijöiden vaikutus saattaa ilmetä ja vaikuttaa asiantuntijoiden näkemykseen eri tavoin, esimerkiksi ankaruutena/lempeytenä. Tästä syystä standardointiprosessin aikana asiantuntijoita johdateltiin eri tavoin kohti yhdenmukaisempaa käsitystä tehtäväosioiden vaativuudesta. Tällaisessa toiminnassa on aina vaarana se, että asiantuntijat menettävät kosketuksensa omaan arviointilinjaansa yrittäessään mukautua asiantuntijaryhmän yleiseen mielipiteeseen.

Standardointiprosessin eri vaiheista saatu aineisto analysoitiin Facets-ohjelmalla, joka perustuu IRT-menetelmiin pohjautuvaan Raschin mallin laajennukseen, MFRM (*Many-Facets Rasch Measurement*)-malliin (ks. Linacre 1989; Eckes 2011). Ensin aineisto analysoitiin 2-facetsin (ulottuvuudet: asiantuntijat ja tehtäväosiot) *rating scale*-mallilla, jonka tuloksissa kiinnitettiin erityisesti huomiota asiantuntijoiden taitotasoasteikon käytön johdonmukaisuuteen sekä siihen, oliko joukossa asiantuntijoita, jotka systemaattisesti määrittelivät tehtäväosioita korkeammalle tai matalammalle taitotasolle verrattuna muihin. Asiantuntijan toiminnan johdonmukaisuudesta tehtiin päätelmiä lähipainotetun keskineliöpoikkeaman (*Infit Mnsq*) avulla, jonka raja-arvot olivat tässä tutkimuksessa 0,5–1,5 (ks. Linacre 2002a, 2003). Asiantuntijoiden ankaruudesta kertovat arvot määriteltiin mallissa siten, että mitä suurempi arvo oli, sitä korkeammille taitotasoille asiantuntija arvioi tehtäväosioita. Koska 2-facetsin mallilla ei vielä kyetty selvittämään, vaihteliko yksittäisten asiantuntijoiden ankaruus tilastollisesti merkittävästi eri vaiheiden välillä, malli laajennettiin 3-facetsin malliksi (ulottuvuudet: asiantuntijat, tehtäväosiot ja standardointiprosessin vaiheet). Tämän pohjalta tehtiin bias-analyysi asiantuntijoiden ja prosessin eri vaiheiden välillä.

4.1.1 Asiantuntijoiden arviointilinjan johdonmukaisuus ja ankaruus

Kuviossa 2 kuvataan asiantuntijoiden arviointilinjan, eli tehtäväosioiden taitotasoille määrittelyn johdonmukaisuutta standardointiprosessin eri vaiheissa. Jokaiselle 12 asiantuntijalle on laskettu kuvioon kolme erillistä *Infit Mnsq*-arvoa. Näiden arvojen olles-

sa punaisen viivan alapuolella (eli *Infit Mnsq* < 1.5) on asiantuntijan arviointilinja kaikissa prosessin eri vaiheissa johdonmukainen. Kuten kuviosta 2 on havaittavissa, asiantuntijoiden työskentely oli prosessin eri vaiheissa johdonmukaista muutamaa poikkeusta lukuun ottamatta. Kahden asiantuntijan (1 ja 2) arviointilinja ei ollut johdonmukainen vaiheessa 2 (*Infit Mnsq* > 1.5), jossa heidän oli ensimmäisen kerran hyödynnettävä tilastollista tietoa tehtäväosioiden vaikeudesta taitotasokuvainten lisäksi. Asiantuntijoiden arviointilinja kuitenkin korjaantui vaiheessa 3. Lisäksi yhdellä asiantuntijalla (9) arviointilinja pysyi johdonmukaisena kahdessa ensimmäisessä vaiheessa, mutta viimeisessä vaiheessa arviointilinja katosi.



KUVIO 2. Asiantuntijoiden arviointilinjain johdonmukaisuus standardointiprosessin eri vaiheissa.

Tulokset arviointilinjasta olivat positiivisia: asiantuntijat pystyivät pitämään tehtäväosioiden taitotasolle määrittelyssä johdonmukaisuuden, vaikkakin koko prosessin ajan heidän käsityksiään ja tulkintojaan pyrittiin muokkaamaan lähemmäksi toisiaan. Kuviossa 2 esiintyvät matalat arvot (*Infit Mnsq* < 0.5), erityisesti vaiheissa 2 ja 3, eivät olleet huolestuttavia, koska ne johtuivat osittain siitä, että asiantuntijat yhä enenevässä määrin antoivat yhdenmukaisia taitotasoarvioita.

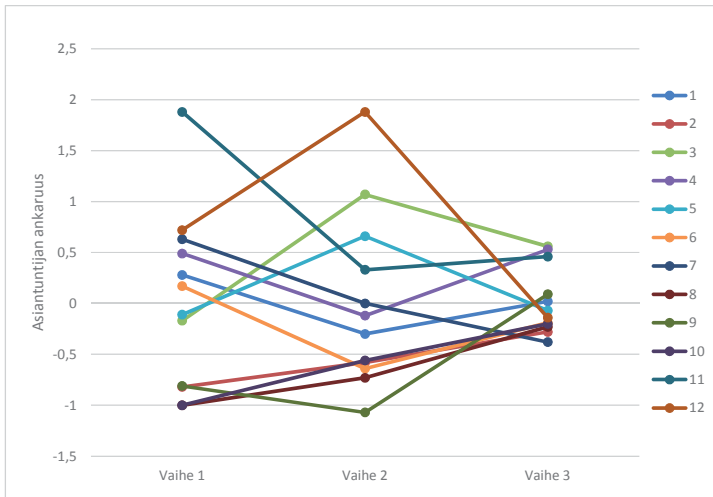
Taulukkoon 1 valittiin Facets-analyyseistä kolme havainnollistavaa tunnuslukua kuvaamaan asiantuntijaryhmän ankaruutta standardointiprosessin eri vaiheissa. Korkeat reliabiliteetin arvot eri vaiheissa kertoivat siitä, että asiantuntijat erosivat ankaruudeltaan toisistaan. Suurimmat erot ankaruusparametreissa olivat vaiheessa 2 ($R=0.95$), jolloin myös *stratan* arvo oli suurin 5.95. Sen mukaan asiantuntijat voitiin jakaa lähes kuuteen ankaruudeltaan toisistaan tilastollisesti eroavaan ryhmään. Edelleen vaiheessa

3 asiantuntijat voitiin jakaa yli kolmeen toisistaan tilastollisesti eroavaan ankaruusryhmään. Ankaruuden suhteen homogeenisen lopputuloksen saavuttamiseen ei lähtökohtaisesti uskottu, mutta keskustelujen toivottiin kuitenkin selvästi pienentävän eroja, ja näin tapahtuikin. Keskustelujen vaikutus näkyi myös siinä, että asiantuntijat olivat taitotasolle määrittelyissään vaihe vaiheelta yhdenmukaisempia, sillä yhdenmukaisuus kasvoi alun 35 %:sta 85 %:in (ks. taulukko 1, *exact agreements*).

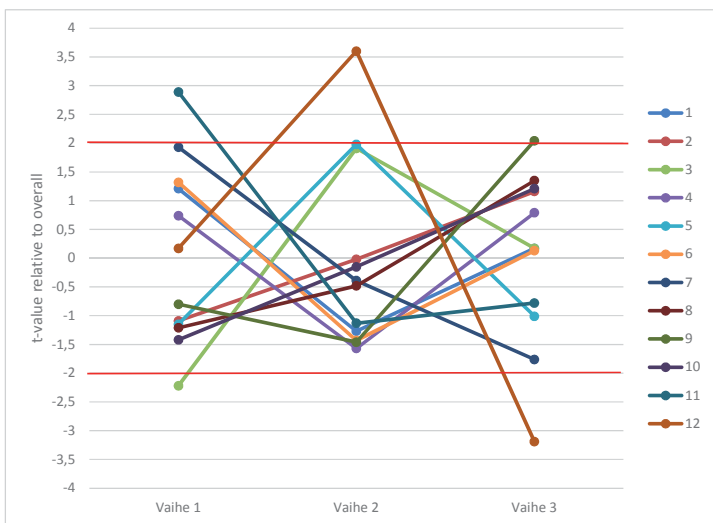
TAULUKKO 1. Asiantuntijoiden (n=12) erottelua kuvaavia yhteenvetolukuja standardointiprosessin eri vaiheissa.

	Vaihe 1	Vaihe 2	Vaihe 3
Strata	4.61	5.95	3.29
Rater separation reliability	0.91	0.95	0.83
Exact agreements	35 %	57 %	85 %

Kuvioissa 3 ja 4 asiantuntijoiden ankaruutta tarkastellaan bias-analyysin tuottamien tulosten avulla. Myös kuvioista 3 nähdään, kuinka erot asiantuntijoiden henkilökohtaisessa ankaruudessa olivat suurimmillaan vaiheessa 2 ja kaventuivat selvästi prosessin edetessä vaiheeseen 3. Kuvio 4 puolestaan havainnollistaa sitä, vaihtelee ko asiantuntijoiden ankaruus tilastollisesti merkitsevästi prosessin eri vaiheissa. Kuvioon 4 on merkitty tulkintaa helpottamaan ylä- ja alarajat, joiden ulkopuolella olevat arvot ovat tilastollisesti merkitseviä merkitsevyydellä 0.05. Kuvio 4 paljastaa, kuinka kolmen asiantuntijan ankaruus vaihtelee prosessin eri vaiheissa tilastollisesti merkitsevästi. Kahdella asiantuntijalla (3 ja 11) vaiheen 1 ankaruus eroaa vaiheista 2 ja 3 siten, että toinen heistä (3) on ollut alussa ankarampi ja toinen (11) lempeämpi verrattuna seuraaviin vaiheisiin. Yksi asiantuntija (12) on muuttanut linjaansa lempeästä ankaraksi vaiheiden 2 ja 3 aikana.



KUVIO 3. Asiantuntijoiden ankaruuserot standardointiprosessin eri vaiheissa.



KUVIO 4. Bias-kuvio asiantuntijoiden ja standardointiprosessin eri vaiheiden välillä.

Tulosten tulkinnessa on huomioitava se, että tulosten laskennassa käytetty malli (*MFRM*) olettaa asiantuntijoiden toimivan toisistaan riippumattomasti. Vaiheessa 1 ja osittain vaiheessa 2 tämä toteutuukin, mutta ei enää vaiheessa 3. Tämä voi aiheuttaa ongelmia malliin (ks. Linacre 2002b), ja siksi varsinkin vaiheen 3 tuloksiin on suhtauduttava varoen. Tarkastelut osoittivat kuitenkin sen, että kaiken kaikkiaan asiantuntijat

määrittivät tehtäväosioita taitotasolle johdonmukaisesti, ja standardointiprosessin edetessä asiantuntijoiden erot ankaruudessa kaventuivat.

4.1.2 Arviointiasteikon käytön laajuus

Asiantuntijat käyttivät vaiheessa 1 ja 2 taitotasoasteikkoa hyvin erilaisella laajuudella. Osa asiantuntijoista käytti arvioinneissaan koko arviointiskaalaa (1–6), kun taas toiset hyödynsivät vain kolmea taitotasoa (2–4). Yleisimmin asiantuntijat määrittivät tehtäväosiot kuitenkin taitotasolle 2–5. Taitotasoasteikon käytön laajuus vaihteli myös prosessin edetessä. Taulukossa 2 esitellään sitä, kuinka laajasti asiantuntijat standardointiprosessin eri vaiheissa käyttivät puheen ymmärtämisen 6-portaista taitotasoasteikkoa määrittellessään tehtäväosioita taitotasolle. Taulukossa 2 käytetyt prosentit viittaavat taitotasoarvioiden kokonaismäärään (12 asiantuntijaa * 35 tehtäväosiota = 420). Taulukosta nähdään, kuinka ennen viimeistä vaihetta (vaihe 3) käyty keskustelu ja tilastollinen palaute sekä omien arviointien vertailu muiden arviointeihin vaikuttivat taitotasoasteikon käyttöön siten, että asiantuntijat luopuivat asteikon ääripäiden tasoista (1 ja 6).

Taulukossa esitetyt taitotasoajakaumat näyttävät varsin samankaltaisilta prosessin eri vaiheissa, mutta todellisuudessa asiantuntijat vaihtoivat käsityksiään tehtäväosioiden vaikeustasosta, etenkin vaiheiden 1 ja 2 välillä. Tämä tarkoittaa sitä, että esimerkiksi asiantuntija, joka oli määritellyt vaiheessa 1 tietyn tehtäväosion taitotasolle 3, määritteli sen vaiheessa 2 taitotasolle 4, ja vastaavasti toinen asiantuntija toimi juuri päinvastoin.

TAULUKKO 2. Asiantuntijoiden taitotasoasteikon käytön laajuus.

Taitotaso	Vaihe 1 (n=420)	Vaihe 2 (n=420)	Vaihe 3 (n=420)
1	2 %	2 %	0 %
2	11 %	15 %	19 %
3	40 %	37 %	34 %
4	33 %	35 %	32 %
5	12 %	9 %	15 %
6	2 %	2 %	0 %

4.1.3 Tehtäväosiokohtaiset tulokset

Standardointiprosessin päämääränä oli määritellä tehtäväosioille mahdollisimman luotettavana pidettävät taitotasot ja löytää näin Yleisten kielitutkintojen osiopankkiin taitotasorajat. Tehtäväosioille määritellyt taitotasot analysoitiin vaiheittain sekä käytetyn *standard setting*-menetelmän mukaisesti että *MRFM*-malleilla, jonka jälkeen tuloksia

vertailtiin keskenään. Tässä artikkelissa näitä tuloksia ei esitellä, vaan kuten jo aiemmin mainitsimme, artikkelin painopiste on asiantuntijoiden toiminta standardointiprosessin aikana. Asiantuntijoiden onnistumista standardoinnissa tarkastellaan kuitenkin sen kautta, miten yhdenmukaisia tehtäväosioiden tasoarviot lopulta olivat, joten näitä tuloksia on myös tarpeellista kuvata lyhyesti.

Ensimmäisessä vaiheessa asiantuntijoiden arviot tehtäväosioiden vaativuudesta jakautuivat lähes poikkeuksetta kolmelle tai neljälle peräkkäiselle taitotasolle. Toisessa vaiheessa jakaumat kaventuivat kahteen tai kolmeen peräkkäiseen taitotasoon, mutta ne keskittyivät jo selkeästi tietyille taitotasolle. Joukossa oli lisäksi tehtäväosioita (28 %), joiden vaikeustasosta asiantuntijoiden mielipiteet jakautuivat varsin tasaisesti kahdelle eri taitotasolle. Vaiheessa 3, viimeisen keskustelun jälkeen, asiantuntijat olivat yksimielisiä 19 osion (54 %) taitotasosta, ja loput 16 tehtäväosiota (46 %) sijoitettiin edelleen kahdelle peräkkäiselle taitotasolle. Näissäkin tapauksissa – kahta osiota lukuunottamatta – selkeä enemmistö asiantuntijoista sijoitti tehtäväosiot kuitenkin samalle taitotasolle.

4.2 Asiantuntijoiden kommentteja prosessista

Kun varsinainen tehtäväosioiden standardointi oli ohi, asiantuntijoilta pyydettiin vielä kirjallista palautetta standardointiprosessista. Koska suurimmalla osalla asiantuntijoista oli kokemusta ennestään tekstin ymmärtämisen taitotasolle asettamisesta, he varsin usein vastauksissaan vertasivat näiden kahden taidon standardointiprosessia keskenään.

- (1) Kuullun ymmärtämisen osiossa oli vaikeampi asettua osallistujan nahkoihin. Tekstin ymmärtämisessä – ehkä harjoittelun vuoksi – jo vähän helpompaa. Mutta myös taitoina erilaiset. (Elli)
- (2) Luetun ymmärtäminen tuntui helpommalta kuin kuullun ymmärtäminen. (Aliisa)

Puheen ymmärtämisen taitotasolle asettamista pidettiin vaikeana sen vuoksi, että sekä itse puheen ymmärtäminen että standardointi ovat prosesseina monitahoisia luonteeltaan. Puheen ymmärtäminen taitona vaatii asiantuntijoista monen asian hallintaa yhtäaikaaisesti, mutta myös taitotasolle asettaminen edellytti usean eri asian ja dokumentin hallintaa.

- (3) Vaikeaa. Liikkuvia palikoita on liian paljon. (Sirkka)
- (4) On niin monta asiaa mitkä pitää ottaa huomioon: kuvaimet, testin taitotaso, kysymysten taitotaso ja arviointikriteerit. (Erja)

- (5) Hankaluutta tietää, mihin oikeastaan tulisi nojata: taitotasokuvaimiin, prosentteihin/tuloksiin vai kokonaisvaltaisempaan tulkintaan tekstistä ja tehtävistä. En oikein osannut ajatella, mikä on tärkein kriteeri määrittelyssä. Minusta ymmärtäminen on niin monitahoinen taito, että tällä tavalla tuntui vaikealta päästä sen ytimeen. (Jaana)

Jotta tehtäväosiot voidaan asettaa taitotasolle, täytyy asiantuntijoiden tuntee asettamisen apuna käytettävät kuvaimet hyvin. Vaikka Yleisten kielitutkintojen asiantuntijoilla oli paljon kokemusta opetustyönsä ja Yleisten kielitutkintojen kautta erilaisten kuvainten käytöstä, he kokivat puheen ymmärtämisen kuvainten käytön vaikeaksi taitotasojen asettamisessa. Vaikeudet liittyivät siihen, että heistä kuvaimissa puheen ymmärtämisen taito kuvattiin varsin yleisellä tasolla, ja toisaalta taas kuunneltavan tekstin ja siihen kohdistuvien osioiden vaikeustasot usein olivat varsin erilaiset, mikä vaikeutti niiden taitotason määrittelyä tietyille tasolle.

- (6) Meni aika intuitiivisesti, monessa kohdin asteikosta ei paljon apua. (Jaana)
- (7) Epävarmuutta aiheuttaa se, että tekstin ja kysymyksen vaikeustaso voivat olla hyvin erilaiset. (Kaisa)

Vaikka määrittelyä ohjaavat kuvaimet, taitotasolle asettamisen taustalla vaikuttivat usein myös asiantuntijan henkilökohtaiset ominaisuudet ja kokemukset. Ennen kaikkea omia kokemuksia oppijoista käytettiin hyväksi prosessin aikana, sillä pääosa asiantuntijoista työskentelee päivätyössään testiin osallistujien kaltaisten oppijoiden kanssa.

- (8) Vaikea päättää, mitä linjaa noudattaa: millaisen testattavan kuvittelee tehtävää tekemään. (Aulikki)
- (9) Monta muuttujaa – teksti, tehtävä, oma kokemus. (Minttu)

Kokemus kielenoppijoista oli myös muokannut asiantuntijoiden käsitystä siitä, mitä tietyllä taitotasolla oleva suomen kielen oppija osaa ja mitä hän ei osaa. Tämä tuli konkreettisesti esiin vaiheen 2 aikana (ks. kuvio 1), jolloin asiantuntijat saivat tietää tehtäväosioiden vaikeusjärjestyksen, joka oli määritelty testiin osallistuneiden kielenoppijoiden vastausten perusteella. Muutamien tehtäväosioiden vaikeustaso yllätti asiantuntijat, sillä he olivat arvioineet tehtävän sen tilastolliseen tietoon verrattuna helpommaksi tai vaikeammaksi. Tilastollinen tieto osioiden vaikeustasosta sai asiantuntijat pohtimaan omaa näkemystään tehtävän vaikeustasosta ja sitä, miksi tehtävä oli ollut vastoin heidän odotuksiaan helppo/vaikea osallistujille.

- (10) Opiskelijoiden tulokset yllättivät. (Kaisa)

- (11) Vaikeaa on löytää katkaisukohtat, koska tehtävät ovat omien arviointien mukaan "väärässä" järjestyksessä. (Elli)
- (12) Vaikka tuntui helpolta, lopputulos oli usein erilainen kuin mitä tilastotieto kertoo. Olo oli jonkin verran ristiriitainen, mutta tuli myös oivalluksia, miksi joku tietty osio olikin lopuksi helpompi tai vaikeampi kuin kuvittelin. (Aliisa)
- (13) Edellistä epävarmemmalta. Onnistumisprosentit sekoittivat pään lopullisesti. (Sirkka)

Vaikka tilastotiedot tehtäväosioista eivät välttämättä aina vastanneet asiantuntijan omaa käsitystä osion tilastollisesta vaikeustasosta, osioista saatu tilastotieto kuitenkin lähes kaikkien asiantuntijoiden mielestä helpotti taitotason määrittelyä.

- (14) Vaikutti paljonkin. Muutti käsitystä joistakin tehtävistä. (Jaana)
- (15) Ohjasi epävarmoissa kohdissa. Antoi ymmärrystä, mikä osio on osoittautunut vaikeaksi itse testitilanteessa. (Ronja)
- (16) Helpotti ehkä hieman päätösten tekemistä. (Tea)
- (17) Helpottavasti, oli helppo tehdä päätös. (Veera)
- (18) Vaikutti yhtenäistävästi ja pohdintaa tarkentavasti. (Aulikki)

Muutamasta asiantuntijasta tilastotieto kuitenkin ohjasi ehkä liikaakin toimintaa, sillä oli vaikea asettaa tilastotietoja vastaan, jos heidän oma käsityksensä tehtäväosion vaikeustasosta oli alun perin ollut toisenlainen.

- (19) Jos kerran 4000 testattavaa on jotain mieltä, pakkohan sitä on uskoa. (Aliisa)
- (20) Asettamisesta tuli "mekaanisempaa" sikäli, että katsoin vaikeustasoarvioita ja vastausprosentteja enkä pohtinut vain tehtävää ja sen helpoutta/vaikeutta. (Elli)
- (21) Oli helpompi siinä mielessä, että järjestys oli tiedossa, tarvitsi vain miettiä, kummalle puolelle rajatapaukset menevät. (Erja)

Taitotasojen asettamisessa keskustelu sekä kuvaimista että tehtäväosioiden vaikeustasoista kuuluvat osioiden tilastollisen tiedon lisäksi keskeisesti standardointiprosessiin. Keskustelujen avulla pyritään siihen, että asiantuntijoilla on samankaltainen käsitys taidon kehittymisestä ja että he tulkitsevat käytettävissä olevat taitotasokuvaimet samalla tapaa. Asiantuntijoista keskustelut olivat hyödyllisiä ja ne myös helpottivat osioiden asettamista taitotasoille.

- (22) Keskustelu antoi näkökulmia, jotka ohjasivat arvioimaan. Hyvä, että on tilaisuus pohdiskella rajoja eri taitotasojen välillä ja muuttaa perustellun keskustelun jälkeen. (Ronja)
- (23) Kompromisseja joutui tekemään sekä muiden vastausten että keskustelun perusteella. (Erja)
- (24) Toisten analyttinen argumentointi vakuutti. (Minttu)
- (25) On hyvä, että pyritään yhdenmukaiseen lopputulokseen ja vaihdetaan näkemyksiä. (Kaisa)

Keskustelu yhtenäisti asiantuntijoiden näkemyksiä osioiden taitotasosta, mutta heitä hämmästyttivät vaikeustasoihin liittyvät näkemyserot. Suurimmassa osassa tehtäväosioita asiantuntijoiden mielipiteet tehtävien vaikeustasoista olivat samankaltaisia, mutta muutamissa osioissa taitotasojakauma oli varsin laaja.

- (26) Hajonta kummastutti. (Veera)
- (27) Kuinka paljon variaatiota arvioijien välillä sallitaan? (Aulikki)

Vaiheessa 3 (ks. kuvio 1) vaikeustason mukaan järjestetyille osioille täytyy määritellä katkaisukohdat, joissa raja taitotasojen välillä sijaitsee. Osalle yhteisen päätöksen hyväksyminen taitotasorajoista oli helppoa, eikä heillä ollut keskustelun jälkeen vaikeuksia taipua toisten kanssa yhteiseen päätökseen.

- (28) Oli helppoa mukautua yhteisiin taitotasorajoihin, ei mitään ongelmia. (Aliisa)
- (29) ”Niska” taipuu jo tässä vaiheessa. (Sirikka)
- (30) Olin jo edellisillä kerroilla oppinut, ettei muutos ole pahasta ☺. (Tea)

Osalle asiantuntijoista taas oman taitotasorajan muuttaminen yhtäläiseksi muiden kanssa herätti ristiriitaisia tuntemuksia. He eivät joko hyväksyneet yhteistä päätöstä tai he eivät kaikilta osin olleet rajasta samaa mieltä, mutta muuttivat sen kuitenkin yleisen keskustelun perusteella.

- (31) Osa muutoksista oli vähän vastentahtoisia, joistakin olin samaa mieltä. Kaikkia en kuitenkaan muuttanut. (Elli)
- (32) Jokin raja-arvo pysyi aika hyvin kohdillaan, toista rajaa muutin toisten mielipiteiden vaikutuksesta. Jäi lopulta kuitenkin hieman epäselväksi, miten raja-arvoja on mielekästä asettaa. (Jaana)

Asiantuntijoiden antamat palautteet vahvistavat useita prosessin aikana saatuja tuloksia, mutta arvioitaessa tehtäviä ja niihin liittyviä osioita ei voida koskaan poissulkea tehtävien ja osioiden sisältöihin liittyviä näkemyksiä eikä niiden vaikutusta vaikeustason määrittelyyn. Vaikka puheen ymmärtämisen tehtävävalikoima koostui tilastollisesti hyvin toimivista ja osallistujia hyvin erottelevista osiosta, asiantuntijat saattoivat nähdä niissä ongelmia, jotka myös vaikuttivat vaikeustason määrittelyyn. Keskustelujen ja palautteen perusteella oli huomattavissa, että asiantuntijat tulkitsevat sisällöllisestikin osiot varsin monella tapaa. Vaikka keskusteluissa tulee keskittyä taitotasoihin, ei tehtävien sisältöön liittyviä kysymyksiä voi täysin ohittaa, koska niillä on myös vaikutusta asiantuntijan arvioihin.

5 Pohdinta

Vaikka taitotasolle määrittelyä ohjaavat taitotasokuvaimet, prosessin taustalla vaikuttivat tutkimuksemme mukaan myös asiantuntijan henkilökohtaiset ominaisuudet ja kokemukset. Henkilökohtaiset erot ja erilaiset tulkinnat käytettävistä arviointikuvaimista näkyivät ennen kaikkea vaiheessa 1, jossa asiantuntijoiden näkemykset tehtäväosioiden vaativuudesta vaihtelivat välillä voimakkaastikin. Tämä näyttäytyi myös asiantuntijoiden välisinä eroina ankaruudessa. Tästä syystä tämän vaiheen tulosten perusteella ei tehtäväosiolle voitu määrittellä luotettavia taitotasoarvioita. Syynä suureen vaihteluun asiantuntijoiden välillä oli todennäköisesti se, että määrittelyä ohjasivat heidän omat tulkintansa taitotasoihin ja että he ajattelivat vaikeustasoa määrittellessään erilaisia kielinoppijoita (*borderline examinee*). Vaiheen merkitystä voidaan kuitenkin perustella sillä, että vaihe konkretisoi hyvin taitotasoihin liittyviä näkemyseroja ja että keskustelu kuvaimista ja osiosta on tarpeellinen, jotta voidaan päästä lähemmäksi samankaltaista tulkintaa osioiden vaativuustasoista.

Vaiheessa 2 asiantuntijoiden tulkinnat tehtäväosioiden vaikeustasoista vaihtelivat vielä suuresti, vaikka asiantuntijoilla oli käytössään osioihin liittyvät tilastotiedot. Eroja voidaan selittää sillä, että tilastotietojen perusteella useat asiantuntijat vaihtoivat käsitystään tehtäväosioiden vaikeustasosta. Lisäksi käytössä ollut menetelmä ohjasi heitä etsimään selkeät katkaisukohtat, joissa peräkkäisten tehtäväosioiden vaikeustaso oli ratkaisevasti erilainen. Tämä aiheutti muutamille asiantuntijoille ongelmia, koska heistä osa tehtäväosioista jäi menetelmän vuoksi väärälle taitotasolle katkaisukohtan takia.

Vaiheessa 3 asiantuntijoiden mielipiteissä tehtäväosioiden vaativuudesta oli vielä eroja, mutta he kuitenkin olivat lähentyneet toisiaan vaihe vaiheelta. Keskusteluilla ja tilastollisella palautteella vaiheesta 2 oli selkeästi merkitystä erojen pienenemiseen, sillä ne herättivät asiantuntijoita pohtimaan omaa näkemystään ja vertaamaan

sitä toisten asiantuntijoiden näkemyksiin. Skorupksi (2012) ja Reckase ja Chen (2012) ovat myös havainneet, että mitä enemmän asiantuntijat keskusteleivat, sitä lähemmäksi heidän näkemyksensä taitotasosta tulevat. Keskustelujen merkitystä arvioitaessa on kuitenkin hyvä muistaa, että keskusteluihin liittyvät aina ryhmädynamiikkaan kuuluvat piirteet. Aktiivisilla ja voimakkailla persoonilla on vaikutusta standardointiprosessissa (esim. Skorupksi 2012; McGinty 2005), ja tämän havaitsivat myös tutkimuksemme osallistuneet asiantuntijat. Osa asiantuntijoista koki taipuvansa keskustelun perusteella muuttamaan näkemystään, mutta toki joukossa oli myös niitä, jotka halusivat pitää oman käsityksensä vaikeustasosta keskustelujen jälkeenkin.

Keskustelua ratkaisevampi merkitys asiantuntijoiden ankaruuserojen kaventumiseen oli osioihin liittyvällä tilastollisella tiedolla (ks. myös Reckase & Chen 2012). Tilastolliseen tietoon suhtauduttiin kuitenkin kaksijakoisesti: se helpotti päätöksentekoa, mutta se myös aiheutti ristiriitaisia tuntemuksia, koska oma näkemys vaikeustasosta ei aina vastannut tilastollisen tiedon antamaa informaatiota. Tulosten näkökulmasta tilastollinen tieto näyttöytyi positiivisena tekijänä, mutta palautteen ja keskustelujen perusteella sen arvo oli ajateltua suurempi. Tilastotiedon saamisen jälkeen osa arvioijista teki selkeästi päätökset pelkästään sen perusteella. Tämä todennäköisesti vaikutti siihen, että taitotasokuvaimet jäivät taitotason määrittelyssä sivurooliin. Tilastotietoon nojautuminen tuntui asiantuntijoista turvallisemmalta, sillä useat heistä kokivat taitotasokuvainten käytön vaikeana, koska taidon kuvaus niissä oli hyvin yleisellä tasolla.

Tarkasteltaessa asiantuntijoiden johdonmukaisuutta ja ankaruutta voidaan todeta standardointiprosessin onnistuneen, vaikka asiantuntijoiden välille jäi ankaruuseroja vielä viimeisenkin vaiheen jälkeen. Ankaruus on kuitenkin pysyvä ominaisuus, ja tästä syystä täydellistä yhdenmukaisuutta ei voida saavuttaa. Hälyttävää sen sijaan oli se, että muutaman asiantuntijan ankaruus vaihteli voimakkaasti prosessin aikana. Todennäköisesti taustalla olivat epävarmuus ja taipumus muuttaa helposti omaa näkemystään eri syiden vuoksi. Vaihtelu toi kuitenkin mukanaan sen kysymyksen, miten vastaisuudessa voimme saada paremmin tietoa siitä, mitkä asiat vaikuttavat asiantuntijan päätöksentekoon kussakin vaiheessa ja miten.

Tutkintojärjestelmän näkökulmasta puheen ymmärtämisen standardointikokeilu oli hyvä alku tulevia käytänteitä ajatellen. Itse prosessi oli selkeä, ja tulokset paranivat vaihe vaiheelta. Muutamat asiat kuitenkin vaativat vielä kehittämistä ennen seuraavaa standardointikertaa.

Ratkaistavat asiat liittyvät ensinnäkin kuvainten yksityiskohtaisuuteen. Tutkimuksemme asiantuntijoiden näkemyksiä kuvainten puutteista ja ongelmista vahvistavat myös aikaisemmat tutkimukset (esim. Papageorgiou 2009; Brunfaut & Harding 2014; Moe 2009; Figueras, Kaftandjieva & Takala 2013). Näissä kritiikki kohdistui mm. siihen, että viitekehukseen pohjautuvat kuvaimet sijoittuvat jokapäiväisen elämän kontekstiin, mutta testitilanne ei ole normaali jokapäiväisen elämän tilanne, ja tästä syystä kuvai-

met eivät välttämättä ole kaikilta osin käyttökelpoisia testitilanteessa. Lisäksi Papageorgiou (2009) tutkimukseen osallistuneet asiantuntijat näkivät kuvaimissa muitakin puutteita, jotka liittyivät kuvainten sanavalintoihin ja sisältöihin. Standardointiprosessin onnistumisen kannalta kuvainten on kuitenkin tärkeä olla yksityiskohtaisuudeltaan sellaiset, että asiantuntijat voivat niihin tukeutua myös tilanteissa, joissa tekstin ja osioiden vaikeustasot eroavat toisistaan paljon tai joissa osion vaikeustaso sijoittuu heistä kahden taitotason rajalle (*borderline case*).

Toinen kehittämistä edellyttävä asia on kahden taitotason välillä olevan ns. *borderline*-suorittajan (*borderline examinee*) taidon tarkempi määrittely. Ongelmat tässä liittyivät ennen kaikkea siihen, millaista oppijaa asiantuntijat ajattelivat, ja sitä kautta luonnollisesti siihen, kuinka he hahmottivat eri taitotasot. Kuvainten selkeyttäminen ja niiden avaaminen keskusteluissa ovat myös tämän ongelman ratkaisemiseksi keskeinen kehittämiskohde.

Suurimmat kysymykset kuitenkin liittyvät eri vaiheiden tulosten käyttöarvoon ja menetelmän valintaan. Aikaisemmat tutkimukset osoittavat, että menetelmän valinnalla on vaikutusta tuloksiin, ja tässä standardointikokeilussa tämä myös tuli selkeästi ilmi. Vaiheen 1 ja 3 tulokset eivät tukeneet toisiaan vaan erosivat toisistaan merkittävästi. Vaiheessa 1 käytetyn menetelmän tuloksissa oli niin paljon hajontaa, että niiden pohjalta ratkaisuja oli mahdotonta tehdä, joten vaihtoehdoksi jäi määritellä taitotasorajat osiopankin vaikeustasoskaalalle vain vaiheen 3 tulosten pohjalta. Pohdittavaksi jääkin, mikä merkitys vaiheella 1 on tulevaisuudessa. Tarvitaanko vaihetta vain konketisoimaan asiantuntijoille heidän välisiä näkemyserojaan ja tutustuttaamaan heidät tehtäviin ja taitotasokuvaimiin, vai voidaanko vaiheesta muokata vertailukelpoinen menetelmä vaiheiden 2 ja 3 rinnalle? Jatkossa on myös mietittävä, miten varmistetaan että asiantuntijat ovat ymmärtäneet standardointiprosessin tehtävänannon.

Standardointikokeilu herätti lisäksi kysymyksiä siitä, parantaako vai heikentääkö uuden menetelmän lisääminen prosessiin tulosten luotettavuutta. Useiden eri menetelmien käyttö standardointiprosessin aikana voi tehdä asiantuntijoiden näkökulmasta prosessista liian raskaan ja monimutkaisen. Koska tässä kokeilussa oli myös paljon hyviäkin puolia, kokeilussa käytettyjä menetelmiä voidaan mahdollisesti vielä soveltaa ja validoida siten, että asiantuntijat jaetaan kahteen eri ryhmään ja näistä kahdesta ryhmästä saatuja tuloksia verrataan toisiinsa. Brunfaut ja Harding (2014) kokeilivat vastaavanlaista lähestymistapaa kahdella eri asiantuntijaryhmällä linkittäessään GEPT:n (General English Proficiency Test) puheen ymmärtämisen testiä Eurooppalaiseen viitekehukseen, ja he saivat kokeilusta rohkaisevia tuloksia.

Kirjallisuus

- Bejar, I. 2008. Standard setting: What is it? Why is it important? *R & D Connections* 7. Princeton: Educational Testing Service. https://www.ets.org/Media/Research/pdf/RD_Connections7.pdf.
- Brunfaut, T. & L. Harding 2014. Linking the GEPT listening test to the Common European Framework of Reference. *LTTC-GEPT Research Reports RG-05*. Taipei: The Language Training and Testing Center. <https://www.ltcc.ntu.edu.tw/ltcc-gept-grants/RReport/RG05.pdf>.
- Buck, G. 2001. *Assessing Listening*. Cambridge: Cambridge University Press.
- Cizek, G. & M. Bunch 2007. *Standard Setting. A guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, Cal: Sage.
- Council of Europe 2009. *Relating language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment (CEFR). A Manual*. Strasbourg: Language Policy Division. http://www.coe.int/t/dg4/linguistic/Source/ManualRevision-proofread-FINAL_en.pdf.
- Eckes, T. 2011. *Introduction to Many-Facet Rasch Measurement. Analyzing and evaluating rater-mediated assessments*. Frankfurt am Main: Peter Lang.
- Figueras, N., F. Kaftandjieva & S. Takala 2013. Relating a reading comprehension test to the CEFR levels: a case of standard setting in practice with focus on judges and items. *The Canadian Modern Language Review*, 69 (4), 359–385.
- Kaftandjieva, F. 2010. *Methods for setting cut scores in criterion-referenced achievement tests. A comparative analysis of six recent methods with an application to tests of reading in EFL*. EALTA. http://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf.
- Linacre, J. 1989. *Many-facet Rasch measurement*. Chicago, Ill: MESA Press.
- Linacre, J. 2002a. What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16 (2), 878. <https://www.rasch.org/rmt/rmt162f.htm>.
- Linacre, J. 2002b. Judge ratings with forced agreement. *Rasch Measurement Transactions*, 16 (1), 857–8. <http://www.rasch.org/rmt/rmt161.pdf>.
- Linacre, J. 2003. Size vs. significance: infit and outfit mean-square and standardized chi-square fit statistic. *Rasch Measurement Transactions*, 17 (1), 918. <https://www.rasch.org/rmt/rmt171n.htm>
- McGinty, D. 2005. Illuminating the “black box” of standard setting: an exploratory qualitative study. *Applied Measurement in Education*, 18 (3) 269–287.
- Messick, S. 1989. Validity. Teoksessa R. Linn (toim.) *Educational measurement* (3. painos). Phoenix, Az: Oryx, 13–104.
- Messick, S. 1994. The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2), 13–23. <http://journals.sagepub.com/doi/pdf/10.3102/0013189X023002013>.
- Moe, E. 2009. Jack of more trades? Could standard setting serve several functions? Teoksessa N. Figueras & J. Noijons (toim.) *Linking to the CEFR levels: research perspectives*. Cito, Institute for Educational Measurement, Council of Europe, European Association for Language Testing and Assessment (EALTA), 131–137.
- Neittaanmäki, R. & T. Hirvelä 2014. Yleisten kielitutkintojen osallistujat taustatietojen valossa. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.) *Yleiset kielitutkinnot 20 vuotta*. Raportit ja selvitykset 2014: 16. Helsinki: Opetushallitus, 46–60.
- Papageorgiou, S. 2009. Analyzing the decision-making process of standard setting participants. Teoksessa N. Figueras & J. Noijons (toim.) *Linking to the CEFR levels: research perspectives*.

- Cito, Institute for Educational Measurement, Council of Europe, European Association for Language Testing and Assessment (EALTA), 75–78.
- Reckase, M. 2009. Standard setting theory and practice: issues and difficulties. Teoksessa N. Figueras & J. Noijons (toim.) *Linking to the CEFR levels: research perspectives*. Cito, Institute for Educational Measurement, Council of Europe, European Association for Language Testing and Assessment (EALTA), 13–20.
- Reckase, M. & J. Chen 2012. The role, format, and impact of feedback of standard setting panelists. Teoksessa G. Cizek (toim.) *Setting performance standards: foundations, methods, and innovations* (2. painos). New York: Routledge, 149–164.
- Rost, M. 1990. *Listening in language learning, applied linguistics and language study*. Singapore: Longman.
- Saarela-Kinnunen, M. & J. Eskola 2001. Tapaus ja tutkimus = Tapaustutkimus. Teoksessa J. Aaltola & R. Valli (toim.) *Ikkunoita tutkimusmetodeihin I*. Jyväskylä: Gummerus, 158–169
- Skorupski, W. 2012. Understanding the cognitive processes of standard setting panelists. Teoksessa G. Cizek (toim.) *Setting performance standards: foundations, methods, and innovations* (2. painos). New York: Routledge, 135–147.
- Tuomi, J. & A. Sarajärvi 2009. *Laadullinen tutkimus ja sisällönanalyysi*. Helsinki: Tammi.
- Törmäkangas, K. & T. Törmäkangas 2009. *Osioanalyysi testien arvioinnissa*. Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.
- Weir, C. 2005. *Language testing and validation. An evidence-based approach*. Lontoo: Palgrave Macmillan.
- Wright, B. & M. Stone 1979. *Best test design*. Chicago, Ill: MESA Press.

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 inhimillisestä 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 lepuhko, 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 kolmesta 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 kokemattomien 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 suomi toisena ja vieraana kieltenä -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: analyttistä 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 jäsenneltyä 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 vaikea 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 vaikea päästä eroon. (Hannele)
- (54) Harvoin tulee sillein eteen kirjoitettua tekstiä, että ei tietäis. 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

Huhta, A. & R. Hildén (toim.) 2016. *Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 110–141.*

Marja Seilonen, Minna Suni, Marita Härmälä & Reeta Neittaanmäki

Jyväskylän yliopisto

Ammatillisen kielitaidon arviointikokeilu terveydenhuollon alalla

This article presents some key findings of the language assessment experiment implemented in the project *Health care Finnish: developing and assessing Finnish proficiency among health care professionals* (2014–2015). The core data includes test performances (n=204) of 35 internationally educated health care professionals in the Finnish National Certificates of Language Proficiency (FNCLP) intermediate level test (B1–B2) and the adjacent professional language assessment module developed for research purposes. The integrated tasks of the latter were completed electronically, and the focus was on such domain-specific communicative functions as giving advice, providing information, requesting action and reporting. The informants' test results were analyzed to figure out how the professional module completed the skill profiles obtained in the FNCLP, and what kind of challenges or added value the assessment of professional language skills might provide for the authorization processes in health sector. Additionally, the lived experiences of the participants were heard in the feedback interviews.

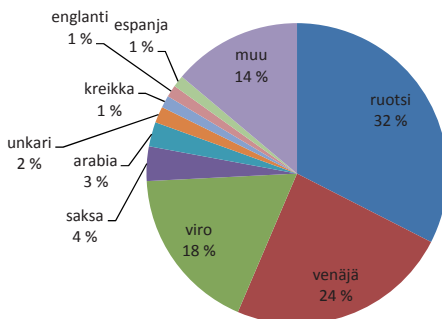
Keywords: Finnish as a second language, language proficiency, professional language skills, assessment, health care, working life

Asiasanat: suomi toisena kielenä, kielitaito, ammatillinen kielitaito, arviointi, terveydenhuolto, työelämä

1 Johdanto

Kansainvälisellä liikkuvuudella on terveydenhuollon alalla erityistä merkitystä. Esimerkiksi kriisialueilla humanitaarisen avun piirissä työskentelee aina eri maista tulleita kansainvälisten avustusjärjestöjen työntekijöitä, ja myös kehitysyhteistyön piirissä terveydenhuolto on keskeinen osa-alue. Kansainvälinen liikkuvuus on vahvistunut myös tavanomaisempien terveydenhuollon työtehtävien osalta (Ailasmaa 2015b). Esimerkiksi Suomesta lähdettiin edellisen syvän taloustaantumana aikaan parempien työmahdollisuuksien perässä erityisesti Isoon-Britanniaan ja Norjaan. Lähtijöitä on edelleen, mutta vielä enemmän on tulijoita: Suomessa on viime aikoina laillistettu noin 600 ulkomailla koulutuksensa saanutta terveydenhuollon ammattilaista vuodessa (OKM 2014).

Suomen terveydenhuollossa on kärsitty työvoimapulasta, mikä on johtanut myös kansainvälisiin rekrytointihankkeisiin ja niihin liittyviin lähtömaakoulutuksiin (Raunio 2013). Näiden hankkeiden myötä Suomeen on rekrytoitu sairaanhoitajia etenkin Filippiineiltä ja Espanjasta, mutta kohdennettua rekrytointia on tehty myös monissa sellaisissa Itä-Euroopan maissa, jotka nyt ovat EU-maita. Vielä mittavampaa on kuitenkin ollut terveydenhuollon ammattilaisten oma-aloitteinen työperusteinen maahanmuutto naapurimaista. Etenkin virolaisista lääkäreistä on tullut olennainen osa alan työvoimatarpeiden vaihteluun vastaajista, ja moni heistä on siirtynyt Suomeen pitkäaikaisiin työsuhteisiin. Euroopan unionin maiden välillä terveydenhuollon tutkinnot tunnustetaan suoraan, mikä on lisännyt ja helpottanut tämäntyyppistä työvoiman liikkuvuutta. Myös Venäjällä syntyneitä ja koulutautuneita terveydenhuollon ammattilaisia on Suomessa paljon, mutta heillä tie koulutustaan vastaaviin tehtäviin on pidempi, koska muut kuin EU/ETA-alueelta tulevat käyvät läpi monivaiheisen laillistamisprosessin. Kuvio 1 havainnollistaa lääkäreiden ammattikunnan kielitaustojen moninaisuutta; ainakin ruotsinkielisiin sisältyy sekä Suomessa että ulkomailla koulutautuneita.



KUVIO 1. Muut kuin suomenkieliset lääkärit Suomessa kielitaustan mukaan v. 2016. Lähde: Lääkärit 2016.

Terveydenhuollon ammateissa tarvittava ja edellytettävä kielitaito on Suomessa nyt kielipoliittisesti visainen ja ajankohtainen kysymys. Kasvava osuus alan ammattilaisista on muualla kuin Suomessa syntyneitä (Ailasmaa 2015a, 2015b), ja riittävän suomen tai ruotsin kielen taidon määrittäminen ja varmistaminen ovatkin keskusteluttaneet paitsi paikallisissa rekrytoinneissa myös valtakunnallisesti. Nykyään EU/ETA-alueen ulkopuolelta tulevien on Valviran laillistusprosessiin hakiessaan osoitettava Yleisissä kielitutkinnoissa tason 3 suomen tai ruotsin kielen taito, mutta työnantajan vastuulla on viime kädessä varmistaa kielitaidon riittävyys työtehtävissä.

Kielitaitolinjauksiin on laajassa asiantuntijatöryhmässä ehdotettu huomattavia muutoksia (OKM 2014):

- Samat kielitaitovaatimukset, jotka koskevat EU/ETA-alueen ulkopuolella koulutuksensa saaneita laillistuksen hakijoita, ulotettaisiin myös EU/ETA-alueella terveydenhuollon koulutuksen saaneisiin (EU:n Ammattipätevyysdirektiivin muutosdirektiivi 2013/55 mahdollistaa tämän; uusi käytäntö tuli Suomessa voimaan tammikuussa 2016, ja vastaava prosessi on harkinnassa eri EU-maissa; ks. Valvira 2016b).
- Vaatimustaso nostettaisiin Yleisten kielitutkintojen tasolta 3 tasolle 4 ja arvioitavaksi otettaisiin myös ammatillinen kielitaito.

Ehdotettuihin kielipoliittisiin muutoksiin liittyy myös laillistusta hakeville tarkoitettujen koulutuspolkujen valtakunnallinen uudistaminen, jota paraikaa pohjustetaan opetus- ja kulttuuriministeriön ja Euroopan sosiaalirahaston rahoittamissa laajoissa yhteistyöhankkeissa (ks. HY+ 2014; Urareitti 2015). Myös ammatillisen kielitaidon arvioinnin tutkimusperusteiseen kokeilemiseen on osoitettu opetus- ja kulttuuriministeriön rahoitusta. Tässä artikkelissa esiteltävä Terveydenhuollon suomi: suomen kielen taidon kehittäminen ja arviointi terveydenhuollon alalla -hanke on toteutettu Jyväskylän yliopistossa tämän kaksivuotisen rahoituksen turvin. Aihepiiriin liittyen on aiemmin toteutettu Helsingin kaupungin toimesta ASKI-hanke (ks. Piikki 2010), jossa keskityttiin kielitaidon arvioinnin kehittämiseen kunkin työntekijän ohjaamiseksi sopivaan kielikoulutukseen, ja pääkaupunkiseudun korkeakoulujen Kotimaisten kielten opetuksen kehittämishanke KOTI, jossa on luotu ammatillinen suomi toisena kielenä -viitekehys mm. sairaanhoitajan tehtäviin (Komppa, Jäppinen, Herva & Hämäläinen 2014).

Tässä artikkelissa esittelemme Terveydenhuollon suomi -hankkeen keskeisiä tuloksia ja päätelmiämme niiden pohjalta. Hankkeessa toteutettiin ammatillisen kielitaidon arviointikokeilu, johon osallistui 35 Suomeen ulkomailta muuttanutta terveydenhuoltoalalle kouluttautunutta henkilöä. Heillä oli lähtömaassa hankittu sairaanhoitajan, lääkärin tai hammaslääkärin koulutus sekä vaihtelevat kielitaustat, kielikoulutuspolut, työllisyystilanteet ja työhistoriat. Osallistujat suorittivat sekä Yleisten kielitutkintojen

keskitason tutkinnon että hankkeessa tutkimuskäyttöön laatimamme ammatillisen kielitaidon arviointimoduulin.

Pyrimme vastaamaan seuraaviin tutkimuskysymyksiin:

1. Miten osallistujat suoriutuivat Yleisten kielitutkintojen suomen kielen keskitason tutkinnossa ja ammatillisen kielitaidon osoittamisessa?
2. Miten ammatillisen kielitaidon arviointi täydentää Yleisten kielitutkintojen antamaa kuvaa osallistujien kielitaidosta?
3. Millaisia ovat osallistujien omat kokemukset ammatillisen kielitaidon arviointimoduulista verrattuna kokemuksiin suomen kielen keskitason tutkintoon osallistumisesta?

Edellisten pohjalta pohdimme myös sitä, mitä lisäarvoa ja haasteita ammatillisen kielitaidon arviointi mahdollisesti toisi terveydenhuollon laillistamisprosesseihin.

Esittelemme ensin terveydenhuoltoalan kielitaidon arvioinnin tutkimuksen antia sekä oman hankkeemme taustaa ja toteutusta. Tämän jälkeen käymme läpi analyysin päätulokset kolmen tutkimuskysymyksen valossa. Päättänessä pyrimme vastaamaan artikkelin tutkimuskysymysten pohjalta vielä ammatillisen kielitaidon arvioinnin lisäarvoa ja haasteita koskevaan kysymykseen kieli- ja koulutuspoliittisia näkökulmia punnitien.

2 Terveydenhuollon alan ammatillisen kielitaidon arviointi ja tutkimus

Ammatillisen ja yleisen kielitaidon välinen raja on häilyvä siinä missä yksittäisen alan yhteinen ammatillinen ydinkin. Tutkimuksessa esiintyykin erilaisia tapoja hahmottaa, mitä ovat ammatillinen ja yleinen kieli terveydenhuollon kontekstissa, ja millainen on kaikkien terveydenhuollon alalla toimivien yhteisesti tunnistama ammatillinen perusta, johon kuuluu tiettyjä toimintoja ja tehtäviä ja joka samalla käsittää myös yhteisen ammatillisen kielen.

Erytisalojen kielitaidon (*language for specific purposes, LSP*) arvioinnissa haasteena on ollut ja on edelleen, missä määrin esimerkiksi terveydenhuoltoalan laillistamisprosessissa vaadittu kielitaito vastaa työpaikan hoitotyössä tarvittavaa kielitaitoa (Elder, Pill, Woodward-Kron, McNamara, Manias, Webb & McColl 2012; Elder 2016). Kielitaitotestit mittaavat kielitaitoa vain epäsuorasti, vaikkakin tehtävät sinällään voivat olla hyvinkin autenttisia. Myös vuorovaikutuksen autenttisuutta on testaustilanteessa vaikeaa – ellei jopa mahdotonta – jäljitellä.

Niin sanottuja *high stakes* -testijärjestelmiä, eli osallistujan elämän kannalta erityisen merkityksellisiä testijärjestelmiä, on terveydenhuollon alalla tältä erää käytössä vain englanninkielisissä maissa. Onkin odotuksenmukaista, että myös tutkimus terveydenhuollon alan ammatillisesta kielitaidosta ja sen arvioinnista on näihin asti ollut paljolti englannin kieleen keskittyvää. Pienemmillä kielillä ja kielialueilla alakohtaisten testijärjestelmien kehittäminen ja ylläpito on suhteessa kalliimpaa, koska käyttäjäkunta on suppeampi, mutta esimerkiksi tehtäväpankin laadinta, tehtävistöjen validointi sekä osallistumis- ja arviointijärjestelmän rakentaminen ja ylläpito vaativat kuitenkin samat työvaiheet ja resurssit kuin suurellakin kielialueella toimittaessa. Pienemmillä kielialueilla terveydenhuollon ammattilaisten kielitaidon arviointi rajautuukin tyypillisesti yleiskielitaitoon, jolle standardoituja testejä on yleisesti olemassa. Tämä on tilanne paitsi Pohjoismaissa myös sellaisilla Euroopan verrattain suurillakin kielialueilla kuin saksan- ja ranskan-kielisissä maissa. Suurille kielialueille ulkomailta töihin hakeutuvilla voidaan yleisesti odottaa olevan jo valmiiksi melko vahva maan valtakielen taito, mikä esimerkiksi suomen kielen kohdalla on ani harvoin tilanne. Terveydenhuoltoalan työ on kaikkialla kielellisesti yhtä vaativaa, ja ammattilaiset suosivatkin maita, joiden kieliä he osaavat ennestään, pikemminkin kuin maita, joihin lähtö vaatii kokonaan uuden kielen haltuunottoa. Kielihaasteiden vastapainoksi suomalaisen terveydenhuolto- ja koulutusjärjestelmän hyvä maine sekä hyvät työehdot ja turvallinen yhteiskunta toimivat kuitenkin vetovoimatekijöinä.

Isossa-Britanniassa hoitoalan kielitaitotarpeiden tutkimus on keskittynyt varsinkin siihen, mikä on tehokasta viestintää, miten se opitaan ja millaisia haasteita hoitoalan viestintä asettaa täysin erilaisista kulttuureista ja terveydenhuoltojärjestelmistä tuleville työntekijöille (Wilkinson, Gambles & Roberts 2002; Philip, Manias & Woodward 2015; Eggenberger, Heimerl & Bennett 2012). Laillistamisprosessissa vaadittu kielitaito osoitetaan akateemisen kielitaidon testillä (*International English Language Testing System*, IELTS). Vaatimustaso terveydenhuollon alalla vastaa karkeasti arvioiden Yleisten kielitutkintojen tasoa 4+/5, ja se on osoitettava kaikissa osataidoissa.

Erityisalojen kielitaitotestejä on viime vuosikymmeninä tutkittu erityisesti Australiassa, ja tuloksia on vastikään julkaistu kootusti *Language Testing* -lehden teemanumerossa. Melbournen yliopistossa kehitetty OET (*Occupational English Test*) on yksi tunnetuimmista nimenomaan hoitoalalla käytössä olevista kielitaitotesteistä. Sitä käytetään Australian lisäksi myös Uudessa-Seelannissa 12:n eri terveydenhuollon ammattiryhmän kielitaidon arviointiin. OET koostuu neljästä osataidosta, joista puheen ja tekstin ymmärtäminen ovat yhteisiä kaikille ammattinimikkeille ja puhuminen ja kirjoittaminen ovat erityisalakohdaisia. Australiassa ammatillinen kielitaito ja substanssiosaaminen arvioidaan erikseen. (McNamara 1996: 40.) Suurin haaste OET:ssä onkin tasapainon löytäminen kielitaidon arvioinnin asiantuntijoiden ja substanssiosaajien kä-

sitysten välillä, jotka koskevat kielitaitotestin sisältöä, suoritusten arviointia ja tulosten tulkintaa (Elder 2016).

Erytysalojen kielitaidon arvioinnissa haasteena on alakohtaisen ja yleisen kielitaidon päällekkäisyys sekä se, tuleeko niitä ylipäätään lainkaan erottaa toisistaan (Davies 2001; ks. myös Kela & Komppa 2011; Härmälä 2008) vai onko kyse pikemminkin jatkumosta hyvin yleisen ja erikoistuneemman kielitaidon välillä. Myös arviointikriteerien tulisi kuvastaa tosielämän kielitaitovaatimuksia (McNamara 1996), eikä kielitaitotesteissä ole reilua vaatia sellaista osaamista, jota ei ns. syntyperäisiltä kielentaitajiltaakaan edellytetä (McNamara 1996, 1997). Esimerkiksi muistiinpanojen tekeminen on lääkärinkierron aikana hyvin erilaista kuin testitilanteessa. Kriteerit, joilla kielitaitoa arvioidaan työelämässä, ovat näin ollen avainasemassa. Niihin on kuitenkin vaikea päästä käsiksi, sillä ne ovat usein implisiittisiä ja ryhmän sisäiseen hiljaiseen tietoon perustuvia (Jacoby 1998), eikä niitä varsinkaan hoitoalalla päästä havainnoimaan suoraan potilastilanteissa. Usein kriteereitä eksplikoidaan vasta hoitotilanteen jälkeen epäsuorasti. Ammatillisen kielitaidon arvioinnissa onkin ongelmana, miten laatia sellaiset kriteerit, jotka vastaavat tosielämän vaatimuksia, ovat helposti operationalisoitavissa ja luotettavasti arvioitavissa (Elder & McNamara 2016).

Toisin kuin yleisen kielitaidon testeissä (esim. Yleiset kielitutkinnot), joissa tehtävät pyritään laatimaan niin, että osallistujan aiemmat tiedot aiheesta eivät helpottaisi tehtävistä selviytymistä, korostuu erityisalojen kielitaitotesteissä nimenomaan substanssiosaaminen. Testi ei ole tällöin enää pelkkä kielitaitotesti, vaan se pyritään räätälöimään kunkin osallistujaryhmän erityistarpeista lähtien. Testissä simuloitu kielenkäytökonteksti (esim. sairaala) ei kuitenkaan ole tosielämässä staattinen, vaan se muuttuu jatkuvasti (Hoekje 2016). Onkin vaikea mallintaa esimerkiksi hoitajan tarvitsemia vuorovaikutustaitoja, koska käytännössä puhelumppani vaihtuu koko ajan. Merkitykset myös syntyvät aina osapuolten välisessä vuorovaikutuksessa (Jacoby & Ochs 1995) eli nekään eivät ole staattisia. Simulaatioiden käyttöä hoitoalan ammatillisen kielitaidon arvioinnissa on kritisoitu esimerkiksi valtasuhteiden keinotekoisuudesta (De La Croix & Skelton 2009). Simulaatiot myös suosivat henkilöitä, joille testausympäristö ja testaus tapa ovat ennestään tuttuja (Atkins, Roberts, Hawthorne & Greenhalgh 2016).

Terveydenhuoltoalan viestintään sisältyy piirteitä, joita ei kielitaitotesteillä voida mitata. Esimerkiksi nonverbaalinen viestintä on tärkeä taito hoitajan ja potilaan välisen luottamuksen ja yhteisymmärryksen rakentamisessa (Pill 2016; Paananen 2015). Koska työelämäviestintä on muutakin kuin vain kielitaitoa, osaamisen tarkastelu kokonaisuutena on nähty relevantimpänä kuin sen erittely osataitoihin. Hoitoalalla esimerkiksi kuunteleminen on sekä psykologinen strategia että taito ymmärtää kuulemaansa tiedon rekisteröimiseksi ja kirjaamiseksi (Macqueen, Pill & Knoch 2016). Myös puhumisen ja kuuntelemisen jyrkkä erottaminen on potilaiden hoidon yhteydessä mahdotonta.

Macqueen ym. (2016) korostavat, että erityisalojen kielitaitotestissä testitulos tulisi nähdä eräänlaisena kynnyksen valmiutena (*boundary object*), jonka saavuttaminen mahdollistaa osaltaan työelämän akkulturaatioprosessin käynnistymisen. Heidän mukaansa työssä tarvittava kielitaito opitaan parhaiten työpaikalla. Työympäristön merkitys ammatillisen kielitaidon kehityksessä on noussut vahvasti esiin myös Suomessa tehdyissä tutkimuksissa (esim. Strömmer 2016; Suni tulossa 2017; Virtanen tulossa 2017).

3 Arviointikokeilun tausta ja toteutus

3.1 Tausta

Opetus- ja kulttuuriministeriön (2014) julkaisemassa työryhmämuistiossa on ehdotettu ammatillisen kielitaidon arviointia osana terveydenhuollon alan laillistamisprosesseja. Arviointitapa, jollaista tulevaisuudessa tavoitellaan, on kuitenkin toteutettava "Yleiset kielitutkinnot huomioiden" (OKM 2014: 48). Tästä syystä Terveydenhuollon suomi-hankkeessa päädyttiin siihen, että ammatillisen kielitaidon arviointikokeilun osallistajat suorittavat sekä Yleisten kielitutkintojen keskitason tutkinnon, joka mittaa säädöksissä nykyisin vaadittua taitotasoa 3 ja myös ehdotettua taitotasoa 4, että erillisen tutkimuskäyttöön laaditun ammatillisen kielitaidon arviointimoduulin.

Yleiset kielitutkinnot on aikuisille tarkoitettu yleiskielitaidon näyttötutkinto, joka perustuu lakiin ja asetukseen (Laki yleisistä kielitutkinnoista 964/2004). Tutkintotodistuksen pääasialliset käyttökohteet ovat kansalaisuuden hakeminen, kielitaidon osoittaminen työnantajalle ja opiskelu. Tutkintokieliä on yhdeksän ja tutkintotasoa kolme (perustaso ~ A1–A2; keskitaso ~ B1–B2; ylin taso ~ C1–C2). Eniten tutkinnon suorittajia on suomen kielen keskitasolla: yli 7000 henkilöä vuosittain (vuonna 2015 heitä oli 7215). Valtaosa heistä käyttää todistusta Suomen kansalaisuuden hakemiseen. Yhtenä kansalaisuuden saamisen edellytyksenä on vähintään keskitason tutkinto suomen tai ruotsin kielessä.

Kielitaitoa arvioidaan Yleisissä kielitutkinnoissa tekstin ja puheen ymmärtämisen, kirjoittamisen ja puhumisen tehtävien perusteella. Tutkintotodistuksessa ilmoitetaan osallistujan kussakin osataidossa saama taitotaso erikseen; yleistaitotasosta luovuttiin vuonna 2011. Tehtävät perustuvat ns. toiminnalliseen kielitaitokäsitykseen eli siihen, miten luontevasti tutkinnon suorittaja selviytyy erilaisissa todellisen elämän kielenkäyttötilanteissa. Työelämä ja työelämän yleiset kielenkäyttötilanteet ovat tutkinnossa mukana yhtenä yhdeksästä sisältöalueesta, ja joitakin niihin liittyviä kuvauksia löytyy siten myös tutkinnon taitotasoasteikosta. Esimerkiksi kirjoittamisen osakokeessa tasolla 3 oleva tutkinnon suorittaja *pystyy kirjoittamaan hyvin lyhyitä raportteja, jotka*

noudattavat tavanomaisia standardimuotoja ja jotka välittävät rutiininomaista asiantietoa. Tasolla 4 kieli on tarkempaa, sanastollisesti varioivampaa ja tutkinnon suorittaja *pystyy kirjoittamaan raportin, jossa tuo esiin keskeisiä seikkoja ja niitä tukevia yksityiskohtia sekä pystyy tekemään synteesejä eri lähteistä poimituista tiedoista.* Yleiset kielitutkinnot on kuitenkin nimensä mukaisesti yleiskielitaitoa mittaava standardoitu testijärjestelmä, eikä siinä siis ole ammattialakohtaisia sisältöjä, vaikka tällaisia yleisiä työelämäteemoja ja -taitoja voidaankin käsitellä.

Terveydenhuollon laillistamisprosessiin pääseminen edellyttää nykyään todistusta Yleisten kielitutkintojen tasolla 3 (Valvira 2016a). Vaikka yleistaitotasoa ei todistuksessa olekaan enää määritelty, Valviralla on oma vakiintunut tapansa määrittää, kenellä tämä taso on. Käytännössä taso tulee osoittaa kaikissa neljässä tutkinnon osataidossa, joita ovat puhuminen, puheenymmärtäminen, kirjoittaminen ja tekstinymmärtäminen, mutta tuloksen voi tarvittaessa koota useammalta kuin yhdeltä osallistumiskerralta (suullinen tiedonanto Maarit Mikkonen, Valvira 6.5.2014).

3.2 Toteutus

Ammatillisen arviointikokeilun pohjaksi tehtiin kansainvälistä tilannekartoitusta, jossa selvitettiin englannin- ja saksankielisten maiden sekä Pohjoismaiden ja joidenkin muita pienempiä kielialueita edustavien maiden arviointikäytänteitä terveydenhuollon ammattilaisten osalta. Direktiivimuutoksen myötä käytänteiden muutosta harkitaan paraikaa muuallakin kuin Suomessa. Lisäksi koottiin yhteen jo olemassa olevaa tietoa terveydenhuollon tehtävissä toimivien lääkäreiden ja hoitajien kielitaidon tarpeesta (Kela & Komppa 2011; Komppa ym. 2014; Mähönen 2014; Heimala-Kääriäinen 2015) sekä laajennettiin saatua kuvaa täydentävin tarveanalyysin. Tätä toteuttivat osin Jyväskylän yliopiston maisterivaiheen suomen kielen opiskelijat työelämän kielikysymyksiä käsittelevällä projektikurssilla.

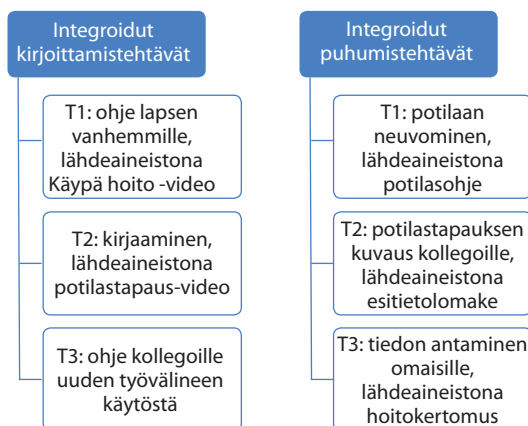
Ennen ammatillisen moduulin laadintaa analysoitiin myös terveydenhuollon koulutuksen saaneiden Yleisten kielitutkintojen suorituksia keskitason tutkinnossa ja havaittiin, että ammatillisuus näkyy taajaan myös näissä yleiskielitaidon arviointia varten laadituissa suorituksissa. Osallistujat hyödyntävät usein ammatillisen kielen resurssiaan myös muissa kuin teemaltaan terveyteen tai hyvinvointiin liittyvissä konteksteissa, mikä kertoo heidän kielitaitonsa kehittyneen pitkälti työympäristössä ja ammatillisen kielenkäytön olevan heidän taitoprofilinsa ydintä (Seilonen 2014; Seilonen & Suni 2015).

Ammatillisen kielitaidon arviointimoduulia varten laadittiin integroituja tehtäviä, jotka pyrkivät sisällölliseen ja funktionaaliseen autenttisuuteen siinä, että ne mukailevat työtehtävissä kohdattavia vuorovaikutustilanteita. Integroitujen tehtävien käyttö arvioinnissa on vielä verrattain harvinaista, joten jo tämä ratkaisu oli luonteel-

taan kokeileva. Tavoitteena oli, että reseptiiviset ja produktiiviset taidot yhdistyvät tehtävissä samaan tapaan kuin työtilanteissa, joissa viestintä usein perustuu johonkin kirjalliseen tai suulliseen materiaaliin: potilaan suullisesti antamat tiedot kirjataan tekstimuotoon, raportoidaan suullisesti eteenpäin tiedot, jotka aiemmin on kirjattu tai välitetään kasvokkaistilanteessa ohje tai tieto, joka itsellä on käytössä kirjallisessa muodossa. Voidaankin ajatella, että kyse on viimeaikaisen tekstitaitojattelu tuomisesta toisen kielen ammatillisen taidon arvioinnin kontekstiin: puhutaan siitä, mitä on luettu tai kirjoitetaan siitä, mitä on kuultu – ja modaaliteetin vaihtuessa vaihdetaan usein myös kielellistä rekisteriä.

Tehtävistö oli sairaanhoitajille ja lääkäreille samanlainen, mutta yksittäiset tehtävät olivat sillä tavoin joustavia, että ne sallivat käsiteltävänä olevan asian tarkastelun oman ammattikunnan näkökulmasta. Samoilla tehtävillä pyrittiin hahmottamaan ns. yhteistä terveydenhuollon ammatillisuutta eli yhteistä ydintä, joka kuuluisi useamman alalla toimivan ammattiryhmän toimintaan ja kielenkäyttöön. Ammatillisen kielitaidon moduulissa suoriutumisen vertailukohtana oli Yleisten kielitutkintojen tuore keskitason tutkinnon suoritus, jonka osa osallistujista oli tehnyt vuoden sisällä ennen arviointimoduulia ja osa teki arviointikokeiluun liittyvänä. Arviointikokeiluun oli alkuaan tarkoitus ottaa mukaan vain EU/ETA-alueen ulkopuolella kouluttautuneita ja ammattiryhmistä vain sairaanhoitajia ja lääkäreitä, joita opetus- ja kulttuuriministeriön selvitys (2014) ensisijaisesti koski, mutta asetelmaa täydennettiin EU-alueella koulutuksensa saaneilla ja myös muutamalla hammaslääkärillä, joille räätälöitiin samojen tehtävätyyppien mukaiset omat tehtävänsä. Näin saatiin mukaan vielä laajemmin ehdotettujen kielipoliittisten muutosten piiriin kuuluvien ryhmien näkökulmia ja samalla kokemusta myös tehtävien erikoisalakohtaisen varioinnin haasteista ja mahdollisuuksista.

Integroiduissa tehtävissä (kuvio 2) suullisesti suoritettavat tehtävät perustuivat kirjalliseen ja kirjallisesti suoritettavat tehtävät suulliseen aineistoon, joten tehtävissä suoriutuminen edellytti myös puheen ja tekstin ymmärtämisen taitoja. Pohja-aineisto oli neljässä tehtävässä kuudesta autenttista: Käypä hoito -ohjeesta kertova video maitoallergian diagnosoinnista, potilasohje EEG-tutkimuksesta, vuodeosaston esitietolomake ja kirjallinen hoitokertomus. Yhdessä tehtävässä aineistona oli simuloitu potilastapaus (potilaan oma suullinen anamneesi), ja yhdessä tehtävänä oli kuvata uuden työssä tarvittavan laitteen toimintaperiaate; laitteen sai valita itse ilman pohja- tai virikeaineistoa.



KUVIO 2. Integroidut tehtävät sairaanhoitajille ja lääkäreille ammatillisen kielitaidon arviointimuodulissa.

Tehtävissä tuli hallita eri tekstilajeja ja diskurssimalleja, kuten kirjaaminen, raportointi ja ohje, sekä terveydenhuollon alalle ominaisten kielellisten funktioiden ilmaisemista, kuten toimintaan kehottaminen ja ohjeen, neuvon ja tiedon antaminen sekä luottamuksen rakentaminen ja rauhoittelu (Moore, Morton, Hall, & Wallis 2015; ABIM 2015; Candlin & Roger 2013: 48). Viestintä tuli myös tehtävästä riippuen kohdentaa joko terveydenhuollon alaa vähemmän tunteville maallikoille – potilaalle ja omaisille – tai alan asiantuntijoille, mihin osaltaan liittyy se, että pohja-aineistona olevan tekstin tai puheen rekisteri tuli muuntaa ammatillisesta yleiseen tai päinvastoin. Kohdentamiseen ja rekisterin muuntamiseen liittyvät myös vastaanottajalle oleellisen ja riittävän tiedon valinta sekä sopivan terminologian ja esitystavan käyttö. Esimerkiksi toimintaan kehoittamisessa tämä tarkoittaa usein yleistävää tai epäsuoraa ilmaustapaa (*tutkimuspäivä saa syödä ja juoda - - pitää olla rentona* NSE3) ja omaisten informoinnissa ammattiterminologian korvaamista tai selittämistä arkikielisemmillä sananvalinnoilla (*siis hän saturoituu elikkä hän hengittää hyvin ilman lisähappee* NSE1; *iso hematooma oli* NSE3 vrt. *iso mustelma on ollut* NLB).¹

Arviointikokeilun toteutuksessa päädyttiin kokonaan tietokoneella suoritettavien tehtävistöjen laadintaan.² Tehtävänannot ja pohja-aineistot näkyivät näytöllä, ja osallistujat tallensivat suulliset vastauksensa tietokoneen omalla kameralla videomuotoon avi-tiedostoiksi ja kirjalliset suoritukset tekstinkäsittelyohjelmalla (MS Word). Asetelma oli haastava rakentaa, ja sen käyttöön liittyi väistämättä epävarmuustekijöitä,

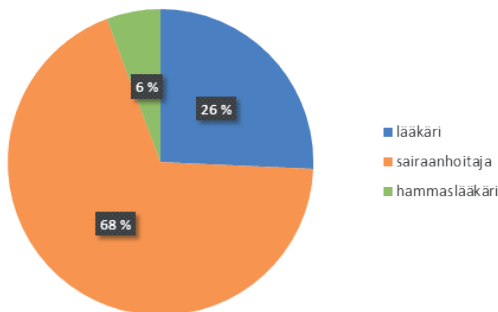
1 Aineisto ja tunnisteet esitellään luvussa 4.

2 Tutkimusavustaja Ella Taivassalolla oli tämän toteutuksessa ja aineiston käsittelyssä tärkeä rooli.

kun osallistujien tietoteknisistä valmiuksista ei ollut ennakkotietoa. Koska kyseessä oli kuitenkin nimenomaan tutkimuksellinen kokeilu, haluttiin sähköinen suoritustapa valita, jotta sen toimivuudesta saataisiin käytännön kokemusta esimerkiksi Yleisten kielitutkintojen kehitystyötäkin silmällä pitäen. Muutoin suoritukset tehtiin kannettavilla tietokoneilla, mutta yksi useiden osallistujien tilaisuus toteutettiin erään oppilaitoksen tietokonehuoneessa, jotta tätäkin vaihtoehtoa päästiin kokeilemaan. Riskit realistuivat siltä osin, että seitsemän osallistujan puhumissuorituksesta tallentui vain kuva ilman ääntä, ja valitettavasti vain kolmella heistä oli käytössä varatallentimena mp3-nauhuri. Neljän osallistujan suulliset suoritukset jäivät siten puuttumaan tutkimusaineistosta. Lisäksi muutaman osallistujan kirjallisten osuuksien suorittamisesta tuli tietokonehuoneessa teknisten ongelmien vuoksi katkonaista; tämä ratkaistiin antamalla heille lisäaikaa.

4 Arviointikokeilun osallistajat

Arviointikokeilun ammatillisen kielitaidon osuus toteutettiin usealla paikkakunnalla eri puolilla Suomea toukokuun ja joulukuun välisenä aikana vuonna 2015, ja siihen osallistui yhteensä 35 terveydenhuollon alalle ulkomailla koulutettanutta henkilöä, joiden äidinkieli oli jokin muu kuin suomi tai ruotsi. Heistä 24 oli sairaanhoitajia, 9 lääkäreitä ja 2 hammaslääkäreitä. Eri ammattikuntiin kuuluvien prosentuaaliset jakaumat ilmenevät kuvioista 3.



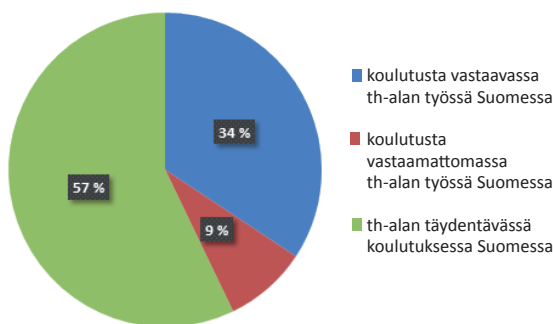
KUVIO 3. Osallistajat (n=35) ammattikunnittain.

Osallistajat olivat kotoisin yhteensä 16 eri maasta. Heistä 9 oli alkuaan koulutettanut terveydenhuollon ammattiin jossakin EU/ETA-maassa ja 26 muualla. Eri äidinkieliä osallistujajoukossa oli 11. Eniten oli mukana venäjänkielisiä (n=9), espanjankielisiä (n=7) sekä filippiinon- ja kiinankielisiä (n=4 molemmissa ryhmissä). Yksi osallistuja ilmoitti kaksi äidinkieltä. Kuudella oli suomenkielinen perheenjäsen. Ammattiin valmistumisesta oli 40%:lla (n=14) osallistujista kulunut alle kymmenen vuotta ja 60%:lla (n=21)

siis tätä kauemmin. Vastaavasti terveydenhuoltoalan työkokemusta oli 40%:lle (n=14) kertynyt jo lähtömaassaan vähintään viisi vuotta, mutta kolmannes (n=11) oli toisaalta muuttanut Suomeen heti ammattiin valmistumisen jälkeen eli ilman oman alansa työkokemusta. Suomeen muuttamisesta oli kahdella kolmesta (n=21) kulunut enintään viisi vuotta; yli 10 vuotta sitten muuttaneita oli joukossa vain neljä henkilöä eli 12%.

Sairaanhoidajista ne, jotka olivat toisessa EU-maassa kouluttautuneita (n=4), työskentelivät täysipäiväisesti jossakin suomalaisessa terveydenhuollon yksikössä. Heistä kolme oli sairaanhoitajana ja yksi lähihoitajana. Muut osallistuivat parhaillaan ammatilliseen täydennyskoulutukseen eivätkä siis olleet arviointikokeilun aikaan työelämässä (n=20). Heistäkin osa oli kuitenkin toiminut Suomessa esimerkiksi lähihoitajana tai henkilökohtaisena avustajana, ja useimmat olivat saaneet tuntumaa esimerkiksi lyhytaikaisessa harjoittelussa tai työkokeilussa terveydenhuollon alalla työskentelemiseen suomenkielisessä työympäristössä. Näistä täydennyskoulutuksessa olleista sairaanhoitajista yhtä lukuun ottamatta kaikki olivat saaneet ammatillisen koulutuksensa nykyisen EU/ETA-alueen ulkopuolella, ja täydennyskoulutuksen myötä he tavoittelivat nimenomaan laillistetun sairaanhoitajan asemaa Suomessa. Kaikki 11 lääkäriä ja hammaslääkäreitä olivat puolestaan työssäkäyviä, ja heistä neljä oli kouluttautunut ammatteensa EU/ETA-alueella ja päässyt heti Suomeen muutettuaan jatkamaan ammatissaan, kun taas seitsemän oli kouluttautunut tämän alueen ulkopuolella ja oli tutkimuksen toteutusajankohtana mukana laillistamisprosessissa tehden alansa töitä toisen ohjauksessa.

Työelämässä oli arviointikokeilun aikaan 43 % (n=15) osallistujista ja koulutuksessa loput. Työttömänä ei siis ollut kukaan. Tarkemmat osallistujien työllisyysasemaa koskevat tiedot on koottu kuvioon 4.

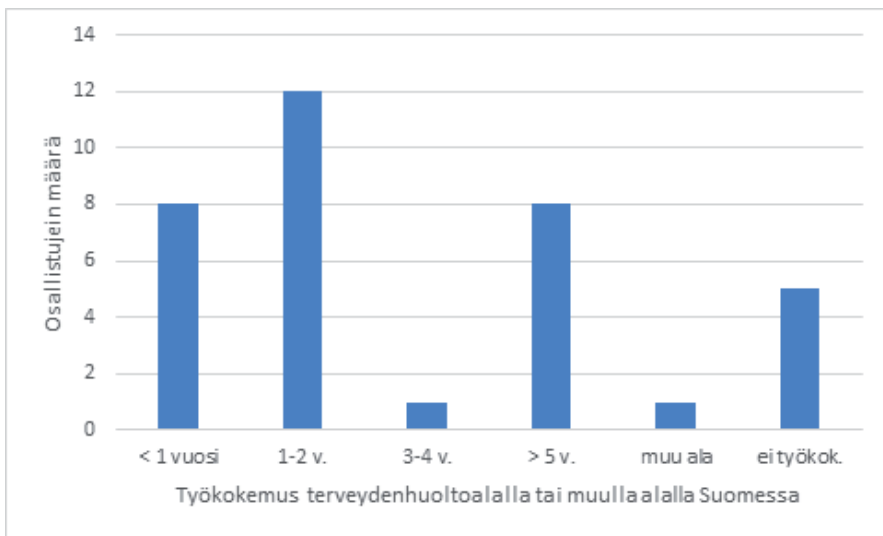


KUVIO 4. Osallistujien (n=35) työllisyysasema.

Kuviosta 4 ilmenee, että jokseenkin kolmannes osallistujista (34%, n=12) oli koulutustaan vastaavassa työssä Suomessa ja 9 % (n=3) teki terveydenhuoltoalan työtä jollakin

muulla nimikkeellä. Viimemainituissa oli käytännössä kyse useimmiten lähihoitajan tehtävissä työskentelevistä ulkomailla sairaanhoitajakoulutuksen suorittaneista henkilöistä. Yli puolet osallistujista (57%, n=20) oli vastikään aloittanut edellä mainitun sairaanhoitajataustan omaaville suunnatun täydentävän koulutuksen, jonka toivottiin edistävän heidän mahdollisuuksiaan saavuttaa sairaanhoitajan pätevyys myös Suomessa. Näitä räätälöityjä koulutuksia on luotu poistamaan sitä ongelmaa, että monet sairaanhoitajakoulutuksen EU/ETA-alueen ulkopuolella saaneista jäävät esimerkiksi lähihoitajan tehtäviin, joihin laillistusta ei vaadita, tai hakeutuvat suorittamaan kokonaista sairaanhoitajatutkintoa uudelleen Suomessa.

Suomessa osallistujat olivat työskennelleet oman alansa työtehtävissä kuviosta 5 ilmeneviä aikoja. Alle vuoden ja vähintään viisi vuotta Suomessa työskennelleitä oli joukossa yhtä paljon eli kahdeksan, ja 1–2 vuotta työskennelleitä oli 12. Osallistujista viisi ei ollut ollut vielä töissä Suomessa ja yhdellä oli vain toisen alan työkokemusta, joskin jo yli viiden vuoden ajalta.



KUVIO 5. Osallistujien (n=35) työkokemus terveydenhuollon alalla Suomessa (suomi työkielenä). Tapaukset, joissa työkokemusta Suomessa ei ole tai sitä on vain muulta alalta, on merkitty erikseen.

Vaikka päähuomiomme tässä artikkelissa on arviointikokeilun kokonaistulosten tarkastelussa eikä yksittäisten osallistujien suoriutumisessa, myös joitakin osallistujakohtaisia tietoja on tulosten esittelyn ja esimerkkien yhteydessä esillä. Osallistujien tunnisteen on muodostettu sukupuolta (N=nainen, M=mies), ammattia (S=sairanhoitaja, L=lääkäri, H=hammaslääkäri) ja kielitaustaa koskevista lyhenteistä (esim. V=venäjä, Vi=viro,

Vt=vietnam), ja samankieliset on erotettu toisistaan numeroin. Siten esimerkiksi NSV1 viittaa venäjänkieliseen naispuoliseen sairaanhoitajaan. Havaintoja eri osallistujien kielenkäytön piirteistä on raportoitu kvalitatiiviselta kannalta toisessa artikkelissa (Seilonen & Suni 2016a).

Palautteen keräämiseksi osallistujille tehtiin lyhyet (10–15 min) teemahaastattelut heti ammatillisen kielitaidon arviointimoduulin suorittamisen jälkeen. Haastattelukielenä oli suomi, ja teemana olivat osallistujan opinto- ja työhistoria sekä hänen kokemuksensa Yleisten kielitutkintojen keskitason tutkinnosta ja ammatillisen kielitaidon arviointimoduulista.

5 Arvioinnin toteutus ja analyysimenetelmät

Arviointikokeilussa tutkimukseen osallistuneet terveydenhuollon ammattilaiset suorittivat aluksi Yleisten kielitutkintojen keskitason tutkinnon ja pian tämän jälkeen, korkeintaan vuoden kuluttua, ammatillisen kielitaidon arviointimoduulin. Tässä alaluvussa esitellään näiden arviointitavat.

Yleisissä kielitutkinnoissa puhumisen ja kirjoittamisen suoritukset arvioidaan holistisesti arviointikriteereihin nojaten. Arvioinnin kohteena ovat esimerkiksi puhumisen ja kirjoittamisen sujuvuus, joustavuus, koherenssi ja ilmaisun ja rakenteiden tarkkuus (ks. Yleiset kielitutkinnot). Arvioinnin suorittavat tehtävään koulutetut Opetushallituksen arvostelijarekisteriin merkityt kielikoulutuksen ammattilaiset.

Ammatillisen arviointikokeilun suoritusten arviointi taas tehtiin kahden hankkeen tutkijan toimesta, ja se kohdistui terveydenhuollon alalle ominaisten kielellisten funktioiden ilmaisemiseen ja pragmaattisiin valintoihin, kuten tekstilajien ja diskurssimallien hallintaan, rekisterinvaihtelun hallintaan, ammatillisen sanaston ja fraseologian käyttöön sekä viestin kohdentamiseen ja impersonaalisen ilmaisutavan hallintaan. Ammatillisen moduulin suorituksia tarkasteltiin myös siltä kannalta, minkä tasoista pohjamateriaalin ymmärtämistä ne osoittivat.

Yleisten kielitutkintojen tulokset saatiin tutkimuskäyttöön sellaisinaan kielitaidon osa-alueittain. Niiden pohjalta muodostettiin tämän tutkimuksen käyttöön myös yleistaso siten, että keskiarvo pyöristettiin lähimpään kokonaislukuun; jos jonkin osion tasoarvio oli alle 3, se laskettiin tässä tasoarvioksi 2 (kyse on siis vain laskennallisesta ratkaisusta; tässä luku 2 ei vastaa Yleisten kielitutkintojen hyväksyttyä perustason suoritusta tasolla 2). Näin esimerkiksi sarja puhuminen 4, puheen ymmärtäminen 4, kirjoittaminen alle 3, tekstin ymmärtäminen 3 tuottaisi keskiarvon 3,25 ja pyöristettynä yleistason 3.

Ammatillisen kielitaidon arviointitavan kehittämisen pohjaksi tehtiin kvalitatiivisia analyyseja, jotka nojasivat alan tutkimuskirjallisuudessa ja tarveanalyysissa tunnistettuihin terveydenhuoltoalan kielenkäytön ja vuorovaikutuksen keskeisiin funktionaalisiin ja pragmaattisiin piirteisiin. Näiden analyysien tuloksia on raportoitu aiemmissa tieteellisissä julkaisuissa (Seilonen & Suni 2016a) ja esitelmissä (mm. Seilonen, Suni & Lammervo 2015; Seilonen & Suni 2016b).

Tehtävien arviointi kohdistettiin aiemman tutkimuskirjallisuuden ja tekemämme kvalitatiivisen tarkastelun pohjalta kolmeen ammatillisen kielitaidon funktionaaliseen ja pragmaattiseen ulottuvuuteen, joihin kuhunkin kiteytettiin toisiinsa liittyviä kielellisten funktioiden ilmaisemisen piirteitä. Näille rakennettiin viisiportainen kriteeristö, jossa tasolla 1 olevassa suorituksessa oli kyseisiä piirteitä niukasti, tasolla kolme jo täysin tunnistettavasti ja tasolla 5 jo luontevasti ja sujuvasti.

Näiden rinnalla arvioitiin vielä integroiduissa tehtävissä osoitettua pohjamateriaalin ymmärtämistä asteikolla 1–5, jossa taso 1 edusti heikkoa ja taso 5 vaivatonta ja tarkkaa ymmärtämisen tasoa. Ymmärtämistaitojen kohdalla olisi ollut kiinnostavaa hyödyntää Yleisten kielitutkintojen kriteeriviitteistä asteikkoa, mutta tätä rajoitti se, että tehtävien kielellinen vaatavuustaso oli abstraktin sisällön sekä kompleksisten kielellisten rakenteiden ja kielen eri rekistereiden läsnäolon (tässä yleiskielen ja yleispuhekielen ohella myös ammatillinen rekisteri) vuoksi kauttaaltaan vähintään ylempää keskitasoa (tasoa 4); yksinkertaisen kielellisen aineksen ymmärtämistä ei ollut käytetyn pohjamateriaalin äärellä mahdollista osoittaa, koska sellaista ei siihen sisältynyt.

Ammatillisen kielitaidon osalta tarkasteltiin ensinnäkin terveydenhuollon tekstilajien ja diskurssimallien hallintaa, mihin sisällytettiin myös tehtävänannon täyttyminen. Toiseksi arvioitiin eri rekisterien hallintaa ja taitoa siirtyä rekisteristä toiseen tilanteen ja vastaanottajan mukaan, mihin kytkeytyvät myös terveydenhuollon alan sanaston ja fraseologian käyttö. Kyse oli ammatillisen ja yleisen rekisterin välisen vaihtelun hallinnasta työelämän vuorovaikutustilanteissa. Kolmanneksi arvioitiin viestinnän kohdentamista erilaisille vastaanottajille ja siihen liittyvää epäsuoran ilmaustavan käyttöä. Ulottuvuudet ovat väistämättä osin limittäisiä: esimerkiksi rekisterin valintaan kytkeytyy myös kohdentamiseen liittyviä valintoja, joihin puolestaan kuuluu myös vastaanottajalle sopivan sanaston käyttö, ja edelleen tekstilajien ja diskurssimallien hallinta on mahdotonta ilman rekisterien variointia. Eri ulottuvuuksien erillinen kriteeripe rustainen arviointi ohjasi kuitenkin suuntaamaan huomion kattavammin eri ilmiöihin kuin suoraan holistisella otteella tehty arviointi olisi tehnyt. Suullisesti ja kirjallisesti suoritetuista tehtävistä muodostettiin omat kokonaisarvionsa, jotka olivat näiden kolmen ammatillisen kielitaidon ulottuvuuden ja pohjamateriaalin ymmärtämisestä annetun arvion keskiarvot. Arviointiasteikko on liitteessä 1. On syytä huomioida, että arviointiasteikko ei kata ns. syntyperäisen kielenpuhujan ammatillisen kielitaidon tasoa, koska siitä ei tässä arviointikokeilussa ollut näytteitä.

Funktionaalista oikeellisuutta tai sopivuutta (*functional adequacy*) pidetään keskeisenä ilmiönä toisen kielen taidossa (esim. Pallotti 2009). Kuikenin ja Vedderin mukaan (2016) toisen kielen kirjoittamistaitoa ei ole edes mahdollista täysin arvioida ilman funktionaalista näkökulmaa. He ovat käyttäneet funktionaalisen oikeellisuuden tarkastelussa arviointiasteikkoa, joka perustuu neljään dimensioon: sisältöön, tehtävän vaatimuksiin, ymmärrettävyyteen sekä koherenssiin ja koheesioon – siis hyvin samantyyppisiin ulottuvuuksiin kuin käyttämämme ammatillisen arviointikokeilun asteikkokin. Sisällön arvioinnissa tarkastellaan, onko tekstin sisältämä informaatioyksiköiden (esim. ajatusten/ideoiden ja käsitteiden) määrä lukijan kannalta sopiva ja relevantti. Tehtävän vaatimuksien täyttymistä arvioidaan sen perusteella, kuinka hyvin suorituksessa edustuvat tavoiteltu tekstilaji, rekisteri ja puheaktit. Ymmärrettävyyden arvioinnin perustana on se, kuinka paljon ponnistelua tekstin tavoitteen ja ajatusten ymmärtäminen vaatii lukijalta. Koherenssin ja koheesio arvioinnissa puolestaan kiinnitetään huomiota sidoskeinojen ja konjunktioiden käyttöön ja toisteisuuteen. (Mt.)

Arviointikokeilun toteuttajista kaksi kuuluu myös Yleisten kielitutkintojen suomen kielen arvostelijarekisteriin, ja he ovat harjaantuneita tämän tutkinnon arviointiasteikkojen käyttäjiä. Puhumisen ja kirjoittamisen arviointiasteikkojen tultua julkiseksi keväällä 2016 päädyttiin niitä hyödyntämään myös osana ammatillisen kielitaidon arviointia. Vaikka tehtävätyypit olivat hyvin erilaisia kuin Yleisissä kielitutkinnoissa ja pohjamateriaalit edustivat abstrakteja ammatillisia aihepiirejä, esimerkiksi sujuvuuden, tarkkuuden ja ilmaisun varioivuuden ilmenemistä eri kielellisten funktioiden suullisen ja kirjallisen ilmaisemisen yhteydessä päästiin havainnoimaan Yleisten kielitutkintojen arviointiasteikkojen valossa. Tämä oli kuitenkin vain varsinaista ammatillisen kielitaidon arviointia täydentävä näkökulma, jota ei tässä artikkelissa käsitellä lähemmin.

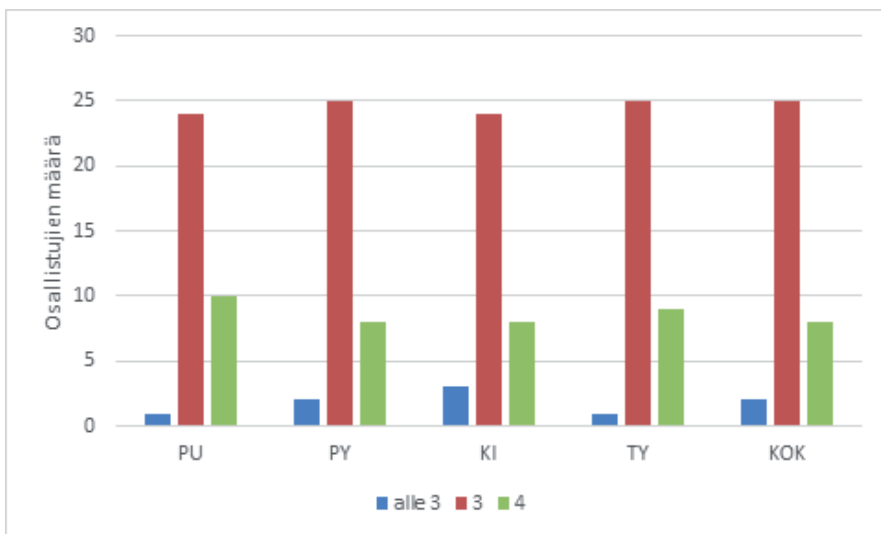
Kun arviointi oli suoritettu, tarkasteltiin tilastollisin menetelmin eri taustamuuttujien vaikutusta Yleisten kielitutkintojen suomen kielen keskitason tutkinnossa ja ammatillisen kielitaidon arviointimoduulissa suoriutumiseen sekä näissä kahdessa suoriutumisen yhteyttä. Analyysi painottui frekvenssijakaumiin ja ristiintaulukointiin. Tarkastellut taustamuuttujat olivat tässä vaiheessa maa (EU/ETA- tai ei-EU/ETA-maa), jossa terveydenhuollon koulutus (sairaanhoitaja, lääkäri tai hammaslääkäri) on saatu, Suomessa asumisen kesto, suomen kielen koulutus ja suomen kielellä saatu koulutus, terveydenhuoltoalan työkokemus Suomessa sekä se, onko osallistujalla joku suomenkielinen perheenjäsen vai ei. Osallistujien rekrytointia oli jatkettu, kunnes oli päästy tilastollisen tarkastelun alarajana pidettyyn 30 osallistujaan ja hieman sen yli. Monien muuttujien (esim. äidinkieli) osalta hajontaa oli kuitenkin niin paljon, ettei tilastollinen tarkastelu ollut mielekäästä. Myös se, että neljältä 35:stä osallistujasta ei ollut tallentunut suullisesti suoritettavaa osiota ja aineisto jäi heidän osaltaan aukkoiseksi, aiheutti omat rajoituksensa. Lisäksi on huomioitava, että osallistujajoukon pienuus vaikuttaa merkit-

tävästi tuloksista tehtävien johtopäätösten yleistettävyyteen, vaikka tilastollisen tarkastelun alaraja täyttyikin.

Tilastolliseen tarkasteluun perustuvia havaintoja peilattiin myös osallistujien omiin kokemuksiin ja havaintoihin, joita oli selvitetty palautehaastatteluissa. Näiden haastattelujen antia lähestyttiin sisällönanalyttisin menetelmin (Tuomi & Sarajärvi 2006). Tällaista kokemusten kartoittamista pidetään nykyään tärkeänä osana kielipoliittisiin kysymyksiin liittyvää tutkimus- ja kehittämistoimintaa, sillä perinteistä ylhäältä alas -politiikkaa ei globalisaation aikana pidetä riittävänä, vaan myös kielipoliittisten ratkaisujen kohteena olevien kannat tulee kuulla ja huomioida (Wright 2004).

6 Tulokset

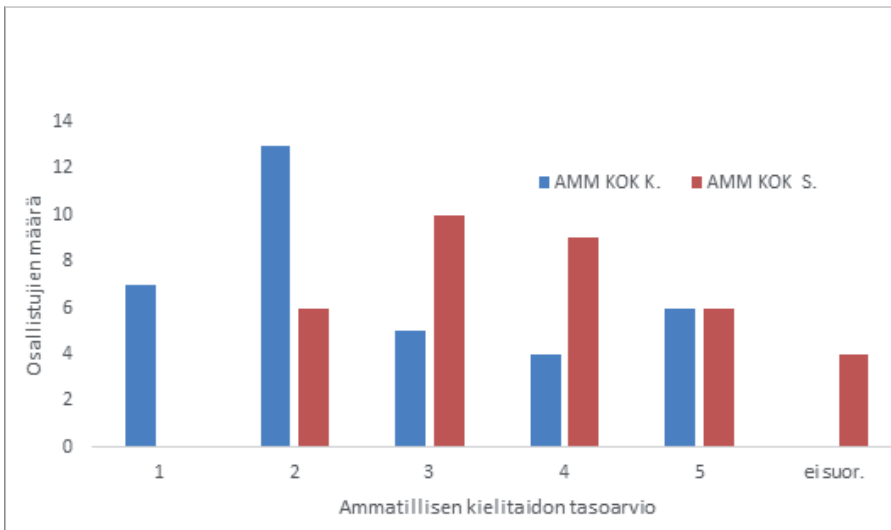
Tutkimuksen osallistujat suorittivat ensin Yleisen kielitutkinnon keskitasolla ja sen jälkeen ammatillisen kielitaidon arviointimoduulin. Kielitutkintosuorituksissa mikään osataito ei näyttänyt selvästi muita vahvempana tai heikompana, vaan jakaumat eri tasoarvioiden välillä olivat hyvin samanlaiset, kun osallistujajoukkoa tarkasteltiin kokonaisuutena. Ainoastaan puhumisessa tasoa 4 ja kirjoittamisessa taas tasoa alle 3 esiintyi hieman enemmän, mutta erot ovat näissäkin vähäisiä. Jakaumat on esitetty kuviossa 6.



KUVIO 6. Suoritusten jakaumat Yleisten kielitutkintojen (YKI) eri osakoikeissa. PU=puhuminen, PY=puheenymmärtäminen, KI=kirjoittaminen, TY=tekstinymmärtäminen ja KOK=kokonais- eli laskennallinen yleistasoarvio.

Tasaisuus näkyi myös yksilötasolla: peräti 18 osallistujaa oli saanut kaikista neljästä osakokeesta tasoarvion 3 ja 6 osallistujaa kaikista tasoarvion 4. Vain seitsemällä oli siis jokin muu kielitaitoprofiili kuin niin sanottu neljän suora. Jakauma poikkeaa varsin paljon Yleisten kielitutkintojen keskitason normaalista taitojakaumasta, sillä kirjoittamisen osakokeessa taitotason alle 3 saaneiden lukumäärä on viime vuosina ollut suomen kielessä noin 40%. Eroa selittää se, että arviointikokeilun osallistujajoukko oli valikoitunutta: heistä moni oli jo laillistamisprosessissa mukana, mikä on edellyttänyt Yleisten kielitutkintojen taitotasoa 3 kaikilla kielitaidon osa-alueilla.

Ammatillisen kielitaidon arviointimoduulissa sen sijaan tuli esiin selvä ero suullisesti ja kirjallisesti suoritettujen integroitujen tehtävien välillä. Kuvio 7 havainnollistaa sitä, että kirjallisissa suorituksissa painottuivat tässä aineistossa alemmat tasoarviot 1 ja 2, kun taas suullisissa suorituksissa oli vain vähän alle tason 3 jääviä tapauksia. On huomattava, että tässä on kyse integroiduista tehtävistä ja eri asteikosta kuin Yleisissä kielitutkinnoissa.



KUVIO 7. Ammatillinen kielitaito (1–5) kirjallisesti (K) ja suullisesti (S) suoritetuissa tehtävissä (n=35).

Alle tason 3 jääviä oli kirjallisista suorituksista 57% (n=20), mutta suullisista vain 19% (n=6). On kuitenkin huomioitava, että neljän osallistujan suullinen suoritus oli jäänyt tallentumatta. Yleisesti nimenomaan kirjoittamisen taidon on havaittu kehittyvän toisen kielen oppijoilla muita osataitoja hitaammin (Tarnanen, Härmälä & Neittaanmäki 2010), mutta tässä ryhmässä ero tuli näkyviin vain ammatillisesti suunnatuissa tehtävissä. Yksityiskohtaisempi tarkastelu osoitti ongelmien painottuneen lievästi arvioinnin

siihen osa-alueeseen, jossa huomion kohteena olivat ammatilliset tekstilajit ja tehtävän täyttyminen; monille tuotti vaikeuksia esimerkiksi allergisen lapsen vanhemmille suunnatun ohjeen laatiminen pelkän asioiden luettelemisen sijaan.

Erotuksena Yleisten kielitutkintojen kirjoittamistehtäviin nähden oli tehtävien ammatillisuuden tuoma sisältöjen ja samalla kieliasun vaatavuus: yksinkertaisella arkikielellä ei selviydytä, kun työssä kirjoittaen käsiteltävät aihepiirit ovat jo lähtökohtaisesti melko abstrakteja ja tilanteet tarkan ammatillisen informaation välittämiseen keskittyviä. Integroiduissa tehtävissä simuloitiin työarkea myös siten, että mukana olivat vielä ymmärtäminen sekä rekisterinvaihdos puhutusta kirjoitettuun. Kovin yksinkertaista kirjoitustehtävää alalta tuskin löytyisikään simuloitavaksi. Itse asiassa myös Tervola, Pajunen, Vainio, Honko ja Mattila (2015) ovat kiinnittäneet huomiota siihen, etteivät Yleisten kielitutkintojen tason 3 läpäisseet onnistu laillistamisprosessiin sisältyvissä yliopiston lääkärikuulusteluissa aina kirjoittamaan eheää ja ymmärrettävää tekstiä koevastauksiinsa. On huomattava, että koevastauskin on tekstilajina vaativa ja lisäksi työelämällekin vieras. Suoriutuminen Yleisissä kielitutkinnoissa tasolla 3 ei siis lupaa suoriutumista näissä tehtävissä.

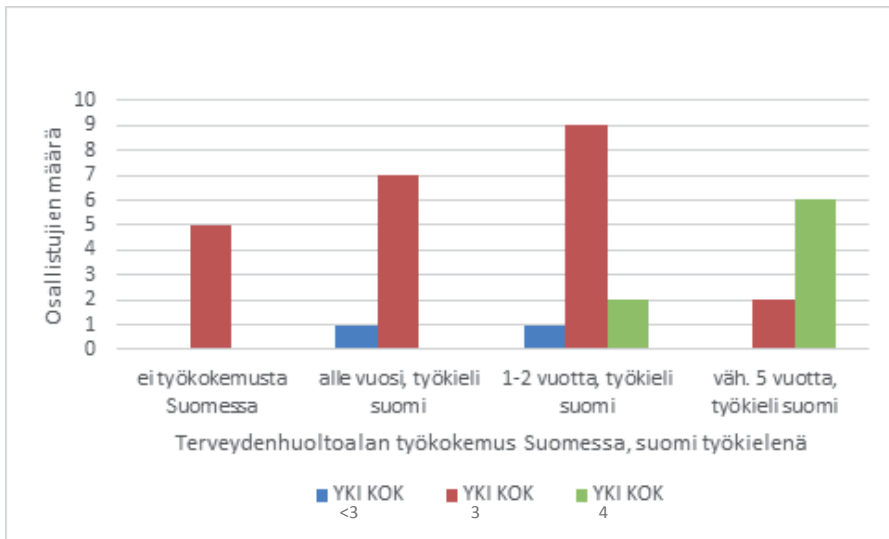
Muuttujien välisiä yhteyksiä tarkasteltiin ristiintaulukoinnin avulla, mutta koska aineisto oli pieni, keskityttiin vain solukohtaisiin jäännöstarkasteluihin. Kummassakaan arviointikokeilun osassa, Yleisten kielitutkintojen suomen kielen tutkinnossa tai ammatillisen kielitaidon arviointimoduulissa, ei osallistujan koulutusmaalla (EU/ETA-maa vai ei), suomea äidinkielenään puhuvilla perheenjäsenillä tai kielikoulutustaustalla ollut yhteyttä suoritusten tasoon.

Sen sijaan osallistujien menestyminen Yleisten kielitutkintojen keskitason tutkinnossa oli yhteydessä asumisaikaan Suomessa: Spearmanin järjestyskorrelaatiokerroin osoitti asumisajan Suomessa ja yleistaitotason välillä keskinkertaista korrelaatiota ($r_s=0,521^*$). Yhteys voitiin todentaa myös kaikissa osataidoissa erikseen (puhuminen $r_s=0,499^{**}$; puheen ymmärtäminen $r_s=0,521^{**}$), kirjoittaminen $r_s=0,420^*$, tekstin ymmärtäminen $r_s=0,552^{**}$); kirjoittamisen taito korreloi maassaoloon nähden heikommin kuin muut osataidot (vrt. Ahola & Hirvelä 2016).

Ammatillisen kielitaidon puolella taas suullisilla suorituksilla oli tilastollisesti merkitsevä yhteys asumisaikaan Suomessa ($r_s=0,369^*$), mutta kirjallisesti suoritetuilla ei ($r_s=0,240$). Kaiken kaikkiaan kirjallisesti suoritetuissa tehtävissä osoitettu ammatillinen taito korreloi Suomessa asumiseen vähemmän kuin suullisesti osoitettu ammatillinen kielitaito ja Yleisten kielitutkintojen tutkinnossa saadut tulokset.

Oman alan työkokemuksella Suomessa taas oli ristiintaulukointien valossa yhteys ammatillisen kielitaidon tasoon, vaikka korrelaatiokertoimien käyttö ei tältä osin ollut luotettavaa. Alle vuosi oman alan työkokemusta Suomessa oli yhteydessä ammatillisen kielitaidon suulliseen tasoon 3, 1–2 vuoden työkokemus tasoon 4 ja vähintään 5 vuoden kokemus tasoon 5.

Kuvioon 8 on koottu osallistujien laskennallinen yleistaso Yleisten kielitutkintojen suomen kielen tutkinnoissa sen mukaan, minkä pituinen terveydenhuoltoalan työkokemus heillä oli Suomessa suomeksi työskennellen, ja kuvioihin 9 ja 10 ammatillinen kielitaito samaan tapaan sekä suullisen että kirjallisen taidon osalta. Kuvioden tarkastelussa on huomattava, että arviointiasteikot ja -kriteerit eivät olleet samat kielitutkinnoissa ja ammatillisessa moduulissa, vaan arviointi kohdistui niissä eri asioihin. Ainoa vain muulla alalla työskennellyt ja ainoa 3–4 vuotta terveydenhuollon alalla työskennellyt on jätetty näistä kuvioista pois.

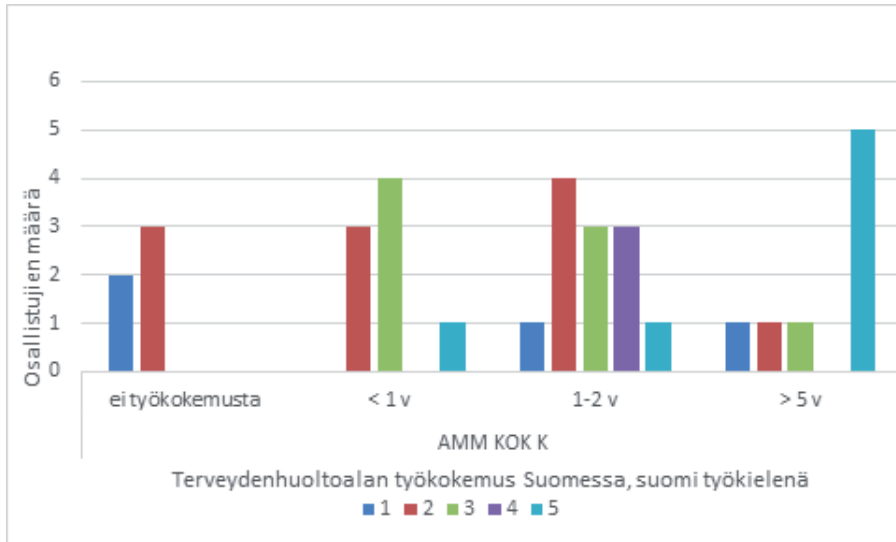


KUVIO 8. Suorituminen Yleisten kielitutkintojen suomen kielen tutkinnoissa ja terveydenhuoltoalan työkokemus Suomessa, suomi työkielenä (n=33).

Kuvio 8 havainnollistaa yleiskielitaidon tasoja suomenkielisessä työympäristössä hankitun työkokemuksen mukaan. Tutkimusasetelman kannalta on huomionarvoista, että Yleisten kielitutkintojen tasolle 4 yltäneitä oli 1–2 vuotta suomeksi työskennelleissä vasta 2, eivätkä kaikki viisi vuotta tai pidempään työskennelleetkään olleet yltäneet tälle tasolle; kahdeksasta osallistujasta kaksi oli tuolloin vielä Yleisten kielitutkintojen tasolla 3. Mikäli Yleisten kielitutkintojen taitotasoa 4 jatkossa edellytettäisiin osana lailistamisprosessia, siitä tulisi siis huomattava viive osalle niistäkin, joilla on mahdollisuus kehittää kielitaitoaan oman alansa työssä.

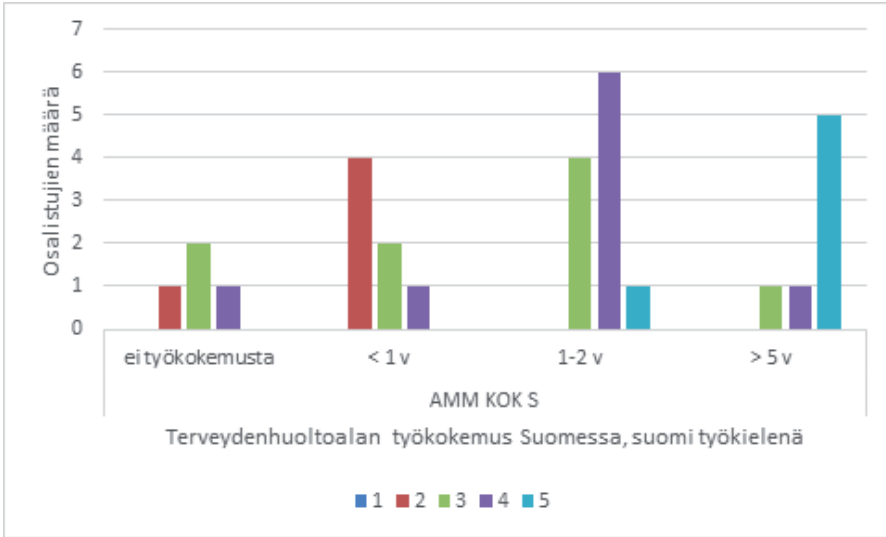
Kirjallisesti suoritettavissa ammatillisen kielitaidon tehtävissä havaitaan samankaltainen tilanne: 1–2 vuoden suomeksi työskentelyn jälkeen viiden osallistujan taidot olivat vasta tasoa 1 tai 2 ja kuusi ylsi tasolle 3 tai 4. Vielä viiden työvuoden jälkeenkään kaikki eivät olleet saavuttaneet tasoa 3 tai 4, vaikka useimmat osoittivatkin silloin jo

vahvaa tason 5 ammatillista osaamista tällä osa-alueella. Kirjallisen kielitaidon kehitys ei siis ole ammatillisissa yhteyksissä kovin nopeaa niilläkään, jotka ovat jo tiiviisti mukana työelämässä, ja kokonaan ilman suomenkielistä työkokemusta olevilla on puolestaan hyvin suuria haasteita tällä osa-alueella. Kuvio 9 havainnollistaa tätä kokonaistilannetta.



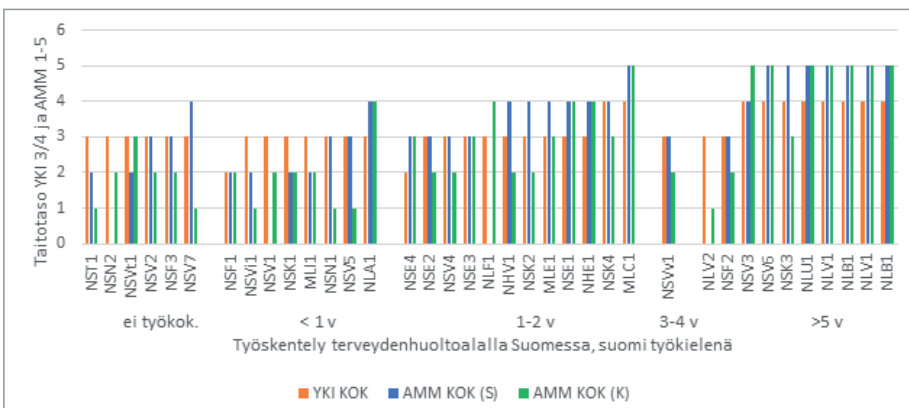
KUVIO 9. Suoriutuminen ammatillisen kielitaidon arviointimoduulin kirjallisesti suoritettavissa tehtävissä ja terveydenhuoltoalan työkokemus Suomessa (suomi työkielenä, n=29).

Kuvio 10 puolestaan osoittaa, että suullisesti suoritettavat ammatilliset tehtävät tuovat työkokemusta suomeksi omalla alallaan jo kartuttaneiden osallistujien kielitaidon vahvuudet esiin. Kaikki vähintään vuoden suomeksi työskennelleet ovat vähintään tasolla 3 ja vähemmänkin aikaa työskennelleistä seitsemästä osallistujastakin jo kolme. Viisi vuotta työskennelleistä lähes kaikki ovat puolestaan tasolla 5. Suullisissa tehtävissä kohtalainen sujuvuus ja luontevuus on siis ammatillisen kielitaidon osalta saavutettu yleensä 1–2 vuoden oman alan työkokemuksen myötä.



KUVIO 10. Suoriutuminen ammatillisen kielitaidon arviointimoduulin suullisesti suoritettavissa tehtävissä ja terveydenhuoltoalan työkokemus Suomessa (suomi työkielenä, n=29).

Kokonaiskuva tuloksista hahmottuu, kun kuvioissa 8–10 esitellyt asiat yhdistetään vielä samaan kuvioon siten, että yksittäiset osallistujat kielitaitoprofileineen ovat erotettavissa. Samalla voidaan löytää selityksiä tyypillisestä kehityskulusta poikkeaville tapauksille. Kuvioon 11 onkin vielä koottu kaikkien osallistujien suoriutuminen Yleisten kielitutkintojen suomen kielen tutkinnossa ja ammatillisen kielitaidon arviointimoduulissa. Osallistujat on järjestetty sen mukaan, kauanko he ovat työskennelleet terveydenhuoltoalan työtehtävissä Suomessa.



KUVIO 11. Osallistujien suoriutuminen Yleisten kielitutkintojen suomen kielen tutkinnossa (YKI KOK, yleistasoarvio 3 tai 4) sekä suullisesti (AMM KOK S) ja kirjallisesti (AMM KOK K) suoritettavissa ammatillisen kielitaidon arviointimoduulissa, jossa tasoarvot 1–5. Osallistujat on ryhmitelty Suomessa hankitun terveydenhuoltoalan työkokemuksen mukaisesti (suomi työkielenä).

Kuviosta ilmenee ensinnäkin se jo edellä todettu seikka, että Yleisten kielitutkintojen tasolla 3 ollaan tyypillisesti vielä siinä vaiheessa, kun työkokemusta on jo karttunut 1–2 vuotta; muutama on tällä taitotasolla pidemmänkin Suomessa-työskentelyn jälkeen. Samoin kertautuu se, että ei lainkaan tai alle vuoden suomeksi omalla alallaan työskennelleistä 14 osallistujasta viiden ammatillinen kielitaito ei vielä yltänyt tasolle 3 edes suullisesti. Sen sijaan vähintään vuoden työskennelleistä kaikki tavoittivat vähintään tämän tason ainakin suullisesti, ja heistä valtaosa ylsi tältä osin jo tasolle 4. Tälle havainnolle löytyy vastakaikua myös aiemmista terveydenhuollon tehtävissä toimiviin kohdistuneista tapaustutkimuksista, joissa noin puolentoista vuoden työskentelyn on raportoitu johtaneen sujuvan pärjäämisen kokemukseen ammatillisissa vuorovaikutustilanteissa (Mähönen 2014; Heimala-Kääriäinen 2015; Suni tulossa 2017).

Yleisten kielitutkintojen suomen kielen tutkinnossa tason 4 saavuttaneissa oli kaksi sellaista osallistujaa, jotka olivat työskennelleet Suomessa terveydenhuollon ammateissa vasta kahden vuoden ajan. Selvä enemmistö tämän yleiskielitaidon tason saavuttaneista oli kuitenkin kartuttanut alan työkokemusta Suomessa vähintään viiden vuoden ajan. Ammatillisen kielitaidon tasolle 5 suullisissa suorituksissa yltäneitä oli puolestaan vain niissä, joilla oli myös Yleisten kielitutkintojen taso 4. On lisäksi muistettava, että Suomessa asumisen kesto korreloi sekä Yleisten kielitutkintojen yleistason että suullisen ammatillisen kielitaidon kanssa.

1–2 vuotta suomeksi työskennelleiden kirjallisesti osoittama ammatillinen kielitaito asettui tasoille 2–4, kun se suullisella puolella oli heillä kaikilla vähintään tasoa 3. Kirjallisissa tehtävissä osoitettu ammatillinen taito on yleensäkin monien osallistujien kohdalla selvästi alempi kuin suullisesti osoitettu; ero on suurimmillaan osallistujalla NSV7, joka oman alansa työkokemuksen puuttumisesta huolimatta on osoittanut vahvaa (taso 4) ammatillista kielitaitoa suullisesti suoritettussa osiossa. Kyseessä on taustaltaan monikielinen henkilö, joka on asunut Suomessa seitsemän vuotta ja suorittanut terveydenhuoltoalan opintoja suomen kielellä vajaan vuoden. Ammatillista kielitaitoa lienee karttunut etenkin opintojen ja niihin todennäköisesti sisältyneiden työelämäjaksojen myötä; kirjallisissa tehtävissä hänelläkin on tosin vielä huomattavia haasteita.

Osallistajat NLA1 ja MLC1 erottuvat muista vahvalla ammatillisen kielitaidon profiilillaan. NLA1 on Yleisten kielitutkintojen yleistasoltaan 3, mutta hän saavuttanut alle kahdessa vuodessa tason 4 ammatillisen kielitaidon sekä suullisesti että kirjallisesti suoritettujen tehtävien valossa. MLC1 taas on samassa ajassa saavuttanut Yleisten kielitutkintojen yleistason 4 ja samalla ammatillisen kielitaidon tason 5 sekä suullisissa että kirjallisissa tehtävissä. Molempien hyvää suoriutumista ammatillisen kielitaidon moduulissa selittää osaltaan se, että he ovat suorittaneet Lääkäriksi Suomeen -koulutuksen, jonka aihepiirit kattavat esimerkiksi Suomen terveydenhuoltojärjestelmää, suomalaisia kansansairauksia, sosiaali- ja terveysturvaa sekä sairausvakuutusta ja oikeuslääketieteen, lääkehuollon ja reseptiopin teemoja sekä lääkärintodistuksiin ja -lau-

suntoihin, lääkärin vuorovaikutustaitoihin ja ammatilliseen suomen kieleen keskittyviä opintoja. Tämäntyyppisten koulutusten vaikuttavuudesta olisi tärkeää saada oma systemaattisesti kerätty tutkimusnäyttönsä, sillä nämä kaksi tapausta antavat jo viitteitä täsmäkoulutusten eduista.

Päinvastainen tapaus on NLV2, joka on työskennellyt Suomessa yli viiden vuoden ajan terveydenhuollon alan tehtävissä suomenkielisessä työympäristössä. Pitkästä työhistoriasta huolimatta hänen kielitutkintosuorituksensa oli tasoa 3 ja ammatillisen moduulin kirjallisissa tehtävissä hän jäi tasolle 1; viimeksimainitussa hänen suulliset suorituksensa eivät olleet tallentuneet. Tuloksia selittää hänen työnsä luonne, sillä hän ei ole ollut kliinisessä työssä vaan tutkimustehtävissä, joissa työkielenä on paljolti englanti suomen kielen rinnalla. Myös palautehaastattelussa hän käytti molempia kieliä ilmaislessaan havaintojaan ja kokemuksiaan.

7 Osallistujien kokemukset

Palautekeskustelussa arviointikokeiluun osallistuneilta kysyttiin mm. heidän ajatuksiaan siitä, kumpi suoritettavana olleista testeistä tuntui helpommalta. Lääkäreistä ja hammaslääkäreistä yhtä lukuun ottamatta kaikki (N=10) pitivät ammatillista arviointimoduulia helpompana kuin Yleisten kielitutkintojen suomen kielen keskitason tutkintoa; yksi osallistujista ei vastannut kysymykseen. He kaikki ovat työelämässä, ja yksi heistä kommentoikin asiaa seuraavasti: *ilman työkokemusta ei voisi tehdä testiä* (NLV1). Myös sairaanhoitajista seitsemän piti ammatillista moduulia helpompana. Yhteistä myös näille vastaajille on se, että he joko ovat parhaillaan työelämässä tai heillä on jo pitkäaikaista kokemusta terveydenhuollon työtehtävistä. He perustelivat kantaansa muun muassa sillä, että ammatillisen moduulin aihepiiri ja sanasto sekä tehtävyytöt olivat työstä jo tuttuja ja niihin tulee harjaannusta työympäristössä päivittäin. Lisäksi, kun aihepiiri on tuttu, myös vieraiden sanojen merkityksiä voi helpommin avata päätellessä.

Osallistuja MLC1 pohti ammatillisen kielitaidon arvioinnin ajankohtaa lääkäreiden näkökulmasta näin:

mun mielestä se olis tosi vaikea jos lääkäri on käyny (-) vain niinkun, yleiskielen kurssilla, eikä saanut kokemusta siitä koska mulle oli helppo koska mä oon jo, nin lukenu ja kirjoittanut mo- niinkun monta niitä tekstiä, mutta jos mä olisin tehny tämän niinkun vuosi sitten tai puoltoist vuotta sitten ei olisi onnistunut niin helposti. se on niinkun siinä mielessä niinkun aika pitkä aika ennen kun kehittää niinkun, niin niin sujuvasti.

MLC1 viittaa tässä edellä jo mainittuun runsaan vuoden työkokemuksen saamiseen eräänlaisena nivelkohtana, jossa tietty ammatillisen kielenkäytön sujuvuus saavute-

taan. Sitä aiemmin ammatillisen kielitaidon arviointimoduulin suorittaminen olisi ollut vaikeaa.

Kaikki niistä sairaanhoitajista, joilla ei ollut kokemusta terveydenhuollon alalla työskentelystä Suomessa ja jotka kommentoivat tätä teemaa (n=15), pitivät Yleisten kielitutkintojen keskitason tutkintoa itselleen helpompana ja ammatillista moduulia vaikeampana. Syyksi tähän monet heistä totesivat sen, että he eivät hallitse ammattialan sanastoa ja ilmaisutapoja, koska eivät ole olleet alan työssä Suomessa eivätkä siten ole saaneet mahdollisuutta oppia ammatillista kieltä: *en ollut Suomessa hoitoalalla sen takia en osanut vastata tilannemukaisesti (NSV3); ei työkokemusta Suomesta ja arkikieli ja tehtävien kieli erilaisia esim. tehtävä jossa pitää kertoa sukulaiselle (NSV6)*. Yleinen kielitutkinto taas oli monen mielestä helpompi siksi, että tutkinnon aiheet ja tehtävissä tarvittava kieli liittyvät arkipäivään ja ovat *elämässä lähempänä (NSVt)*. Helpoiksi tehtäviksi koettiin nimenomaan ne, joissa *ei tarvitse ymmärtää ammattikieltä - - vaikeat ne joissa piti käyttää ammattisanajoja ja puheita (NSV7)*. Kaksi sairaanhoitajista ei kommentoinut asiaa tai piti molempia yhtä vaikeina.

Ammatillisen kielitaidon arviointimoduuliin suhtauduttiin kaiken kaikkiaan hyvin myönteisesti. Tehtävät koettiin lähellä omaa ammattiarkea oleviksi, ja osa koki voitavansa siinä osoittaa nimenomaan sitä kielitaitoa, jota osaa parhaiten: *tätä harjoittelee joka päivä (NSE3); helpompi koska minun töissä, lähellä minun päivapäivä (NSE2)*. Tässä mielessä osallistuminen koettiin jopa voimauttavana. Täydennyskoulutuksessa olleista yksi toi myös esiin, että osallistuminen auttoi orientoitumaan siihen, mitä työelämässä on edessä: *minusta se oli kiinnostava lukea te(k)stiä, joka myöhemmin tulee työelämässäni (NSN2)*. Toiselle ammatillisen kielitaidon arviointimoduuli konkretisoi sitä, mitä hänen juuri aloittamassaan täydennyskoulutuksessa tulee ehkä esiin: *testi on hyvä pohja kun voi tietää mitä odottaa kurssilla (NSF3)*.

Kriittiset huomiot koskivat lähinnä ajankäyttöä: osassa tehtäviä useampi pohjaaineistona olleen videon katsomiskerta tai pidempi kirjoitusaika olisi ollut muutaman osallistujan mielestä tarpeen. Tietokonealueissa suorituksia tehneistä osa myös kohdasi teknisiä ongelmia, jotka välillä veivät huomiota pois itse suorituksesta. Tarjotun lisäajan myötä testitilaisuus kesti heillä suunniteltua pidempään. Yksi osallistuja esitti perustellun toiveen, että testitilanne sisältäisi myös kaksisuuntaista potilasvuorovaikutusta eli simuloitun tilanteen potilaan roolissa olevan henkilön kanssa. Kliinisen osaamisen arvioinnissa tämä onkin melko yleisesti käytetty ratkaisu, mutta vain kielitaidon arviointiin valjastettavaksi hyvin kallis toteuttaa.

8 Päätäntö

Terveydenhuollon suomi -hankkeen tulokset osoittavat, että osallistujien suoriutuminen ammatillisen kielitaidon moduulissa on yhteydessä siihen, miten pitkään he ovat kartuttaneet oman alansa työkokemusta suomenkielisessä ympäristössä. Myös yhteys Suomessa-asumisen kestoon on kuitenkin olemassa; sen merkittävyys näkyy muilla osa-alueilla kuin kirjallisesti suoritetuissa ammatillisen kielitaidon tehtävissä.

Niillä, jotka eivät ole päässeet oman alansa työtehtäviin Suomessa tai joilla on niistä kokemusta vasta alle vuoden verran, ei ammatillinen kielitaito ole yleensä pääsyt kehittymään kovin pitkälle. Ammatillisen kielitaidon kehityksessä voidaan kuitenkin tunnistaa merkittävää kehitystä jo 1–2 vuoden työssäolon myötä: kaikki tässä työuran vaiheessa olleet olivat saavuttaneet vähintään Yleisten kielitutkintojen tason 3 ja samoin jo kohtalaisena pidettävän tason 3 ammatillisessa kielitaidossa suullisissa tehtävissä ja kirjallisissakin vähintään tason 2. Siihen, että ammatillisen kielitaidon taso 3 olisi toimiva kynnystasona, viittaavat tällä tasolla olevien osallistujien omat kokemukset työssä ja ammatillisen moduulin tehtävissä pärjäämisestään. Yli viisi vuotta omalla alallaan Suomessa työskennelleistä taas useimmat saavuttivat Yleisten kielitutkintojen tason 4, ja vain tämän tason saavuttaneissa oli myös ammatillisessa kielitaidossa jo korkealle tasolle 5 yltäneitä.

Jos näitä havaintoja suhteutetaan opetus- ja kulttuuriministeriön esityksiin siitä, miten laillistamisprosessien yhteydessä tehtävää kielitaidon arviointia tulisi kehittää, on tärkeää kiinnittää huomiota taitojen kehittymiseen kuluvaan aikaan. Yleiskielitaidoltaan tasolla 3 olevat kokivat pärjäävänsä terveydenhuoltoalan työtehtävissään kohtalaisen sujuvasti jo siinä vaiheessa, kun heillä oli työkokemusta 1–2 vuotta. Heidän ammatillinen kielitaitonsa ei yleensä ollut vielä vahva, mutta se oli kuitenkin vähintään asteikon tasoa 3 suullisesti ja tasoa 2 kirjallisesti.

Yleiskielitaidon kehittyminen tasolta toiselle vie vuosia, ja kielitaidon kehittyminen myös korreloi paljolti asumisajan kanssa. Tämän tutkimuksen osallistujista Yleisten kielitutkintojen tasolla 4 olivat tyypillisesti ne, jotka olivat olleet Suomessa töissä ainakin viisi vuotta. Vain kahdella työkokemusta oli tätä vähemmän, ja heistäkin toinen oli itse asiassa asunut maassa jo seitsemän vuotta. Yleiskielitaidon kriteerin kiristäminen tasolta 3 tasolle 4 hidastaisi laillistamisprosessiin pääsyä siis oleellisesti – etenkin, jos tämä taso vaadittaisiin jo prosessin alussa. Ajankohdan tulisi olla selvästi myöhemmässä vaiheessa, ja silloinkin on tarkkaan harkittava, onko mielekästä vaatia myös kirjoittamisen taidon olevan tasolla 4. Kyseinen osataito kehittyy tämänkin tutkimuksen tulosten mukaan vielä hitaammin. Kirjoittamisen taito on työssä tärkeä, mutta siihen on toisaalta mahdollista saada myös apua työyhteisöissä – ja ainakin tapaustutkimusten valossa juuri siinä sitä pisimpään myös tarvitaan ja käytetään (ks. Mähönen 2014; Heimala-Kääriäinen 2015).

Ammatillisen kielitaidon arviointi täydentää Yleisten kielitutkintojen antamaa kuvaa osallistujien kielitaidosta ennen kaikkea laadullisesti: se auttaa inventoimaan sitä, miten osallistujat selviytyvät tyypillisissä työelämän viestintätilanteissa, kuten potilaan ohjeistamisessa ja hänen tietojensa kirjaamisessa tai omaisten tai kollegojen informoinnissa. Sen avulla pystytään siis tekemään näkyväksi nimenomaan sellaista työelämässä tarvittavaa tehtäväkeskeistä kielitaitoa, jonka yleiskielitaidon arviointi ohittaa ja jonka nimenomaan oman alansa työelämässä vähintään vuoden Suomessa olleista kokevat kielitaitonsa vahvimmaksi puoleksi. Hamp-Lyonsin (1991) esiin nostama arvioinnin reiluuden kokemus siis toteutuu heidän kohdallaan: he pääsevät näyttämään osaamisestaan puolia, joita yleiskielitaidon arviointi ei niinkään tavoita.

Yksi ammatillisen kielitaidon arvioinnin moduulin käyttötapa olisi täydentää sillä yksilön taitoprofilia erityisesti tilanteissa, joissa esimerkiksi kirjoittamisen taito jää Yleisissä kielitutkinnoissa vielä alle asetetun vaatimustason eikä pääsyä laillistamisprosessiin siten nykyisellään ole, vaikka työkokemusta hoito- tai hoiva-alalta jo olisi karttunut ja ammatillista kielitaitoa sen mukana. Potilasturvallisuuden näkökulmasta näyttö ammatillisen kielitaidon kohtalaisesta hallinnasta olisi todennäköisesti informatiivisempi kuin kirjoittamisen taitotaso 3 Yleisissä kielitutkinnoissa.

Ammatillisen kielitaidon arvioinnin mielekkyys voidaan kuitenkin tulosten valossa myös kyseenalaistaa. Ainakin tällaisessa melko suppeassa arviointikokeiluaineistossa yhteys työssäolon ja ammatillisen kielitaidon välillä oli niin ilmeinen, että periaatteessa voidaan olettaa riittävän pitkäkestoisen työkokemuksen saamisen johtavan ammatillisen kielitaidon kehitykseen: ainakin suullisesti suoritettavissa integroiduissa tehtävissä tavoitetaan 1–2 vuodessa taso 3 ja noin viidessä vuodessa yleisesti jo taso 5. On siis kriittisesti pohdittava, onko tarkoituksenmukaista panostaa valtakunnallisen ammatillisen kielitaidon testausjärjestelmän rakentamiseen ja ylläpitoon, jos tavoiteltava tulos todennäköisesti saavutetaan yksinkertaisesti varmistamalla jokaiselle laillistamisprosessissa olevalle edellytykset kehittää ammatillista kielitaitoa olemalla mukana suomenkielisessä työelämässä ainakin noin vuoden tai kahden vuoden ajan. Tämä on siis tilanne, jos tämän tutkimuksen asteikon tasoa 3 ajatellaan melko itsenäiseen ammatilliseen suoriutumiseen riittävänä kynnystasona. Kuten todettua, kirjoittamisen osalta tuen tarvetta esiintyy selkeästi tätä pidempään.

Tutkimus antaa myös viitteitä siitä, että ammattialakohtaista räätälöityä ammatillisäältä ja alakohtaista kielikoulutusta yhdistävät intensiivikoulutukset, kuten Lääkäriksi Suomeen -koulutus, ovat tuottaneet ainakin yksilötasolla hyviä tuloksia. Paraikaa käynnissä olevan alan koulutuspolkujen valtakunnallisen uudistamisen ja niiden pilotoinnin yhteydessä kannattaisikin tehdä systemaattista tutkimustyötä koulutuksen vaikuttavuudesta niiden ensisijaisen kohderyhmän eli EU/ETA-alueen ulkopuolelta tulevien parissa.

Osallistujien kokemukset arvioinnista jakautuivat hyvin selkeästi sen mukaan, oliko heillä alansa työelämäkokemusta Suomesta vai ei. Alan työssä olevat ja työkokemusta Suomessa jo kartuttaneet kokivat ammatillisen kielitaidon moduulin itselleen sopivaksi ja Yleisten kielitutkintojen keskitason tutkintoa helpommaksi, kun taas muut pitivät sitä helpompana ja ammatillisen kielitaidon moduulia vaikeana nimenomaan siksi, että heillä ei vielä ole tuntumaa siinä tarvittaviin ilmauksiin ja kielellisiin käytänteisiin. Lääkäri NLB tiivisti tämän eron palautehaastattelussa seuraavasti:

mä luulin että että tuo testi jos ois tehty heti kun olin tullut ois tullut, vaikea, koska tää on myös, näyttää se su- suomalainen tapa työskennellä, ja se on se semmonen kun on ihan sekin vieras ja vielä vieras kieli tulee ehkä haastavampi, se ehkä alussa on semmonen YKI-(tentin) ihan alussa alussa ehkä, semmonen yleisen elämään liittyvät asiat ehkä on semmosia, läheisempiä

Ammatillisen kielitaidon edellyttäminen niiltä, joilla ei ole ollut pääsyä työelämän todellisiin vuorovaikutustilanteisiin, ei siis ole tutkimuksen tuottamien havaintojen valossa reilua eikä realistista. Alalle luotavissa koulutusmalleissakin on varmistettava, että osallistujille turvataan pääsy mukaan työyhteisöihin, koska ammatillisen kielitaidon osa-alueita on vaikeaa tai jopa mahdotonta ottaa haltuun pelkästään luokkahuoneopetukseen osallistuen. Osallistujista MLC1 tiivisti kantansa tällä tavoin: *mun mielestä ammattikieli kehittyy niinkun vain töissä, täs tää ei saa, ei pääse niinkun oppimaan, kursilla*. Hän oli osallistujien joukossa kaikkein nopeimmin edistyneitä, ja työkokemuksen lisäksi hän oli ollut myös ammattikunnalleen suunnatussa täsmäkoulutuksessa.

E erityisen kiinnostavaa olisi tehdä ammatillista kielitaidon arviointia dynaamisen arvioinnin periaattein eli pyrkien ennustamaan ja samalla tukemaan tulevaa kehitystä sen sijaan, että keskitytään vain tiettyyn hetkeen mennessä haltuun otetun kielitaidon erittelyyn numeerisen arvion tuottamiseksi (ks. Poehner 2008). Hankkeessa koottu rikas aineisto näyttäisi tarjoavan mahdollisuuksia tunnistaa tyypillisiä kehityskulkuja ja kehittää niiden pohjalta myös alan kielikoulutusta palvelevia pedagogisia työkaluja. Tämä onkin yksi suunta, johon voisi edetä. Ammatillisen kielitaidon dynaamista arviointia kannattaisi todennäköisesti tehdä osana koulutusta nimenomaan siinä vaiheessa, kun työelämäkokemusta on jo alkanut karttua. Tällöin olisi tarkoituksenmukaista ohjata osallistujien ja myös heidän ohjaajiensa huomio ammatillisessa kielitaidossa kehittymässä oleviin puoliin sekä tapoihin, joilla kehittää juuri näitä puolia erilaisissa työtilanteissa ja työyhteisön tuella.

Kirjallisuus

- ABIM 2015. *American Board of Internal Medicine* [online] [luettu 27.08.2016] Saatavissa: www.abim.org.
- Ahola, S. & T. Hirvelä 2016. Mikä merkitys osallistujan taustatekijöillä on menestymiseen kielitestissä? *Kieli, koulutus ja yhteiskunta*, syyskuu 2016. <http://www.kieliverkosto.fi/article/mika-merkitys-osallistujan-taustatekijoilla-on-menestymiseen-kielitestissa/>.
- Ailasmaa, R. 2015a. *Terveys- ja hyvinvointipalvelujen henkilöstö 2013*. Tilastoraportti 26/2015. Helsinki: Terveiden ja hyvinvoinnin laitos.
- Ailasmaa, R. 2015b. *Terveys- ja hyvinvointipalvelujen henkilöstön liikkuvuus 2012*. Tilastoraportti 12/2015. Helsinki: Terveiden ja hyvinvoinnin laitos.
- Atkins, S., C. Roberts, K. Hawthorne & T. Greenhalgh 2016. Simulated consultations: a sociolinguistic perspective. *BMC Medical Education*, 16 (16). DOI: 10.1186/s12909-016-0555-2.
- Candlin, S. & P. Roger 2013. *Communication and professional relationships in healthcare practice*. Sheffield: Equinox.
- Davies, A. 2001. The logic of LSP. *Language Testing*, 18 (2), 133–149.
- De La Croix, A. & J. Skelton 2009. The reality of role-play: interruptions and amount of talk in simulated consultations. *Medical Education*, 43 (7), 695–703. DOI: 10.1111/j.1365-2923.2009.03392.x.
- Eggenberger, E., K. Heimerl & M. Bennett 2012. Communication skills training in dementia care: a systematic review of effectiveness, training content, and didactic methods in different care settings. *International Psychogeriatrics*, 23 (3), 345–358. DOI: <http://dx.doi.org/10.1017/S1041610212001664>.
- Elder, C. 2016. Exploring the limits of authenticity in LSP testing: the case of a specific-purpose language test for health professionals. *Language Testing*, 33 (2), 147–152. DOI: 10.1177/0265532215607397.
- Elder, C. & T. McNamara 2016. The hunt for “indigenous criteria” in assessing communication in the physiotherapy workplace. *Language Testing*, 33 (2), 153–174. DOI: 10.1177/0265532215607397.
- Elder, C., J. Pill, R. Woodward-Kron, T. McNamara, E. Manias, G. Webb & G. McColl 2012. Health professionals’ views of communication: Implications for assessing performance on a health-specific English language test. *TESOL Quarterly*, 46 (2), 409–419.
- Hamp-Lyons, L. 1991. Pre-text: Task-related influences on the writer. Teoksessa L. Hamp-Lyons (toim.) *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex, 127–151.
- Heimala-Kääriäinen, E. 2015. *Seurassa parempaa sairaalasuomea: ammatillisen kielitaidon oppiminen ja kollegojen kielellinen tuki sairaalaympäristössä*. Maisterintutkimus. Jyväskylän yliopiston kielten laitos. Saatavissa: <http://urn.fi/URN:NBN:fi:ju-201506192388>.
- Hoekje, B. 2016. “Language,” “communication,” and the longing for the authentic in LSP testing. *Language Testing*, 33 (2), 289–299. DOI: 10.1177/0265532215607397.
- HY+ 2014. Valtakunnallinen koulutusmalli, ETA-alueen ulkopuolella lääkärin tutkinnon suorittaneiden modulaarinen valmennusohjelma. [luettu 30.11.2015]. Saatavissa: <http://hyplus.helsinki.fi/hanke/valtakunnallinen-koulutusmalli-eta-alueen-ulkopuolella-laakarintutkinnon-suorittaneiden-modulaarinen-valmennusohjelma/>.
- Härmälä, M. 2008. *Riittääkö Ett ögonblick näytöksi merkonomilta edellytetystä kielitaidosta? Kielitaidon arviointi aikuisten näyttötutkinnoissa*. Jyväskylä Studies in Humanities 101. Jyväskylä: Jyväskylän yliopisto. Saatavissa: <http://urn.fi/URN:ISBN:978-951-39-3340-1>.

- Jacoby, S. 1998. *Science as performance: socializing scientific discourse through the conference talk rehearsal*. Unpublished doctoral dissertation. Los Angeles: University of California.
- Jacoby, S. & E. Ochs 1995. Co-construction: an introduction. *Research on Language and Social Interaction*, 28 (3), 171–183.
- Kela, M. & J. Komppa 2011. Sairaanhoidajan työkieli: yleiskieltä vai ammattikieltä? Funktionaalinen näkökulma ammattikielen oppimiseen toisena kielenä. *Puhe ja kieli*, 31 (4), 173–192. <http://ojs.tsv.fi/index.php/pk/article/view/4752>.
- Komppa, J., T. Jäppinen, M. Herva & T. Härmäläinen 2014. *Korkeakoulutuksen ammatilliset suomi toisena kielenä -viitekehykset*. Metropolia Ammattikorkeakoulun julkaisusarja Aatos-artikkelit 16. Helsinki: Metropolia Ammattikorkeakoulu. Saatavissa: http://www.metropolia.fi/fileadmin/user_upload/Julkaisuutoiminta/Julkaisusarjat/AATOS/PDF/METROPOLIA_AATOS_16.pdf.
- Kuiken, F. & I. Vedder 2016. Functional adequacy in L2 writing: towards a new rating scale. *Language Testing*, 33 (4), 1–16. DOI: 10.1177/0265532216663991.
- Lääkärit 2016. Suomen Lääkäriliitto. <http://www.laakariliitto.fi/tilastot/> [luettu 13.9.2016].
- Macqueen, S., J. Pill & U. Knoch 2016. Language test as a boundary object: perspectives from test users in the healthcare domain. *Language Testing*, 33 (2), 271–288. DOI: 10.1177/0265532215607397.
- McNamara, T. 1996. *Measuring second language performance*. London: Longman.
- McNamara, T. 1997. Problematizing content validity: the Occupational English Test as a measure of medical communication. *Melbourne Papers in Language Testing*, 6 (1), 19–43.
- Moore, T., J. Morton, D. Hall & C. Wallis 2015. *Literacy practices in the professional workplace: Implications for the IELTS reading and writing tests*. IELTS Research Reports Online Series, 1. London: British Council.
- Mähönen, E. 2014. *Maahanmuuttajalääkärin kokemuksia ammatillisen kielitaidon kehittämisestä: "No se mun kurssi oli töissä"*. Maisterintutkielma. Jyväskylän yliopiston kielten laitos. Saatavissa: <http://urn.fi/URN:NBN:fi:juu-201501201153>.
- Opetus- ja kulttuuriministeriö 2014. *Kielitaidon määrittäminen sekä kielitaidon ja EU/ETA-alueen ulkopuolella hankitun koulutuksen täydentäminen terveysalalla*. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2014: 5. Helsinki: Opetus- ja kulttuuriministeriö. Saatavissa: http://www.minedu.fi/OPM/Julkaisut/2014/kielitaito_terveysalalla.html.
- Paananen, J. 2015. Kuinka eleet helpottavat yhteisymmärrystä? Ikoniset ja deiktiset eleet monikulttuurisilla lääkärin vastaanotoilla. *Puhe ja kieli*, 35 (2), 73–95.
- Pallotti, G. 2009. CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, 30 (4), 590–601.
- Philip, S., E. Manias & R. Woodward 2015. Nursing educator perspective of overseas qualified nurses' intercultural clinical communication: barriers, enablers, and engagement strategies. *Journal of Clinical Nursing*, 24 (17–18), 2628–2637. DOI: 10.1111/jonc.12879.
- Piikki, A. 2010. Terveystieteiden alalla riittävä suomen taito – mitä se on? *Sutina*, 1, 30–33.
- Pill, J. 2016. Drawing on indigenous criteria for more authentic assessment in a specific-purpose language test: Health professionals interacting with patients. *Language Testing*, 33 (2), 175–193. DOI: 10.1177/0265532215607400.
- Poehner, M. E. 2008. *Dynamic assessment: a Vygotskian approach to understanding and promoting second language development*. Berlin: Springer Publishing.
- Raunio, P. 2013. *Työhön Suomeen? Tutkimus työperusteiseen maahanmuuttoon liittyvistä koulutusprosesseista*. Acta Electronica Universitatis Tampereensis 1331. Tampere: Tampereen yliopisto. Saatavissa: <http://urn.fi/URN:ISBN:978-951-44-9209-9>.
- Seilonen, M. 2014. "Tuntuuko, jos tästä painaa?" Epäsuoruus, toimintaan kehottaminen ja ammatillisuus. *Lähivertailuja – Lähivördlusi*, 24, 221–243. DOI:10.5128/LV24.08.

- Seilonen, M. & M. Suni 2015. Puhetta työstä: kansainvälisten terveydenhuollon ammattilaisten suomen kielen puhumistaidon arviointi. *Ammattikasvatuksen aikakauskirja*, 17 (2), 27–43.
- Seilonen, M. & M. Suni 2016a. Ohjeita, tietoa ja turvaa kielen keinoin. Ulkomailta palkatut sairaanhoitajat ammatillista suomen kielen taitoaan osoittamassa. *LähiVertailuja – Lähivördlusi*, 26, 450–480. DOI: <http://dx.doi.org/10.5128/LV26.15>.
- Seilonen, M. & M. Suni 2016b. *Assessing professional Finnish language skills in health care*. 5th International Conference Applied Linguistics and Professional Practice (ALAPP) 2015. Esitelmä. 5.–7.11.2015. Milano.
- Seilonen, M., M. Suni & T. Lammervo 2015. *Health care Finnish: developing and assessing Finnish proficiency among health care professionals*. Language Testing Research Colloquium (LTRC). Poster. 18.–20.3.2015. Toronto.
- Strömmer, M. 2016. Affordances and constraints: second language learning in cleaning work. *Multilingua: Journal of Cross-Cultural and Interlanguage Communication*, 35 (6), 697–721. <http://dx.doi.org/10.1515/multi-2014-0113>. DOI:10.1515/multi-2014-0113.
- Suni, M. (tulossa 2017). Working and learning in a new niche: ecological interpretations of work-related immigration. Teoksessa J. Angori, M. Marra & J. Holmes (toim.) *Negotiating boundaries at work*. Edinburgh: Edinburgh University Press.
- Tarnanen, M., M. Härmälä & R. Neittaanmäki 2010. Aikuisten kielitaito ja Yleiset kielitutkinnot. Teoksessa M. Garant & M. Kinnunen (toim.) *AFinLA-e: Soveltavan kielitieteen tutkimuksia 2*. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 59–71. Saatavissa: <http://journal.fi/afinla/article/view/3876>.
- Tervola, M., A. Pajunen, S. Vainio, M. Honko & K. Mattila 2015. Maahanmuuttajataustaisten lääkäreiden suomen kielen taito laillistamiskäytännössä. *Duodecim*, 131 (4), 339–346. <http://urn.fi/URN:NBN:fi:uta-201602251249>.
- Tuomi J. & A. Sarajärvi 2006. *Laadullinen tutkimus ja sisällönanalyysi*. Helsinki: Tammi.
- Urareitti 2015. Urareitti – korkeasti koulutetun maahan muuttaneen osaamisen tunnistamisen ja tunnustamisen viitekehys -ESR-hanke. [luettu 30.11.2016]. Saatavissa: <http://www.hamk.fi/tyoelamalle/hankkeet/urareitti/hankkeen-esittely/Sivut/default.aspx>.
- Valvira 2016a. EU/ETA-valtioiden ulkopuolella koulutetut lääkärit. Kielitaitovaatimus. [luettu 30.11.2016]. Saatavissa: http://www.valvira.fi/terveydenhuolto/ammattioikeudet/hakemusohjeet/eu_eta_valtioiden_ulkopuolella_koulutetut/laakarit.
- Valvira 2016b. *Language requirements*. [luettu 30.11.2016]. Saatavissa: http://www.valvira.fi/web/en/healthcare/professional_practice_rights/language_requirements.
- Virtanen, A. (tulossa 2017). The multivoicedness of workplace literacies: an international nursing student documenting in a second language. *European Journal of Applied Linguistics* (ahead of print): <http://www.degruyter.com/view/j/eujal.ahead-of-print/eujal-2015-0015/eujal-2015-0015.xml?format=INT>.
- Wilkinson, S. M., M. Gambles & A. Roberts 2002. The essence of cancer care: the impact of training nurses' ability to communicate effectively. *Journal of Advanced Nursing*, 40 (6), 731–738.
- Wright, S. 2004. *Language policy and language planning: from nationalism to globalisation*. Basingstoke: Palgrave Macmillan.
- Yleiset kielitutkinnot. Kielitutkintojen arviointi ja todistus. [luettu 18.9.2016]. Saatavissa: <https://www.jyu.fi/hum/laitokset/solki/yki/yleista/tietoakielitutkinnoista/arviointi>.

Suulliset lähteet

Mikkonen, M. 2014. Suullinen tiedonanto Maarit Mikkonen, Valvira 6.5.2014.

Lait ja asetukset

Laki yleisistä kielitutkinnoista 964/2004

LIITE 1.

Tekstilajin/diskurssimallin hallinta ja tehtävänannon täyttyminen
1 Ei noudata tekstilajin/diskurssimallin konventioita vaan tyytyy esim. luettelemaan asiat. Suoritus liittyy tehtävän aiheeseen, mutta ei seuraa tehtävänantoa.
2 Joitakin tunnistettavia tekstilajin/diskurssimallin piirteitä. Tehtävänanto täyttyy heikosti (alle puolet osatehtävistä täyttyy).
3 Monia tunnistettavia tekstilajin/diskurssimallin piirteitä. Tehtävänanto täyttyy kohtalaisesti (puolet osatehtävistä täyttyy).
4 Noudattaa melko hyvin tekstilajin/diskurssimallin konventioita. Suurin osa tehtävänannon osatehtävistä täyttyy.
5 Noudattaa hyvin ammatillisen tekstilajin/diskurssimallin konventioita. Tehtävänanto täyttyy kokonaisuudessaan.

Rekisterin hallinta, sanaston ja fraseologian käyttö
1 Ei huomioi tilannetta rekisterin valinnassa. Sanaston ja fraseologian hallinta heikkoa; toistaa ilmauksia suoraan pohja-aineiston mallin mukaan.
2 Osoittaa tiedostavansa jonkin verran rekisterieroja. Sanaston ja fraseologian käytössä vain vähän pyrkimystä etääntyä pohja-aineiston ilmaustavoista.
3 Vaihtelee rekisteriä tilanteen mukaan jonkin verran. Sanaston ja fraseologian käytössä pyrkimystä tietoiseen vaihteluun.
4 Vaihtelee rekisteriä tilanteen mukaan melko sujuvasti ja luontevasti. Käyttää sanastoa ja fraseologiaa itsenäisen kielenkäyttäjän tavoin.
5 Vaihtelee rekisteriä sujuvasti ja luontevasti vastaanottajan ja tilanteen mukaan. Käyttää sanastoa ja fraseologiaa idiomaattisesti varioiden ja osoittaa hyvää konventioiden tunte- musta.

Kohdentaminen ja epäsuora ilmaustapa
1 Ei kohdenna viestiään tietylle vastaanottajalle. Käyttää yksittäisiä kiteytyminä oppimiaan epäsuoria ilmauksia. Rauhoittelua ja luottamuksen synnyttämistä ei ilmene.
2 Joitakin viestin kohdentamisen elementtejä. Pyrkimystä epäsuoraan ilmaustapaan, mutta ilmaisu on vielä epäluontevaa. Käyttää joitakin yksittäisiä kielellisiä keinoja rauhoittelun ja luottamuksen synnyttämisen funktioissa.
3 Pyrkimystä kohdentaa viesti vastaanottajalle, mutta toteutuu epäjohdonmukaisesti. Osoittaa tiedostavansa epäsuoran ja suoran ilmaustavan merkityseron, vaikka niiden käyttö horjuu. Käyttää useammanlaisia ilmaustapoja luottamuksen synnyttämisessä ja rauhoittelussa.
4 Kohdentaa sanottavansa vastaanottajalle melko luontevasti. Osoittaa hallitsevansa epäsuoran ja suoran ilmaustavan merkityseron esimerkiksi pyynnöissä ja toimintaan kehottamisessa ja vaihtelee ilmauskeinoja luottamuksen synnyttämisessä ja rauhoittelussa.
5 Kohdentaa viestin tietoisesti vastaanottajalle, mikä ilmenee esimerkiksi puhuttelun tavoissa ja terminologian avaamisessa. Vaihtelee epäsuoraa ja suoraa ilmaustapaa luontevasti esimerkiksi pyynnöissä ja kehotuksissa ja käyttää idiomaattisia ilmaustapoja luottamuksen synnyttämisessä ja rauhoittelussa.

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 142–166.

Marita Härmälä¹, Raili Hildén² & Dmitri Leontjev¹

¹Jyväskylän yliopisto, ²Helsingin yliopisto

Yhdeksäsluokkalaisten opiskelukäytänteet englannin, ruotsin ja ranskan oppitunneilla ja vapaa-ajalla

The article reports on a quantitative investigation of the differences in Finnish 15–16 year-old students' self-reported foreign/second language (L2) study practices and use across three languages: English ($n=3,476$), Swedish ($n=1,679$), and French ($n=1,023$). The data set consisted of (a) a questionnaire about students' background and L2 study practices/use and (b) their L2 speaking and writing proficiency. The differences between the languages were generally small although the students reported to do their homework in Swedish considerably less regularly than in English and French and English was used more outside classroom than Swedish and French. The differences in the background variables and the correlations between the language practices/use and the language proficiency were generally small, too, although doing homework in French correlated moderately with proficiency. The findings imply paths to enhance coherence across language education integration of language learning in and outside school.

Keywords: study practices, evaluation of learning outcomes, English, Swedish, French

Asiasanat: opiskelukäytänteet, oppimistulosten arviointi, englanti, ruotsi, ranska

1 Johdanto

Perusopetuksen opetussuunnitelman perusteissa 2004 koulutuksen keskeisiksi tavoitteiksi asetettiin ihmisenä kasvu, kulttuurinen identiteetti ja kansainvälisyys, mediataidot ja viestintä sekä yksilön suhde tieto- ja viestintäteknologiaan (Opetushallitus 2004). Samat teemat toistuvat myös syksyllä 2016 käyttöön otetuissa Perusopetuksen opetussuunnitelman perusteissa (Opetushallitus 2014: 20–24), joskin oppiminen nähdään niissä entistä vielä laaja-alaisempaan itsestä huolehtimisena ja arjen taitoina. Kielikasvatuksen lähtökohtana on kielen tai kielten käyttö eri tilanteissa ja kielitietoisuus. Oppilaan monikielinen kompetenssi kehittyy kotona, koulussa ja vapaa-ajalla läpi koko elämän (Opetushallitus 2014: 324). Näiden laaja-alaisen oppimistavoitteiden tulisi heijastua kaikkeen opetukseen ja opiskeluun.

Vuoden 2004 opetussuunnitelman aikana tehdyssä tutkimuksessa (Hildén, Härmälä, Rautopuro, Huhtanen, Puukko & Silverström 2014) tuotiin esiin havaintoja eräiden painoalueiden toteutumisesta kielikasvatuksessa. Modernia tieto- ja viestintäteknologiaa tulisi huomioida enemmän (Taalas 2005), samoin koulun ulkopuolista kielen käyttöä (Luukka, Pöyhönen, Huhta, Taalas, Tarnanen & Keränen 2008). Huolestuneisuutta ilmaistiin myös suullisen harjoittelun (Harjanne & Tella 2011) ja oppilaan autonomiaa tukevan harjoittelun (Hildén & Salo 2011) vähäisyydestä. Tällä opetussuunnitelma-kaudella korostui myös se, että opetuksen tulee tukea oppilaan kulttuurista herkkyyttä (Kaikkonen 2004; Larzén 2005).

Tässä artikkelissa raportoitava tutkimus kohdentuu myös vuoden 2004 perusteeseen, jolloin vieraiden kielten oppimistavoitteisiin perusopetuksessa (Opetushallitus 2004) sisältyivät paitsi kielitaitoon liittyvät tavoitteet myös kulttuuritaidot ja opiskelustrategiat. Opiskelustrategiset valmiudet ovat metakognitiivisia taitoja, kuten esimerkiksi oman työskentelyn suunnittelu ja edistymisen arviointi. Kielenopetuksen tavoitteissa korostuu lisäksi tehokkaiden opiskelutapojen ja pitkäjänteisen viestinnällisen harjoittelun merkityksen ymmärtäminen (Opetushallitus 2004). Tässä artikkelissa opiskelustrategioista käytetään nimitystä opiskelukäytänteet.

Tutkimuskysymykset ovat:

1. Mitä opiskelukäytänteitä oppilailla on englannin, ruotsin ja ranskan kielissä?
2. Mitä oppilaiden taustamuuttujiin liittyviä eroja opiskelukäytänteissä havaitaan?
3. Mitä yhteyksiä opiskelukäytänteillä on oppilaiden englannin, ruotsin ja ranskan kielen puhumis- ja kirjoittamistaitoihin?

Tarkastelemme artikkelissa opiskelukäytänteitä englannin, ruotsin ja ranskan A-oppimäärässä oppilaan näkökulmasta. Paneudumme opiskelukäytänteiden ja kielitaidon yhteyksiin kuvailemalla, analysoimalla ja vertailemalla niitä oppilaan sukupuolen, koulun opetuskielen, oppilaan jatko-opintosuunnitelmien ja vanhempien koulutustason suhteen. Artikkelin pohjautuu Vieraiden kielten ja A-ruotsin oppimistulokset perusopetuksen päättövaiheessa 2013 -hankkeen (jatkossa VK2013) yhteydessä kerättyyn aineistoon (Hildén & Rautopuro 2014; Härmälä, Huhtanen & Puukko 2014; Härmälä & Huhtanen 2014). Pyrimme tässä artikkelissa luomaan VK2013:ssa raportoituja kielikohtaisia havaintoja kokonaisvaltaisemman kuvan oppilaiden opiskelukäytänteistä koulussa ja vapaa-ajalla, esimerkiksi selvittämällä, mitkä käytänteet ovat kaikille tutkituille kielille yhteisiä, ja mitä käytänteitä harjoitetaan toisissa kielissä kenties enemmän kuin toisissa.

Artikkelin rakenne on seuraava. Aluksi luodaan lyhyt katsaus kieltenopetuksen ja -opiskelun suuntauksiin Suomessa, minkä jälkeen raportoidaan aiemmissa oppimistulosten arviointihankkeissa havaittuja yhteyksiä opiskelukäytänteiden ja oppimistulosten välillä. Sen jälkeen esitellään aineisto, keskeiset analyysimenetelmät ja saadut tulokset. Lopuksi pohditaan tulosten antia perusopetuksen kielikasvatukselle.

2 Kieltenopetuksen ja -opiskelun suuntauksia Suomessa

Suomalaisen kieltenopetuksen ja -opiskelun suuntaukset noudattelevat pienellä viiveellä yleisiä länsimaisia virtauksia. Kirjallista kielenkäyttöä ja sen kieliopillista virheetömyyttä vaalivan perinteisen lähestymistavan rinnalla esiintyi jo 1960-luvulla piirteitä audiolingvaalisesta menetelmästä, jonka katsottiin palvelevan paremmin suullisen taidon kehitystä (Scherer & Wertheimer 1964).

Yhtenäiskoulussa lisääntynyt kieltenopiskelu rakentui sekin aluksi klassisten kielten opetuksen muotopainotteiselle ihanteelle, vaikka jo peruskoulun ensimmäinen opetussuunnitelma (POPS 1970) korosti kielen ja sen harjoittelun viestinnällisyyttä. Viestinnällisyyden tunnusmerkkejä olivat tuolloin kuten edelleenkin mielekäs ja merkityspainotteinen kielenkäyttö oppilaan elämään luontevasti liittyvissä yhteyksissä. Oppimiskäsitysten vaihtelut heijastuivat myös kieltenopetukseen ja -opiskeluun: behavioristiseen näkemykseen tukeutuvat lähestymistavat saivat väistyä ensin kognitiivisen ja sosiokonstruktivistisen (Opetushallitus 1994) ja sittemmin sosiokulttuurisen (Opetushallitus 2004) oppimiskäsityksen tieltä. Parhaan metodin metsästyksestä on vihdoin luovuttu niin maailmalla kuin Suomessakin (Kumaravadelu 2006), ja tämän päivän kieliluokissa menneisyys ja moderni elävät rinnan monikerroksisessa opiskelutodellisuudessa, jonka kautta opetus ja kieliaines välittyvät oppilaan muistiin, mieleen ja kielenkäyttöön. Myös laajemmat sosiopoliittiset painotukset Suomen kansainvälistymisen ja talouden globalistumisen myötä vauhdittivat esimerkiksi suullisen kielitaidon ja

kulttuuritaidon opetusta 1990-luvulla. Viime vuosikymmeninä myös kieltenopettajia on kannustettu erityisesti tieto- ja viestintäteknologian pedagogiseen hyödyntämiseen (Opetushallitus 2004, 2014).

Suomalainen kieltenopetus on nykyään lähtökohtaisesti merkitys- ja viestintälähtöistä ja tehtäväperustaista (Ellis, 2012, 2014; Larsen-Freeman 2015; Hildén 2009). Tehtävällä (task) tarkoitetaan useimpien vakiintuneiden määritelmien mukaan merkityskeskistä kielenkäyttötehtävää, joka mahdollistaa oppilaiden kielellisten ja ei-kielellisten resurssien hyödyntämisen jonkin oppilaalle tärkeän ongelman ratkaisemiseksi (Prabhu 1987), merkitysneuvottelun toiminnan aikana (Pica & Doughty 1985) ja monisuuntaisen vuorovaikutuksen opettajan ja oppilastoverien kesken (Gass 2005).

Tehtävällä on siis selkeä tavoite, joka ei rajoitu ainoastaan kieliaineksen oppimiseen (Ellis & Shintani 2014: 135–137), vaan parhaimmillaan mahdollistaa oppilaan mielekkään osallistumisen tärkeisiin arjen ja oman elämänpiiriin toimintoihin. Kielelliset muodot käsitellään viestinnällisessä yhteydessään ja periaatteessa oppilaan tarpeiden mukaan (Ellis 2012: 272). Opettajan antama oikea-aikainen tuki auttaa oppilasta muokkaamaan tehtävän sisältämää kieliainesta lähikehityksen vyöhykkeelleen sopivaksi ja siten edistämään siitä oppimista (Vygotsky 1982; Aljaafreh & Lantolf 1994). Opettaja voi tukea oppilaan kielitaidon kehitystä käyttämällä opetuksessaan kohdekieltä mahdollisimman paljon (Ellis 1984). Opettajan käyttämän kielen hyödyllisyys riippuu tilanteesta ja tavoitteesta, ja välittävän näkemyksen mukaan kielioppiasiat voidaan käsitellä tehokkaimmin äidinkielellä (Ellis 2012: 126–129).

Tehtäväperustaisen opetuksen ja harjoittelun laatua ei kuitenkaan määritä se, miten tehtävän on suunniteltu etenevän, vaan ennen kaikkea tehtäväsuunnitelman toteutuminen toimintana (Seedhouse 2005). Esimerkiksi parikeskusteluissa tämä tarkoittaa, että vuorovaikutukseltaan tasa-arvoiset ja vastavuoroiset keskustelut tuottavat parempaa oppimista kuin epäsuhtaiset ja toisen osapuolen dominoimat keskustelut (Storch 2002). Pari- ja ryhmätyöskentely tarjoaa myös tilaisuuksia oppimista auttavaan vertaistukeen (Donato & McCormick 1994; Ohta 2000).

Viestinnällisen kielenopetuksen tavoitteena on myös kehittää oppilaan kykyä vaalia ja kartuttaa taitojaan kouluopetuksen ulkopuolella osana omaa autenttista elämänpiiriään. Opiskelun strategiaan tavoitteisiin lukeutuu taito suunnitella opiskeluaan, soveltaa monipuolisia opiskelumuotoja ja arvioida omaa työskentelyään ja sen tuloksia (Oxford ym 2014; Cohen & Macaro 2007; Plonski 2011). Opiskelustrategioiden tehokkuus on mitä suurimmassa määrin yksilöllistä ja tilanteesta riippuvaista, joten niidenkin paremmuudesta on turha kiistellä. Tosin eräät strategiat ovat Suomessakin saaneet muita enemmän huomiota oppimateriaalien tai aktiivisten opettajankouluttajien kautta, kuten esimerkiksi sanaelaborointi (Kristiansen 1998), suullinen kielitaito (Hildén 2000) ja portfoliotyöskentely eurooppalaisen kielisalkun avulla (Kohonen & Pajukanta 2003).

Missä määrin kielten opiskelu koulussa sitten vastaa nykyoppilaan kielitaitotarpeita? Laajassa perusopetuksen 9. luokkalaisten äidinkielen ja vieraiden kielten lukemisen ja kirjoittamisen käytänteitä selvittäneessä hankkeessa (Luukka ym. 2008) todettiin muun muassa, että oppilaat sosiaalistuvat nyky-yhteiskunnan uusiin tekstikäytänteisiin enemmän vapaa-ajallaan kuin formaalissa opetuksessa. Koulun ja vapaa-ajan käytänteet eivät ole tasa-arvoisessa asemassa, sillä oppilas voi valita vapaa-ajan käytänteet omien intressiensä mukaan, kun taas koulussa opettaja useimmiten määrittelee tekstit, joiden kanssa työskennellään. Tutkimuksessa havaittiin myös eroja oppilaiden ja opettajien raportoimien työskentelytapojen välillä. Oppilaiden mielestä opettajajohtoista työskentelyä oli tunneilla enemmän kuin opettajien mielestä. (Luukka ym. 2008, 236–241.)

3 Opiskelukäytänteiden yhteyksiä osaamiseen

Opiskelukäytänteiden yhteyksistä oppimistuloksiin on hankittu tietoa kansallisissa oppimistulosten arvioinneissa, jotka ovat kohdistuneet varsin kattavasti perusopetuksessa opiskeltaviin oppiaineisiin. Viimeisimmissä, 2010-luvulla tehdyissä arvioinneissa on havaittu muun muassa tehtävien säännöllisen tekemisen selittävän tyttöjen menestymistä matematiikassa (Julin & Rautopuro 2016) ja olevan yhteydessä paitsi osaamiseen, myös asenteiden myönteisyyteen esimerkiksi historiassa ja yhteiskuntaopissa (Quakrim-Soivio & Kuusela 2012). Sen sijaan läksyihin käytetyn ajan pituus ei näyttäisi ainakaan terveystiedossa liittyvän osaamisen laatuun (Summanen 2014).

Oppilaan harrastuneisuudella (esim. kirjojen lukemisella, elokuvien katselulla, pelien pelaamisella) ja osaamisella on havaittu olevan myönteinen yhteys suoriutumiseen historiassa ja yhteiskuntaopissa (Ouakrim-Soivio & Kuusela 2012) sekä äidinkielessä ja kirjallisuudessa (Lappalainen 2011; Harjunen & Rautopuro 2015). Myös kotitaloudessa kohtalainenkin harrastuneisuus vapaa-ajalla oli yhteydessä parempiin oppimistuloksiin (Venäläinen 2015).

Kielten oppitunneilla käytetyt työtavat näyttäisivät muuttuneen vain vähän vuosien mittaan. Esimerkiksi toisen kielen (S2) oppitunneilla opetus keskittyi kielen rakenteisiin, tekstien tuottamiseen ja ymmärtämiseen. Opettajajohtoinen opetuskeskustelu oli tyypillisin opetusmenetelmä. Vaikka pelejä ja leikkejä käytettiin melko yleisesti, tieto- ja viestintäteknologian hyödyntäminen oli vähäistä. Työtavoista ainoastaan pelien ja leikkien maltillinen käyttö (joskus tai vain harvoin) tuotti keskimäärin parempaa osaamista. (Kuukka & Metsämuuronen 2016.)

Myös VK2013:ssa havaittiin että vapaa-ainana tapahtuneella englannin harrastamisella oli kohtalainen yhteys oppilaan kielitaitoon (Härmälä ym. 2014: 162), muissa arvioiduissa kielissä yhteys oli heikompi. Englannin oppimistulosten edellisessä arvioin-

nissa vuonna 1999 (Tuokko 2000) yhteys oli ollut vain pieni (korrelaatiot: kirjeenvaihtoveri 0,28; lehtien lukeminen 0,23; internet 0,21). Englanninkielisiä internetsivuja selaili tuolloin oppilaista noin joka toinen (51 %), samoin noin puolet pelasi tietokonepelejä (50 %), pojat huomattavasti tyttöjä enemmän.

Eurooppalaisessa kielitaidon arviointitutkimuksessa (European Commission 2012) havaittiin positiivinen yhteys oppilaiden ja vanhempien kielitaidon välillä sekä sillä, missä määrin opettaja ja oppilaat käyttivät kohdekieltä kielen oppitunneilla. Myös perinteisen ja nykyaikaisen median käytön todettiin olevan yhteydessä oppilaiden parempaan kielitaitoon. Yllättävä havainto oli kuitenkin, että tieto- ja viestintäteknologian käytöllä oli negatiivinen yhteys kotona tapahtuvaan kielenopiskeluun (European Commission 2012: 96). Vaikka opettajat käyttivät tieto- ja viestintäteknologiaa oppituntien ulkopuolella, he toivat nettisisältöjä oppitunneilleen vain harvoin. Osasyyski ilmoitettiin materiaalin huono saatavuus. Oppilaat sen sijaan käyttivät nettiä koulun ulkopuolella moniin eri tarkoituksiin, eivät kuitenkaan vieraan kielen läksyjen tekemiseen. Erityisesti englannin opetuksen ongelmana on, miten kuroa umpeen kuilu (ks. myös Olsson & Sylvén 2011) koulussa tapahtuvan ja koulun ulkopuolisen (Sundqvist 2013) englannin kielen käytön välillä. Kuilun on havaittu heikentävän 13–16-vuotiaiden opiskelumotiivaatiota englannin tunneilla.

4 Tutkimusaineisto ja analyysimenetelmät

Artikkelin aineisto on kerätty VK2013:n yhteydessä, ja se koostuu oppilaskyselyistä ja oppilaiden tuloksista kielitaitoa mittaavissa tehtävissä (puhuminen ja kirjoittaminen). Oppilasmäärät tarkasteltavissa kielissä on ryhmitelty taulukossa 1 koulun opetuskielen mukaan. Sulkeissa ilmoitetaan arviointiin osallistunut oppilasmäärä kielittäin. Oppimistulosten arvioinneissa otoskoulut valitaan niin, että ne edustavat tasapuolisesti maan eri osia (AVI-alueet) ja toimintaympäristöjä (kaupunki, taajama, maaseutumainen kunta). VK2013:ssa arviointi suoritettiin samanaikaisesti viidessä eri kielessä ja kahdessa oppimäärässä. Arviointiin osallistui samasta koulusta korkeintaan kaksi oppiainetta.

TAULUKKO 1. Oppilasmäärät kielittäin.

Kieli ja oppimäärä	suomenkieliset	ruotsinkieliset	yhteensä
A-englanti	2 966 (95 %)	510 (99 %)	3 476
A-ruotsi	1 679 (90 %)	-	1 679
A-ranska	1 023 (86 %)	-	1 023

Oppilaskyselyn taustatiedoissa kysyttiin äidinkielen lisäksi oppilaan kotikieli, jatko-opintosuunnitelmat, vanhempien koulutustaso, oppilaan A1-kieli, viimeisin arvosana arvioitavassa oppiaineessa ja oppilaan käsitys arvosanan osuvuudesta. Lisäksi tiedusteltiin oppilaan läksyihin käyttämää aikaa ja tehtävien teon säännöllisyyttä.

Opiskelukäytänteisiin luettiin oppitunneilla käytetyt opiskelukäytänteet sekä oppilaan vapaa-ajallaan harjoittamat toiminnot arvioitavalla kielellä. Opiskelukäytänteiden monipuolisuuden ja erityisesti koulun ulkopuolisen harrastuneisuuden määrän ja laadun katsottiin kertovan oppilaan valmiuksista tarttua ympäristössään tarjoutuviin oppimismahdollisuuksiin. Oppilaskyselyn käytänteitä koskevat kysymykset laadittiin vastaamaan teoreettista tietoa tehtäväperustaisen opetuksen ja opiskelun tunnusmerkeistä. Valtaosa kyselyn käytänteistä kuvaa varsin perinteisiä kielen opetuksessa ja opiskelussa käytettyjä työtapoja, osa pohjautuu jonkin verran uudempiin käsityksiin. Oppilaat vastasivat oppitunneilla harjoitettuja käytänteitä kartoittaviin kysymyksiin (21 osiota) 5-portaisella Likert-asteikolla (*ei koskaan–vain harvoin–joskus–usein–lähes aina*) ja koulun ulkopuolisiin käytänteisiin (8 osiota) 4-portaisella asteikolla (*ei koskaan–joskus–viikoittain–päivittäin*). Kysymykset on esitelty liitteessä 1.

Kyselyaineisto analysoitiin SPSS-ohjelmalla. Johtuen aineiston luonteesta (Likert-asteikolliset osiot) sekä siitä, että oppilaiden vastausten histogrammijakaumat olivat erittäin vinoutuneita valtaosassa kyselyosiota, analyysit tehtiin ei-parametrisilla menetelmillä. Myös taustamuuttujien osalta jouduttiin tekemään useita ratkaisuja analyysien mielekkyyden ja luotettavuuden varmistamiseksi. Oppilaiden jatko-opintosuunnitelmista otettiin mukaan ainoastaan lukio ja ammatillinen koulutus, sillä muita vaihtoehtoja (välivuosi ja töihin meno) valitsi vain noin 1 % ja 0,2 % oppilaista. Lisäksi yhdistettiin kysymykset, joissa tiedusteltiin erikseen oppilaan isän tai äidin ylioppilastustaa (0 = ei kumpikaan ylioppilas; 1 = äiti ylioppilas; 2 = isä ylioppilas; 3 = molemmat ylioppilaita). Myös tapaukset, joissa oppilaat olivat vastanneet "en tiedä", poistettiin. Sellaisissa koulun ja vapaa-ajan opiskelukäytänteissä, joissa oli enemmän kuin kaksi luokkaa (esim. kolme kieltä), ryhmien välisiä eroja analysoitiin Kruskal-Wallisn testeillä. Mann-Whitney U -testillä puolestaan selvitettiin, oliko ryhmien välillä merkitseviä eroja esimerkiksi sukupuolten välillä. Erojen raportoinnissa käytettiin seuraavia raja-arvoja: hyvin pieni, jos $r < 1$; pieni, jos $r \geq 1$; kohtalainen, jos $r = 3$; suuri, jos $r \geq 5$ (tai suurempi).

Koulun ja vapaa-ajan opiskelukäytänteiden välistä yhteyttä kielitaitoon selvitettiin ei-parametrisen korrelaatioanalyysien avulla. Oppilaiden kielitaito (kuullun ja luetun ymmärtäminen, puhuminen ja kirjoittaminen) oli päätutkimuksessa (esim. Härmälä & Huhtanen 2014) kuvattu Eurooppalaisen viitekehyksen suomalaisella soveluksella (Opetushallitus 2004), jossa käytettiin taitotasoja A1.1–B2.1. Ranskan kielessä luetun ja kuullun ymmärtämistehtävissä luotettavasti todennettu asteikko oli kuitenkin lyhyempi kuin englannissa ja ruotsissa (A-ranska: A1.3–B1.1, muissa A1.1–B2.1), ja siksi korrelaatioanalyysi rajoitettiin koskemaan ainoastaan puhumisen ja kirjoittamisen

osakokeita. Koulun opetuskielen yhteyttä käytänteiden harjoittamiseen tarkasteltiin vain englannin kielen osalta, sillä kaikki ruotsin ja A-ranskan arviointiin osallistuneet oppilaat olivat suomenkielisistä kouluista. Saatujen tulosten valtavan määrän vuoksi tässä artikkelissa raportoidaan vain koko otosta koskevat havainnot. Muista mielenkiintoisista tuloksista tehdään yhteenveto ja esitetään tulokset, joissa korrelaatiot ja efektiivisyydet olivat vähintään pieniä.

5 Tulokset

Seuraavassa kootaan tutkimuksen päätulokset ja vastataan tutkimuskysymyksiin. Ensin kuvaillaan oppilaiden opiskelukäytänteitä sekä oppitunneilla että vapaa-ajalla. Sen jälkeen tarkastellaan koulun ja vapaa-ajan opiskelukäytänteiden yhteyttä oppilaan sukupuoliin, koulun opetuskielen, oppilaan jatko-opintosuunnitelmiin ja vanhempien koulutustasoon sekä oppilaan kielitaitoon. Kielitaidosta selvitettiin yhteydet puhumiseen ja kirjoittamiseen, joissa taitotasoluokitus oli yhteismitallinen. Opiskelukäytänteiden yhteydet kuullun ja luetun ymmärtämistaitoihin on kuvattu tarkemmin kielikohtaisissa raporteissa (Hildén & Rautopuro 2014; Härmälä & Huhtanen 2014; Härmälä ym. 2014). Kaikille kielille yhteisten tulosten lisäksi annetaan myös esimerkkejä muutamista kielikohtaisista havainnoista.

5.1 Koulun ja vapaa-ajan opiskelukäytänteet

Koulun ja vapaa-ajan opiskelukäytänteissä olevien erojen selvittämiseksi (tutkimuskysymys 1), laskettiin ensin tilastolliset tunnusluvut ja frekvenssit kyselylomakkeen osi-
oille kielittäin. Tuloksista (Liite 2) ilmeni esimerkiksi, että oppilaat ilmoittivat tekevänsä ruotsin tehtävät huomattavasti harvemmin kuin englannin ja ranskan tehtävät. Siksi oli tarpeen selvittää, oliko oppilaiden vastauksissa myös muita kielten välisiä eroja, jotka Kruskal-Wallis testillä pystyy paljastamaan. Seuraavassa raportoidaan näistä eroista ne, joissa efektiivisyys oli vähintään pieni ($r \geq 0,1$).

Kuten odotettua, kielten väliset erot tehtävien teon säännöllisyydessä olivat merkitseviä ($\chi^2(2, N=6\,128)=2\,925,58, p<,001$). Ero englannin ja ruotsin välillä oli suuri ($n=5\,109, Z=-53,241, p<,001, r=,74$), samoin ruotsin ja ranskan välillä ($n=2\,683, Z=-37,421, p<,001, r=,51$). Käytännössä tämä tarkoitti, että oppilaat ilmoittivat tekevänsä ruotsin tehtävät harvemmin kuin englannin ja ranskan tehtävät. Vaikka ero englannin ja ranskan välillä oli merkitsevä, se oli käytännössä melko pieni ($n=4\,464, Z=-5,911, p<,001, r=,09$) ja jäljitettävissä ainoastaan suuresta otoskoosta johtuvaksi.

Kaikki muut kielten väliset erot olivat pieniä, eivätkä kaikki edes tilastollisesti merkitseviä. Osiot, joissa kahden kielen väliset erot olivat havaittavissa (vaikka pieniä), on koottu taulukkoon 2. Taulukosta ilmenee, että ranskan kielen tunneilla harjoitetaan ainakin osin erilaisia opiskelukäytänteitä kuin englannin ja ruotsin tunneilla. Ranskan kielen ääntämisen ja oikeinkirjoituksen erityisluonteen vuoksi on varsin ymmärrettävää, että tunneilla kuunnellaan esimerkiksi enemmän musiikkia ja tuotetaan enemmän tekstiä kirjallisesti. Ranskan oppilaat kokivat myös opettajan kannustavan heitä englantia ja ruotsia enemmän ranskan käyttöön vapaa-ajalla. Englannin ja ruotsin tunneilla tehdään puolestaan ranskaa enemmän sanaston elaborointiharjoituksia.

TAULUKKO 2. Kyselylomakkeen osiot, joissa on kahden kielen välisiä eroja.

Taustakyselyn osio	<i>n</i>	Z-arvo	<i>p</i> -arvo	efektikoko (<i>r</i>)
Enemmän elokuvia ja lauluja (17):				
ranskassa kuin englannissa	4 462	-13,239	< ,001	,20
ranskassa kuin ruotsissa	2 682	-8,294	< ,001	,16
Enemmän sanastoharjoituksia (13):				
englannissa kuin ranskassa	4 431	-13,846	< ,001	,21
ruotsissa kuin ranskassa	2 664	-9,978	< ,001	,19
Enemmän kirjallista tuottamista (31):				
ranskassa kuin englannissa	4 416	-11,403	< ,001	,17
ranskassa kuin ruotsissa	2 661	-7,681	< ,001	,15
Opettaja kannustaa kielen käyttämiseen vapaa-ajalla enemmän ranskassa (32):				
kuin englannissa	4 429	-11,385	< ,001	,17
kuin ruotsissa	2 664	-7,813	< ,001	,15
Internetiä käytetään vähemmän tiedonhakuun (23):				
ruotsissa kuin englannissa	5 079	-9,230	< ,001	,13
ranskassa kuin englannissa	2 679	-8,769	< ,001	,17
Kielioppia harjoitellaan tietokoneella enemmän (22):				
ruotsissa kuin ranskassa	2 680	-10,895	< ,001	,21

Tulokset koulun ulkopuolisesta kielen käyttämisestä olivat varsin odotuksenmukaisia, ja ne tukevat aikaisempia tutkimustuloksia suomalaisnuorten englannin kielen käytös-

tä (Luukka ym. 2008; Leppänen, Pitkänen-Huhta, Nikula, Kytölä, Törmäkangas, Nissinen, Kääntä, Virkkula, Laitinen, Pahta, Koskela, Lähdesmäki & Jousimäki 2009). Myös tässä tutkimuksessa tulos oli, että yhdeksäsluokkalaiset käyttivät englantia vapaa-ajalla merkittävästi enemmän kuin ruotsia ja ranskaa. Erot englannin ja ruotsin välillä vaihtelivat kohtalaisesta (pienin ero kielen käyttäminen turistien kanssa, $n=5\ 076$, $Z=-20,828$, $p<,001$, $r=,29$) suureen (musiikin kuunteleminen, $n=5\ 087$, $Z=-57,687$, $p<,001$, $r=,81$). Ainoa poikkeus oli kielen käyttäminen ystävien ja sukulaisten kanssa, missä ero englannin ja muiden kielten välillä oli vain pieni ($n=5\ 084$, $Z=-13,852$, $p<,001$, $r=,19$). Tulokset olivat samansuuntaiset englannin ja ranskan välillä, missä erot vaihtelivat kohtalaisesta (kielen käyttäminen ystävien ja perheen kanssa: $n=4\ 442$, $Z=-19,847$, $p<,001$, $r=,29$) suureen (musiikin kuunteleminen: $n=4\ 444$, $Z=-48,868$, $p<,001$, $r=,73$).

Kaikki ruotsin ja ranskan väliset erot kielen käyttämisessä vapaa-ajalla olivat pieniä. Esimerkkinä havaittavissa olevista eroista voidaan mainita musiikin kuunteleminen ja sosiaalisen median käyttäminen. Oppilaat kuuntelivat musiikkia enemmän ranskassa ($n=2\ 677$, $Z=-5,037$, $p<,001$, $r=,10$), mutta käyttivät sosiaalista mediaa puolestaan enemmän ruotsissa ($n=2\ 675$, $Z=-5,521$, $p<,001$, $r=,11$). Ruotsia käytettiin myös ranskaa enemmän ystävien ja perheen kanssa ($n=2\ 674$, $Z=-7,544$, $p<,001$, $r=,15$), mikä onkin ymmärrettävää ruotsin kielen virallisen aseman vuoksi.

Yhteenvetona voidaan todeta, että vaikka tutkitut kolme kieltä olivat oppimistavoitteiltaan yhteneväisiä, oli oppilaiden ilmoittamissa opiskelukäytänteissä jonkin verran myös kielikohtaisia, kielen erikoisluonteen huomioon ottavia eroja, kuten esimerkiksi ääntämisen ja oikeinkirjoituksen säännöllisempi harjoittelu ranskassa. Odotuksenmukaista sen sijaan oli, että oppilaiden vapaa-ajan käytänteissä oli selkeitä eroja. Osa eroista selittyy ainakin osaksi englannin asemalla median ja joukkoviestinten valtakielinä. Ruotsin osalta olisi kuitenkin tärkeää selvittää, miksi A-ruotsin valinneet oppilaat ilmoittivat käyttävänsä kieltä vapaa-ajalla varsin vähän ja tekevänsä tehtävät jonkin verran harvemmin kuin muissa tutkituissa kielissä.

5.2 Käytänteiden yhteys taustatekijöihin

Tutkimuskysymyksen 2 tarkoittamista oppilaan taustatekijöistä analysoitiin sukupuolen, koulun opetuskielen, oppilaan jatko-opintosuunnitelmien ja vanhempien koulutustason yhteyttä opiskelukäytänteiden harjoittamiseen. Seuraavassa selostetaan päätulokset sekä muutamia kielten välillä havaittuja eroja. Tilastollisesti merkitsevien erojen kielentämisessä ohjenuorana oli, että lähellä arvoa $r=,30$ olevat arvot indikoivat kohtalaisia eroja, sen alle jäävät pieniä.

Vaikka Mann-Whitney U -testit paljastivat joitakin tyttöjen ja poikien välisiä eroja, olivat erot hyvin pieniä (suurin niistä $r=,20$). Myös kielten väliset erot olivat pieniä. Efektikooltaan suurimmat erot havaittiin tehtävien tekemiseen käytetyn ajan pituudes-

sa, nekin kuitenkin vain suuren otoskoon ansiosta (englanti: $n=3\ 431$, $Z=-11,695$, $p<,001$, $r=,20$; ruotsi: $n=1\ 664$, $Z=-4,959$, $p<,001$, $r=,12$; ranska: $n=1\ 016$, $Z=-4,447$, $p<,001$, $r=,10$). Vaikka tytöt käyttivät kaikissa tutkituissa kielissä poikia enemmän aikaa kotitehtäviin, ei oppilaan sukupuolella ollut juurikaan yhteyttä oppituntien opiskelukäytänteisiin.

Toinen mainitsemisen arvoinen kielikohtainen tulos oli, että tytöt kertoivat käyttävänsä ruotsin kieltä enemmän vapaa-ajallaan kuin pojat. Vaikka kaikissa koulun ulkopuolista kielen käyttöä koskevissa osioissa havaittiin sukupuolten välillä tilastollisesti merkitseviä eroja, olivat ne käytännössä pieniä (arvot $r=,10$:n ja $r=,20$:n välillä). Tytöt ilmoittivat esimerkiksi kuuntelevansa vapaa-ajallaan ranskalaista musiikkia jonkin verran useammin kuin pojat ($n=1\ 015$, $Z=-7,234$, $p<,001$, $r=,22$).

Kuten luvussa 4 mainittiin, koulun opetuskielen yhteyttä opiskelukäytänteiden harjoittamiseen selvitettiin ainoastaan englannin kielen osalta, sillä ruotsin ja A-ranskan arviointiin ei osallistunut ruotsinkielisten koulujen oppilaita. Englannin osalta päätulos oli, että erot suomenkielisten ja ruotsinkielisten koulujen välillä olivat pieniä tai jopa hyvin pieniä. Suurin yksittäinen ero liittyi suullisiin pariharjoituksiin, joita tehtiin suomenkielisissä kouluissa useammin kuin ruotsinkielisissä kouluissa ($n=3\ 429$, $Z=-13,977$, $p<,001$, $r=,24$). Ruotsinkielisissä kouluissa pidettiin puolestaan kirjallisia kokeita päivän läksystä useammin kuin suomenkielisissä. Ero kieliryhmien välillä oli kohtalainen ($n=3\ 403$, $Z=-20,722$, $p<,001$, $r=,36$).

Samoin kuin sukupuolen ja koulun opetuskielen, myös oppilaan jatko-opintosuunnitelmien osalta kielten väliset erot olivat pieniä, lähes olemattomia. Ainoa kohtalainen ero havaittiin englannin kielessä, missä lukioon jatkavat oppilaat ilmoittivat tekevänsä kotitehtävät säännöllisemmin kuin ammatilliseen koulutukseen aikovat ($n=3\ 358$, $Z=-16,300$, $p<,001$, $r=,28$). Tässä yhteydessä on syytä myös todeta, että A-ruotsin ja A-ranskan valinneissa oppilaisissa oli ammatilliseen koulutukseen aikovia kokonaisuudessaan vähemmän (ruotsi 23 %, ranska 10 %) kuin englannissa (43 %).

Vanhempien koulutustason vaikutusta opiskelukäytänteiden harjoittamiseen selvitettiin Kruskal-Wallis-testillä. Kyselylomakkeessa oppilaat ilmoittivat vanhempien koulutustason valitsemalla vaihtoehdoista "kumpikaan vanhemmista ei ylioppilas", "äiti ylioppilas", "isä ylioppilas", "molemmat vanhemmat ylioppilaita" Huomionarvoista on, että merkitseviä, vanhempien koulutustasoon yhteydessä olevia eroja ilmeni ainoastaan englannin kielen käyttämisessä vapaa-ajalla. Kahdeksasta koulun ulkopuolista kielen käyttöä koskevasta osiosta ainut, jolla ei englannissa ollut yhteyttä vanhempien koulutustasoon, oli englannin käyttäminen sosiaalisessa mediassa.

Englannin ohella ainoastaan yhdessä ruotsin osiossa (ruotsinkielisten televisio-ohjelmien katseleminen) oppilaiden vastaukset vaihtelivat vanhempien koulutustason mukaan. Kahden ylioppilasvanhemman lapset ilmoittivat katsovansa enemmän ruot-

sinkielistä televisiota kuin ei-ylioppilasvanhempien lapset. Tosin ero ryhmien välillä oli jälleen hyvin pieni ($n=837$, $Z=-2,687$, $p=.04$, $r=.09$).

Lähes samankaltainen trendi oli osiossa, jossa tiedusteltiin englannin käyttämistä koulun ulkopuolella. Tässä osiossa oli myös suurin ero ei-ylioppilasvanhempien ja kahden ylioppilasvanhemman lasten välillä. Ylioppilasvanhempien lapset käyttivät englantia koulun ulkopuolella merkitsevästi enemmän kuin ei-ylioppilasvanhempien lapset. Ero oli kuitenkin efektikooltaan pieni ($r=.09$ vs. $r=.15$).

Mielenkiintoista oli, että englanninkielisten televisio-ohjelmien katseleminen ja erilaisten englanninkielisten tekstien lukeminen vaihtelivat merkitsevästi sen mukaan, oliko ylioppilasvanhempi oppilaan isä vai äiti. Oppilaat, joiden isä oli ylioppilas, katsoivat enemmän englanninkielisiä ohjelmia televisiosta ja he myös lukivat enemmän englanninkielisiä tekstejä kuin oppilaat, joiden äiti oli ylioppilas ($n=1\ 029$, $Z=-3,397$, $p=.004$, $r=.11$ (pieni) ja $n=1\ 027$, $Z=-3,743$, $p=.001$, $r=.12$). Ero ryhmien välillä oli kuitenkin edelleen pieni.

Kaiken kaikkiaan oppilaan sukupuolen, jatko-opintosuunnitelmien ja vanhempien koulutustason yhteys opiskelukäytänteiden harjoittamiseen oli pieni, eikä niiden pohjalta voida tehdä laajakantoisia johtopäätöksiä. Mielenkiintoinen yksittäinen havainto oli kuitenkin, että koulun opiskelukäytänteissä esiintyi jonkin verran eroja suomen- ja ruotsinkielisten koulujen oppilaiden välillä, tosin eroja voitiin selvittää tässä tutkimuksessa ainoastaan englannin kielen osalta.

5.3 Käytänteiden yhteys kielitaitoon

Seuraavaksi tarkastellaan opiskelukäytänteiden ja oppilaiden puhumis- ja kirjoittamistehtävissä saamien tulosten välistä yhteyttä tutkimuskysymyksen 3 mukaisesti. Ensin raportoidaan koulussa harjoitettujen käytänteiden ja kielitaidon väliset yhteydet kieli-kohtaisten esimerkkien avulla. Sen jälkeen selvitetään, mikä yhteys oppilaiden vapaaajalla tapahtuvalla kielen harrastamisella oli kielitaitoon, ja annetaan esimerkkejä kieli-kohtaisista eroista.

Useimpien oppitunneilla harjoitettujen opiskelukäytänteiden yhteys kielitaitoon oli heikko tai jopa lähes olematon. Kielten välillä havaittiin kuitenkin useita eroja. Yksi niistä liittyi tehtävien teon säännöllisyyteen, mikä ranskan kielessä oli kohtalaisessa yhteydessä sekä puhumiseen ($r_s=.410$, $p<.001$) että kirjoittamiseen ($r_s=.493$, $p<.001$). Englannissa tehtävien teon säännöllisyys korreloi kirjoittamiseen kuitenkin vain heikosti ($r_s=.217$, $p<.001$) ja puhumiseen jopa hyvin heikosti ($r_s=.197$, $p<.001$). Ruotsin kielessä tehtävien teon säännöllisyyden yhteys sekä puhumiseen että kirjoittamiseen oli heikko ($r_s=.146$ ja $r_s=.160$) ja merkitsevää ainoastaan suuren otoskoon vuoksi.

Toinen ero kielten välillä liittyi opettajan antamaan kannustukseen. Ruotsissa opettajan kannustus käyttää kieltä myös koulun ulkopuolella oli heikossa yhteydessä

oppilaiden kirjoittamistaitoon ($r_s = .212, p < .001$). Yhteys ruotsin kielen puhumiseen oli hyvin heikko ($r_s = .170$). Ranskassa opettajan kannustuksella oli yhteyttä oppilaiden puhumistaitoon ($r_s = .203, p < .001$), kun taas yhteys kirjoittamiseen oli jonkin verran heikompi ($r_s = .199$). Englannissa sen sijaan opettajan kannustamisen ja kielitaidon yhteys oli lähes olematon ($r_s = .090$ ja $r_s = .059$). Oli myös mielenkiintoista havaita, että oppilaiden harjoittama itsearviointi ranskan kielessä oli heikossa yhteydessä kielitaitoon ($r_s = .235, p < .001$ ja $r_s = .279, p < .001$). Vastaavaa yhteyttä ei ollut havaittavissa englannin tai ruotsin kielissä.

Myös kielen käyttäminen vapaa-ajalla oli vain heikosti yhteydessä oppilaan kielitaitoon, joskin yhteys oli englannin kielessä hiukan vahvempi kuin ruotsissa ja ranskassa. Englannissa esimerkiksi television katsominen vapaa-ajalla korreloi oppilaan kielitaitoon ($r_s = .370, p < .001$ and $r_s = .358, p < .001$ puhuminen ja kirjoittaminen). Yhteys oli huomattavasti heikompi ruotsissa ($r_s = .286, p < .001$ ja $r_s = .273, p < .001$) ja ranskassa ($r_s = .226, p < .001$ and $r_s = .274, p < .001$). Samankaltaisia trendejä oli myös muissa vapaa-ajan opiskelukäytännöissä. On kuitenkin syytä huomata, että tulos oli toinen, jos oppilas ilmoitti käyttävänsä kieltä ystävien ja sukulaisten kanssa. Silloin korrelaatio oli ruotsissa korkeampi ($r_s = .315, p < .001$ ja $r_s = .324, p < .001$ puhuminen ja kirjoittaminen) kuin englannissa ($r_s = .290, p < .001$ ja $r_s = .280, p < .001$) ja ranskassa ($r_s = .215, p < .001$ and $r_s = .251, p < .001$).

Yhteenvedon voidaan todeta, että tehtävien säännöllinen tekeminen kannattaa kielissä, joiden käyttömahdollisuudet koulun ulkopuolella ovat rajalliset. Näissä kielissä, opettajan kannustus ja opastus tarjoumien löytämiseksi ovat ensiarvoisen tärkeitä. Tulisi myös miettiä keinoja, joilla vapaa-ajan opiskelukäytännöt voitaisiin nykyistä paremmin nivoa osaksi koulussa tapahtuvaa kielenopiskelua (ks. myös Luukka ym. 2008).

6 Johtopäätökset

Artikkelissa raportoitiin perusopetuksen päättövaiheessa olevien englantia, ruotsia ja ranskaa opiskelleiden opiskelukäytäntöjä sekä koulussa että vapaa-ajalla. Aineistona käytettiin VK2013:ssa kerättyjä oppilaiden taustakyselyjä ja heidän kielitaidon eri osa-alueita mitanneissa tehtävissä saamiaan taitotasoja. Kielitaidon osa-alueista valittiin puhuminen ja kirjoittaminen, sillä niissä taitotasoasteikot olivat yhteismitalliset kaikissa kolmessa kielessä. Analyysin rajoittaminen tuottamistaitoihin oli perusteltua myös siksi, että kuullun ja luetun ymmärtämisen yhteydet opiskelukäytäntöisiin on kuvattu yksityiskohtaisesti VK2013:n kielikohtaisissa raporteissa (Hildén & Rautopuro 2014; Härmälä ym. 2014; Härmälä & Huhtanen 2014).

Tutkimuksen päätarkoitus oli kuvata ja analysoida, mitä opiskelukäytänteitä oppilaat harjoittivat eri kielissä, mitä oppilaan taustamuuttujiin liittyviä eroja käytänteiden harjoittamisessa oli, ja mitä yhteyksiä opiskelukäytänteillä oli oppilaiden suulliseen ja kirjalliseen kielitaitoon. Päätutkimuksessa (esim. Härmälä ym. 2014), opiskelukäytänteitä oli analysoitu kielittäin ilman, että eri kielissä saatuja tuloksia olisi vertailtu keskenään. Kaikki kolme tutkittua oppimäärää olivat pitkiä eli oppilaan ensimmäiseksi vieraaksi kielekseen valitsemia. Siksi niiden opiskelu luo perustaa myöhemmille kielipinnoille ja työtaivoille. Tutkittujen kolme kielen pitkän oppimäärän tavoitteet olivat samat lukuun ottamatta hyvän osaamisen tasoa, joka oli englannissa määritelty yhtä tasoaskelmaa korkeammaksi kuin ranskassa ja ruotsissa.

Tulokset osoittivat, että strategisten taitojen (joihin tutkitut työskentely ja käytänteet kuuluvat) ja kulttuurisen osaamisen tavoitteiden toteutumisessa ei ollut huomattavia eroja kielten välillä. Täten opetussuunnitelman perusteiden tarkoittama yhtenäisyys kielten pitkien oppimäärien tavoitteiden saavuttamisessa toteutui suureksi osaksi saatujen tulosten perusteella.

Erot englannin, ruotsin ja ranskan pitkän oppimäärän opiskelijoiden opiskelukäytänteissä olivat yleensä hyvin pieniä ja jäljitettävissä suuresta otoskoosta johtuviksi. Tehtävien teon osalta havaittiin, että ruotsin kielessä oppilaat ilmoittivat tekevänsä tehtävät huomattavasti harvemmin kuin englannissa ja ranskassa. Odotuksenmukainen havainto aiemman tutkimuksen perusteella (esim. Sundqvist 2013) puolestaan oli, että englantia käytettiin koulun ulkopuolella useammin kuin ruotsia ja ranskaa. Erot opiskelukäytänteissä oppilaan sukupuolen mukaan ryhmiteltyinä jäivät pieniksi, eivätkä yleensä olleet käytännössä merkitseviä.

Opiskelukäytänteiden yhteydestä kielitaitoon (puhuminen ja kirjoittaminen) mielenkiintoisin tulos oli tehtävien teon säännöllisyyden ja kirjoittamis- ja puhumistaidon välinen yhteys ranskan kielessä. Ranskassa myös opettajan antama kannustus kielen käyttöön vapaa-ajalla oli yhteydessä oppilaiden puhumistaitoon ja heikosti myös kirjoittamistaitoon. Englannissa tätä yhteyttä ei ollut. Sen sijaan vapaa-ajan kielen käytöllä, esimerkiksi englanninkielisten televisio-ohjelmien ja videokatkelmien katselemisella oli yhteyttä oppilaan kielitaitoon. Ruotsissa puolestaan kielen käyttäminen sukulaisten ja ystävien kanssa oli yhteydessä sekä puhumis- että kirjoittamistaitoon. Havainto selittynee sillä, että ruotsin käyttö koulun ulkopuolella liittyy usein henkilöiden väliseen kanssakäymiseen kotimaan kaksikielisissä ympäristöissä, kun taas englantia on tyypillisesti käytössä mediavälitteisissä kansainvälisissä yhteyksissä.

Lähes olemattomat erot koulun opiskelukäytänteiden harjoittamisessa olivat osin odotuksenmukaisia aiemman tutkimuksen perusteella (esim. Härmälä ym. 2014; Hildén & Rautopuro 2014; Härmälä & Huhtanen 2014). Kielestä riippumatta tavallisimmat työtavat oppitunneilla olivat oppikirjan suulliset pariharjoitukset ja sanakokeet, vaikka esimerkiksi englannin tunneilla olisi mahdollista hyödyntää oppilaiden kielitai-

toa ja mielenkiinnon kohteita nykyistä monipuolisemmin (esim. media). Vaikka viestinnälliseen tehtävään perustuva kielenopetus on ollut opettajankoulutuksen valtavirtaa jo parin vuosikymmenen ajan, sen tunnusmerkkien toteutuminen käytännössä on vaihtelevaa, kuten tämäkin tutkimus osoitti. Kansainvälisestikin todettu ilmiö opettajien menetelmällisten ihanteiden ja arkikäytännössä harjoittamien opetustapojen välillä (Macaro 2001) pätee myös Suomessa sekä kansallisten oppimistulosarviointien (Härmälä ym. 2013; Hildén & Rautopuro 2014) että muun tutkimuksen valossa (esim. Harjanne, Reunamo & Tella 2015: 921; Luukka ym. 2008).

Toisaalta myös erilaisten opiskelustrategioiden tehokkuus ja opetettavuus riippuvat monista tilannetekijöistä eikä ole likikään opettajan vallassa, miten oppilaat ohjattuinaakaan lopulta haluavat opiskella ja mitä käytänteitä omaksua. Esimerkiksi Plonskyn (2011) suorittama meta-analyysi osoitti vain pieniä tai kohtalaisia vaikutuksia strategiaopetuksen ja oppimistulosten välillä.

Suullisten pariharjoitusten suosio suomalaisessa kielenopetuksessa on vankka kaikissa kielissä ja oppimäärissä (Hildén ym. 2014). Niiden yhteydet oppilaan kielitaitoon tai muihin tavoitteisiin ovat kuitenkin satunnaisia, eivätkä ne aina ole tehokkaita edes suullisen kielitaidon kehittymisen kannalta. Esimerkiksi suomenkielisissä kouluissa tehdään enemmän paritöitä englannin tunneilla kuin ruotsinkielisissä kouluissa, mutta ruotsinkielisten oppilaiden englannin taito oli silti parempi kuin suomenkielisissä kouluissa (Härmälä ym. 2014). Paritehtävät tarjoavat toki peruslähtökohdan vuorovaikutteiselle opiskelulle ja vertaistuelle, mutta tehtävien laatu ja järjestäminen kaipaisivat pedagogista terävöittämistä. Huomion arvoista olisi tutkia, mikä on erityyppisten ja eri määrin ohjattujen pariharjoitusten todellinen vaikuttavuus esimerkiksi kuuntelu- ja myötäilytaitoihin, ilmaisuvaraston rikastumiseen ja viestintärutiinien automaattistumiseen.

Koulun ulkopuolisten käytänteiden osalta oli odotettua, että englantia käytetään huomattavasti enemmän kuin muita kieliä. Yllättävä kielikohtainen tulos oli kuitenkin, että yhdeksäsluokkalaiset käyttivät ruotsia lähes yhtä vähän koulun ulkopuolella kuin ranskaa. Muiden kielten kuin englannin käyttöä koulun ulkopuolella tulisi tukea ohjaamalla sitä tarkemmin: tunneilla tehtäville tulisi asettaa selkeä tavoite ja aikaraja, kotiläksyiksi pitäisi antaa nykyistä enemmän ja monimuotoisempia tiedonhankintatehtäviä verkossa tai lähiympäristössä esimerkiksi kaksikielisillä seuduilla. Multimedia ja sisältöjen tuottaminen ovat monien oppilaiden arkea, jonka koulu ja sen kielenopetus ilmeisesti vielä paljolti sivuuttavat ymmärtämättömyyttään tai osaamattomuuttaan. Uusimmat opetussuunnitelman perusteet kiinnittävät paljon huomiota oppilaan tieto- ja viestintätekniisiin taitoihin, kulttuuriseen sensitiivisyyteen ja itseohjautuvuuden kehittymiseen. Nämä kaikki koskevat myös kielten opetusta mitä suurimmassa määrin, ja kaikkien kohentamiseksi on paljon tehtävää sekä tutkituissa että muissa perusopetuksessa opetettavissa kielissä.

Tässä raportoitu tutkimus antoi yhden näkökulman oppilaiden kokemuksiin vieraiden kielten tuntityöskentelystä. Vaikka kyselylomakkeen osiot keskittyivät pääosin varsin perinteisiin vieraiden kielten opiskelukäytänteisiin, nousi tuloksista kuitenkin esiin asioita, joita opettajankoulutuksessa tulisi pohtia. Miten esimerkiksi ruotsin kielen opiskelu toisena kotimaisena kielenä saataisiin paremmin nivottua oppilaiden vapaa-aikaan? Miten kielten oppitunneilla uskallettaisiin hyödyntää entistä ennakkoluulottomammin erilaisia tekniikan suomia mahdollisuuksia ja sitä kautta ohjata oppilaita itsenäiseen opiskeluun ja vastuunottoon omasta oppimisestaan? Kuten englannin tulokset osoittivat, ei tuntityöskentely juurikaan näy oppilaiden kielitaidossa, vaan kieltä opitaan itselle mielekkäissä, autenttisissa kielenkäyttötilanteissa. Koulun ulkopuolella opitun kielitaidon tunnustaminen voi osaltaan olla edesauttamassa monikielisyyden kehittymistä. Oppilaiden itsensä valitsemien tekstien ja materiaalien pohjalta syntyvien tilanteiden luominen ja niiden vaikuttavuuden analyysi tulevat jatkossa olemaan kielikasvatustutkimuksen keskeisiä aiheita.

Kirjallisuus

- Aljaafreh, A. & P. Lantolf 1994. Negative feedback as regulation and second language learning in the zone of proximal development. *The Modern Language Journal*, 78 (4), 465–483. DOI: 10.1111/j.1540-4781.1994.tb02064.x.
- Cohen, A. D. & E. Macaro (toim.). 2007. *Language learner strategies: thirty years of research and practice*. Oxford: Oxford University Press.
- Donato, R. & D. McCormick 1994. A sociocultural perspective on language learning strategies: the role of mediation. *The Modern Language Journal*, 78 (4), 453–464. DOI: 10.1111/j.1540-4781.1994.tb02063.x
- Ellis, R. 1984. *Classroom second language development: a study of classroom interaction and language acquisition*. Oxford: Pergamon.
- Ellis, R. 2012. *Language teaching research and language pedagogy*. Malden, Mass: Wiley-Blackwell.
- Ellis, R. & N. Shintani 2014. *Exploring language pedagogy through second language acquisition research*. Abingdon, Oxon: Routledge. Saatavilla osoitteessa <http://helka.linneanet.fi/cgi-bin/Pwebrecon.cgi?BBID=2659411>.
- European Commission 2012. *First European survey on language competences: Final report*. Directorate-General for Education and Culture. Luxembourg: Publications Office of the European Union.
- Gass, S. 2005. Task-based interactions in classroom and laboratory settings. *Language Learning*, 55 (4), 575–611. DOI: 10.1111/j.0023-8333.2005.00318.x
- Harjanne, P. & S. Tella 2011. Kieliluokkien todellisuus kielenopettajan näkökulmasta: KIELO-hankkeen alustavia tutkimustuloksia. Teoksessa L. Tainio, K. Juuti, A. Kallioniemi, P. Seitamaa-Hakkarainen & A. Uitto (toim.) *Näkökulmia tutkimusperustaiseen opetukseen*. Suomen Ainedidaktisen Tutkimusseuran Julkaisuja. Ainedidaktisia tutkimuksia, 1, 97–113.
- Harjanne, P., J. Reunamo & S. Tella 2015. Finnish foreign language teachers' views on teaching and study reality in their classes: The KIELO project's rationale and results. *Journal of Language Teaching and Research*, 6 (5), 913–923. DOI: <http://dx.doi.org/10.17507/jltr.0605.02>.

- Harjunen, E., & J. Rautopuro 2015. *Kielenkäytön ajattelua ja ajattelun kielentämistä. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2014: keskiössä kielentuntemus ja kirjoittaminen*. Helsinki: Kansallinen koulutuksen arviointikeskus. Saatavilla osoitteessa <http://helka.linneanet.fi/cgi-bin/Pwebrecon.cgi?BBID=2840876>.
- Hildén, R., M. Härmälä, J. Rautopuro, M. Huhtanen, M. Puukko & C. Silverström 2014. *Kielten oppimistulokset perusopetuksen päättövaiheessa 2013* (tiivistelmä). Informaatioaineistot 2014:1. Helsinki: Kansallinen koulutuksen arviointikeskus/Opetushallitus.
- Hildén, R. & J. Rautopuro. 2014. *Ruotsin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Helsinki: Kansallinen koulutuksen arviointikeskus. Saatavilla osoitteessa <http://helka.linneanet.fi/cgibin/Pwebrecon.cgi?BBID=2669581>
- Hildén, R. 2000. *Att tala bra, bättre och bäst*. Tutkimuksia 217. Helsingin yliopiston opettajankoulutuslaitos: Helsinki.
- Hildén, R. 2009. Transforming language curricula through a research and development project – a case from Finland. Teoksessa E. Ropo & T. Autio (toim.) *International conversations on curriculum studies: subject, society and curriculum*. Rotterdam: Sense Publishers, 235–256.
- Hildén, R. & O. Salo 2011. *Kielikasvatus tänään ja huomenna: Opetussuunnitelmat, opettajan-koulutus ja kieltenopettajan arki*. Helsinki: WSOYpro.
- Härmälä, M., & M. Huhtanen 2014. *Ranskan kielen A- ja B-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Helsinki: Kansallinen koulutuksen arviointikeskus. Saatavilla osoitteessa <http://helka.linneanet.fi/cgibin/Pwebrecon.cgi?BBID=2669590>.
- Härmälä, M., M. Huhtanen & M. Puukko 2014. *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Helsinki: Kansallinen koulutuksen arviointikeskus. Saatavilla osoitteessa <http://helka.linneanet.fi/cgibin/Pwebrecon.cgi?BBID=2669591>.
- Julin, S. & J. Rautopuro 2016. *Läksyt tekijäänsä neuvovat. Perusopetuksen matematiikan oppimistulosten arviointi 9. vuosiluokalla 2015*. Julkaisut 20: 2016. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Kaikkonen, P. 2004. *Vierauden keskellä: vierauden, monikulttuurisuuden ja kulttuurienvälisen kasvatuksen aineksia*. Jyväskylä: Jyväskylän yliopisto, opettajankoulutuslaitos.
- Kohonen, V. & U. Pajukanta 2003. *Eurooppalainen kielisalkku 2: EKS-projektin päätösvaiheen tuloksia*. Tampere: Tampereen yliopisto.
- Kristiansen, I. 1998. *Tehokkaita oppimisstrategioita: esimerkinä kielet*. Helsinki: WSOY.
- Kumaravadivelu, B. 2006. *Understanding language teaching: from method to postmethod*. Mahwah NJ: Lawrence Erlbaum Associates.
- Kuukka, K. & J. Metsämuuronen 2016. *Perusopetuksen päättövaiheen suomi toisena kielenä (S2) -oppimäärän oppimistulosten arviointi 2015*. Julkaisut 13: 2016. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Lappalainen, H.-P. 2011. *Sen edestään löytää. Äidinkielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2010*. Koulutuksen seurantaraportit 2011: 2. Helsinki: Opetushallitus.
- Larsen-Freeman, D. 2015. Research into practice: grammar learning and teaching. *Language Teaching*, 48 (2), 263–280.
- Larzen, E. 2005. *In pursuit of an intercultural dimension in EFL-teaching: exploring cognitions among Finland-Swedish comprehensive school teachers*. Åbo: Åbo Akademis förlag.
- Leppänen, S., A. Pitkänen-Huhta, T. Nikula, S. Kytölä, T. Törmäkangas, K. Nissinen, L. Kääntä, T. Virkkula, M. Laitinen, P. Pahta, H. Koskela, S. Lähdesmäki & H. Jousimäki 2009. *Kansallinen kyselytutkimus englannin kielestä Suomessa: käyttö, merkitys ja asenteet*. Jyväskylä Studies in Humanities; 132. Jyväskylä: Jyväskylän yliopisto.
- Opetushallitus 1994. *Lukion opetussuunnitelman perusteet 1994*. Helsinki: Opetushallitus.

- Luukka, M.-R., S. Pöyhönen, A. Huhta, P. Taalas, M. Tarnanen & A. Keränen 2008. *Maailma muuttuu – mitä tekee koulu. Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla*. Jyväskylän yliopisto, soveltavan kielentutkimuksen keskus. Saatavilla osoitteessa <https://jyx.jyu.fi/dspace/handle/123456789/36607>.
- Macaro, E. 2001. Analysing student teachers' codeswitching in foreign language classrooms: theories and decision making. *The Modern Language Journal*, 85 (4), 531–548. DOI: 10.1111/0026-7902.00124
- Ohta, A. S. 2000. Rethinking recasts: a learner-centred examination of corrective feedback in the Japanese language classroom. Teoksessa J. Hall & L. Verplaeste (toim.) *Second and foreign language learning through classroom interaction*. Mahwah, NJ: Erlbaum, 47–71.
- Olsson, E. & L. K. Sylvén 2011. Extramural English and academic vocabulary. A longitudinal study of CLIL and non-CLIL students in Sweden. *Apples – Journal of Applied Language Studies*, 9, 77–103. DOI: <http://dx.doi.org/10.17011/apples/urn.201512234129>.
- Opetushallitus 2004. *Perusopetuksen opetussuunnitelman perusteet 2004*. Helsinki: Opetushallitus.
- Opetushallitus 2014. *Perusopetuksen opetussuunnitelman perusteet 2014*. Määräykset ja ohjeet 2014:96. Helsinki: Opetushallitus.
- Oxford, R., J. Rubin, A. Uhl Chamot, K. Schramm, R. Lavine, P. Gunning & C. Nel 2014. The learning strategy prism: perspectives of learning strategy experts. *System*, 43, 30–49. DOI: <http://dx.doi.org/10.1016/j.system.2014.02.004>.
- Pica, T. & C. Doughty 1985. The role of group work in classroom second language acquisition. *Studies in Second Language Acquisition*, 7 (2), 233–248. DOI: <https://doi.org/10.1017/S0272263100005398>.
- Plonsky, L. 2011. The effectiveness of second language strategy instruction: a meta analysis. *Language Learning*, 61 (4), 993–1038. DOI: 10.1111/j.14679922.2011.00663.x.
- POPS 1970. Peruskoulun opetussuunnitelmakomitean mietintö II. *Opetussuunnitelman perusteet: oppiaineiden opetussuunnitelmat*. Helsinki: Opetusministeriö.
- Prabhu, N. S. 1987. *Second language pedagogy*. Oxford: Oxford University Press.
- Quakrim-Soivio, N. & J. Kuusela 2012. *Historian ja yhteiskuntaopin oppimistulokset perusopetuksen päättövaiheessa 2011*. Koulutuksen seurantaraportit 2012: 3. Helsinki: Opetushallitus.
- Seedhouse, P. 2005. "Task" as research construct. *Language Learning*, 55 (3), 533–570. DOI: 10.1111/j.0023-8333.2005.00314.x
- Scherer, G. & M. Wertheimer 1964. *A psycholinguistic experiment in foreign-language teaching*. McGraw-Hill.
- Storch, N. 2002. Patterns of interaction in ESL pair work. *Language Learning*, 52 (1), 119–158. DOI: 10.1111/1467-9922.00179.
- Summanen, A.-M. 2014. *Terveystiedon oppimistulokset perusopetuksen päättövaiheessa 2013*. Koulutuksen seurantaraportit 2014: 1. Helsinki: Opetushallitus.
- Sundqvist, P. 2013. Classroom vs. extramural English: teachers dealing with demotivation. *Language and Linguistics Compass*, 7 (6), 329–338. DOI:10.1111/Inc3.12031.
- Taalas, P. 2005. *Change in the making: strategic and pedagogical aspects of technology integration in language teaching. Towards multimodal language learning environments*. Jyväskylä: Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus.
- Tuokko, E. 2000. *Peruskoulun 9. vuosiluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi 1999*. Oppimistulosten arviointi 3/2000. Helsinki: Opetushallitus.
- Venäläinen, S. 2015 *Arjen tiedot ja taidot hyvinvoinnin perustana. Kotitalouden oppimistulokset perusopetuksen päättövaiheessa 2014*. Julkaisut 2015: 5. Helsinki: Kansallinen koulutuksen arviointikeskus.
- Vygotsky, L. S. 1982. *Ajattelu ja kieli*. Espoo: Weilin Göös.

LIITE 1.

Oppilaskyselyn osiot koulussa ja vapaa-ajalla harjoitetuista opiskelukäytännöistä.

Yläkoulun (vuosiluokat 7-9) englannin/ruotsin/ranskan tunneilla...	1 = ei koskaan, 2 = vain harvoin, 3 = joskus, 4 = usein, 5 = lähes aina
12. pidämme suullisia esityksiä englanniksi/ruotsiksi/ranskaksi.	
13. harjoitteleminen sanastoa tekemällä omia lauseita suullisesti.	
14. harjoitteleminen kielioppiasiaa tekemällä omia lauseita.	
15. teemme oppikirjasta suullisia pariharjoituksia.	
16. opettaja käyttää englannin/ruotsin/ranskan kieltä puhuessaan koko luokalle.	
17. katsomme/kuuntelemme englannin/ruotsin/ranskankielisiä elokuvia, lauluja tms.	
18. luemme englannin/ruotsin/ranskankielisiä lehtiä tai kirjoja.	
19. tallennamme puheharjoituksia videolle.	
20. suunnittelen omaa työskentelyäni (esim. Mitä opiskelen ensi viikolla?)	
21. kokoamme kielisalkkua tai kirjoitamme opiskelupäiväkirjaa.	
22. harjoitteleminen kielioppia tietokoneella (pelit tms.).	
23. käytämme internetiä tiedonhakuun.	
24. saamme itse valita, mitä kotitehtäviä teemme.	
25. annamme palautetta toisen oppilaan englannin/ruotsin/ranskan osaamisesta (esim. puhetehtävässä).	
26. arvioin omaa englannin/ruotsin/ranskan kielen osaamistani (esim. pohtimalla, miten onnistuin, tai ehdottamalla arvosanaa).	
27. keskustelen opettajan kanssa omasta edistymisestääni.	
28. teemme sanakokeita (vastauksina yksittäiset sanat).	
29. meillä on kirjallisia kokeita päivän läksystä.	
30. opettaja kuuntelee ja antaa palautetta, kun keskustelemme parin kanssa englanniksi/ruotsiksi /ranskaksi.	
31. kokeissa on vapaata kirjallista tuottamista englanniksi/ruotsiksi/ranskaksi (ei vain käännöstä tai täydennystä).	
32. opettaja kannustaa käyttämään englantia/ruotsia/ranskaa koulun ulkopuolella.	
Kohdekielen käyttö koulun ulkopuolella	1 = ei koskaan, 2 = joskus, 3 = viikoittain, 4 = päivittäin
33. katselen englannin/ruotsin/ranskankielisiä elokuvia tai videokatkelmia.	
34. kuuntelen englannin/ruotsin/ranskankielistä musiikkia.	
35. seuraan englannin/ruotsin/ranskankielisiä keskustelupalstoja (YouTube, Facebook jne).	

36. osallistun englannin/ruotsin/ranskankielisiin verkkokeskusteluihin (blogit, Facebook, chat, Twitter jne).
37. luen englannin/ruotsin/ranskankielisiä verkkolehtiä tai muita verkkotekstejä.
38. kirjoitan englannin/ruotsin/ranskankielisiä tekstejä (tekstiviestit, runot jne.).
39. puhun englantia/ruotsia/ranskaa esim. turistien kanssa.
40. käytän englantia/ruotsia/ranskaa kavereiden tai sukulaisten kanssa.

LIITE 2.

Kuvailevat tunnusluvut kyselylomakkeen osioille kielittäin.

		Keski-arvo	Hajonta	Mediaani
Kuinka paljon käytät aikaa englannin/ruotsin/ ranskan/ruotsin/ranskan läksyihin	Englanti	1,99	0,481	2
	Ruotsi	1,97	0,444	2
	Ranska	1,96	0,523	2
Kuinka säännöllisesti teet englannin/ruotsin/ ranskan/ruotsin/ranskan opiskeluun liittyvät tehtävät	Englanti	4,07	1,030	4
	Ruotsi	1,97	0,438	2
	Ranska	3,81	1,193	4
KYS12 Pidämme suullisia esityksiä englanniksi/ ruotsiksi/ranskaksi.	Englanti	2,38	0,947	2
	Ruotsi	2,34	0,972	2
	Ranska	2,28	0,948	2
KYS13 Harjoitteleme sanastoa tekemällä omia lauseita suullisesti.	Englanti	2,71	0,971	3
	Ruotsi	2,83	0,989	3
	Ranska	2,90	1,048	3
KYS14 Harjoitteleme kielioppiasiaa tekemällä omia lauseita.	Englanti	3,23	0,970	3
	Ruotsi	3,29	0,971	3
	Ranska	3,34	1,006	3
KYS15 Teemme oppikirjasta suullisia pariharjoituksia.	Englanti	3,86	0,924	4
	Ruotsi	3,86	0,919	4
	Ranska	4,00	0,896	4

KYS16 Opettaja käyttää englannin/ruotsin/ranskan kieltä puhuessaan koko luokalle.	Englanti	3,90	1,055	4
	Ruotsi	3,85	1,000	4
	Ranska	4,07	0,895	4
KYS17 Katsomme/kuuntelemme englannin/ruotsin /ranskankielisiä elokuvia, lauluja jne.	Englanti	2,73	0,852	3
	Ruotsi	2,85	0,858	3
	Ranska	3,14	0,843	3
KYS18 Luemme englannin/ruotsin/ranskankielisiä lehtiä tai kirjoja.	Englanti	1,71	0,799	2
	Ruotsi	1,81	0,905	2
	Ranska	1,73	0,792	2
KYS19 Tallennamme puheharjoituksia videolle.	Englanti	1,15	0,474	1
	Ruotsi	1,10	0,372	1
	Ranska	1,13	0,436	1
KYS20 Suunnittelen omaa työskentelyäni.	Englanti	1,54	0,798	1
	Ruotsi	1,54	0,807	1
	Ranska	1,58	0,846	1
KYS21 Kokoamme kielisalkkua tai kirjoitamme opiskelupäiväkirjaa.	Englanti	1,13	0,505	1
	Ruotsi	1,14	0,544	1
	Ranska	1,10	0,473	1
KYS22 Harjoittelemme kielioppia tietokoneella.	Englanti	2,04	0,932	2
	Ruotsi	1,85	0,884	2
	Ranska	2,26	1,002	2
KYS23 Käytämme internetiä tiedonhakuun.	Englanti	2,19	0,916	2
	Ruotsi	1,94	0,917	2
	Ranska	2,27	0,955	2
KYS24 Saamme itse valita, mitä koti-tehtäviä teemme.	Englanti	1,44	0,657	1
	Ruotsi	1,37	0,590	1
	Ranska	1,45	0,716	1
KYS25 Annamme palautetta toisen oppilaan englannin/ruotsin/ranskan osaamisesta.	Englanti	1,74	0,879	2
	Ruotsi	1,70	0,838	1
	Ranska	1,78	0,928	2
KYS26 Arvioin omaa englannin/ruotsin/ranskan kielen osaamistani.	Englanti	2,30	0,991	2
	Ruotsi	2,20	0,971	2
	Ranska	2,24	1,002	2

KYS27 Keskustelen opettajan kanssa omasta edistymisestääni.	Englanti	1,81	0,82	2
	Ruotsi	1,83	0,809	2
	Ranska	1,85	0,827	2
KYS28 Teemme sanakokeita.	Englanti	3,65	0,990	4
	Ruotsi	3,53	0,993	4
	Ranska	3,14	1,052	3
KYS29 Meillä on kirjallisia kokeita päivän läksystä.	Englanti	1,98	0,924	2
	Ruotsi	1,88	0,84	2
	Ranska	1,71	0,796	2
KYS30 Opettaja kuuntelee ja antaa palautetta, kun keskustelemme parin kanssa englanniksi/ ruotsiksi/ranskaksi.	Englanti	2,70	1,038	3
	Ruotsi	2,72	1,059	3
	Ranska	3,05	1,048	3
KYS31 Kokeissa on vapaata kirjallista tuottamista englanniksi/ ruotsiksi/ ranskaksi.	Englanti	3,05	1,182	3
	Ruotsi	3,17	1,161	3
	Ranska	3,53	1,201	4
KYS32 Opettaja kannustaa käyttämään englantia/ruotsia /ranskaa koulun ulkopuolella.	Englanti	2,95	1,156	3
	Ruotsi	3,05	1,156	3
	Ranska	3,43	1,146	3
KYS33 Katselen englannin/ruotsin/ ranskankielisiä elokuvia tai videokatkelmia.	Englanti	3,29	0,780	3
	Ruotsi	1,66	0,594	2
	Ranska	1,73	0,560	2
KYS34 Kuuntelen englannin/ruotsin / ranskankielistä musiikkia.	Englanti	3,69	0,641	4
	Ruotsi	1,63	0,689	2
	Ranska	1,78	0,746	2
KYS35 Seuraan englannin/ruotsin/ ranskankielisiä keskustelupalstoja (YouTube, Facebook).	Englanti	3,36	0,905	4
	Ruotsi	1,41	0,702	1
	Ranska	1,32	0,605	1
KYS36 Osallistun englannin/ruotsin / ranskankielisiin verkkokeskusteluihin (blogit, Facebook jne.).	Englanti	2,49	1,120	2
	Ruotsi	1,24	0,567	1
	Ranska	1,14	0,435	1
KYS37 Luen englannin/ruotsin / ranskankielisiä verkkolehtiä tai muita verkkotekstejä	Englanti	2,36	0,978	2
	Ruotsi	1,33	0,586	1
	Ranska	1,22	0,480	1

KYS38 Kirjoitan englannin/ruotsin/ ranskankielisiä tekstejä (tekstiviestit, runot jne.)	Englanti	2,08	0,979	2
	Ruotsi	1,36	0,647	1
	Ranska	1,25	0,513	1
KYS39 Puhun englantia/ruotsia/ rans- kaa esim. turistien kanssa	Englanti	1,91	0,693	2
	Ruotsi	1,51	0,631	1
	Ranska	1,39	0,546	1
KYS40 Käytän englantia/ruotsia/ rans- kaa kavereiden tai sukulaisten kanssa.	Englanti	2,07	0,898	2
	Ruotsi	1,73	0,817	2
	Ranska	1,49	0,658	1

LIITE 3.

Taustamuuttujien frekvenssit.

Koulun kieli (vain englanti)			
		Oppilasmäärä	%
	suomi	2966	85,3
	ruotsi	510	14,7
	Yhteensä	3476	
Sukupuoli			
Kieli		Oppilasmäärä	%
Englanti	poika	1766	51.0
	tyttö	1697	49.0
	Yhteensä	3463	
Ruotsi	poika	686	40.9
	tyttö	993	59.1
	Yhteensä	1679	
Ranska	poika	318	31.1
	tyttö	703	68.9
	Yhteensä	1021	

Perusopetuksen jälkeen			
Kieli		Oppilasmäärä	%
Englanti	pyrin ensisijaisesti lukioon.	1903	55.5
	pyrin ensisijaisesti ammatilliseen koulutukseen.	1466	42.7
	hankin lisävalmiuksia jatko-opintoihin	50	1.5
	aion mennä töihin tai pitää vapaavuoden.	12	0.3
	Yhteensä	3431	
Ruotsi	pyrin ensisijaisesti lukioon.	1266	76.1
	pyrin ensisijaisesti ammatilliseen koulutukseen.	387	23.3
	hankin lisävalmiuksia jatko-opintoihin	9	0.5
	aion mennä töihin tai pitää vapaavuoden.	2	0.1
	Yhteensä	1664	
Ranska	pyrin ensisijaisesti lukioon.	924	90.6
	pyrin ensisijaisesti ammatilliseen koulutukseen.	94	9.2
	hankin lisävalmiuksia jatko-opintoihin	1	0.1
	aion mennä töihin tai pitää vapaavuoden.	1	0.1
	Yhteensä	1020	
Onko äiti/isä ylioppilas			
Kieli		Oppilasmäärä	%
Englanti	ei kumpikaan	1038	38.0
	äiti	842	30.8
	isä	190	7.0
	molemmat	661	24.2
	Yhteensä	2731	
Ruotsi	ei kumpikaan	305	21.9
	äiti	409	29.4
	isä	139	10.0
	molemmat	538	38.7
	Yhteensä	1391	
Ranska	ei kumpikaan	89	9.6
	äiti	230	24.8

166 YHDEKSÄSLUOKKALAISTEN OPISKELUKÄYTÄNTEET

	isä	63	6.8
	molemmat	546	58.8
	Yhteensä	928	

Heidi Harju-Luukkainen¹, Mirja Tarnanen² & Kari Nissinen²

¹Helsingin yliopisto, ²Jyväskylän yliopisto

Monikieliset oppilaat koulussa: eri kieliryhmien sisäinen ja ulkoinen motivaatio sekä sen yhteys matematiikan osaamiseen PISA 2012 -arvioinnissa

For the first time in PISA, the Finnish students with an immigrant background were oversampled in 2012. This makes the data more representative and increases the precision of statistical analyses. In this article 691 first-generation and 603 second-generation students with an immigrant background are studied closer as a group. We examine the level of students' intrinsic and extrinsic motivation and its relation to the educational outcome in PISA 2012 from the perspective of students' linguistic background. According to the results, there are statistically significant differences in students' motivation level in different language groups. All of the non-native language groups represented in PISA 2012 had a higher level of motivation compared to native language groups. Secondly, the results reveal that the correlation between motivation and educational outcome within different language groups is not straightforward. Finally, we discuss some implications concerning the role of motivation and native languages of the student with migrant background in learning and how learning could be supported by taking these factors into consideration when teaching in multicultural and multilingual settings.

Keywords: motivation, PISA assessment, students with migrant background, L1 of students with migrant background

Asiasanat: motivaatio, PISA-arviointi, maahanmuuttajataustaiset oppilaat, oma äidinkieli



1 Johdanto

Maahanmuuttajien määrä on kymmenkertaistunut vuosien 1990 ja 2010 välillä Suomessa (Tilastokeskus 2014), mikä näkyy myös suomalaisissa kouluissa kulttuurisen monimuotoisuuden ja monikielisuuden rikastumisena (OPH 2016). Maahanmuuttajataustaisiin lapsiin ja nuoriin kohdistuvaa tutkimusta ja siitä tulevaa tietoa tarvitaan yhä enenevässä määrin esimerkiksi koulujärjestelmän tarpeisiin, koska koululla on merkittävä asema maahanmuuttajataustaisten oppilaiden kouluun ja yhteiskuntaan integroimisessa. Kulttuurinen monimuotoisuus ja kielitietoisuus, joilla viitataan koulun kielellisen moninaisuuden tukemiseen ja kielitietoisuuden vahvistamiseen yli oppiainerajojen, ovatkin keskeisiä koulun toimintakulttuurin kehittämistä ohjaavia periaatteita. Peruskoulun tehtävänä on oppilaiden oman kulttuuri- ja kieli-identiteetin rakentamisen tukeminen. (OPH 2014.) Perusopetuksen opetussuunnitelman perusteiden (2014: 28) mukaan kielitietoisessa yhteisössä keskustellaan kieliin ja kielyhteisöihin kohdistuvista asenteista ja ymmärretään kielen keskeinen merkitys oppimisessa, vuorovaikutuksessa ja yhteistyössä sekä identiteettien rakentumisessa ja yhteiskuntaan soisaalistumisessa. Tutkimusten valossa vahvan etnisen identiteetin on osoitettu olevan yhteydessä parempaan koulumenestykseen ja oppimistuloksiin (esim. Altschul, Oyserman & Bybee 2006; Zirkel 2008). Koulun monikulttuurisuuden ja monikielisuuden vahvistamiseen pyrkiminen on suotavaa, koska näin koulu voi tukea oppilaidensa monikulttuuristen ja monikielisten identiteettien kehittymistä (Cummins 2000; Hélot & Young 2002; Martin-Jones 2007).

Koulun opetuskieli on maahanmuuttajataustaisille oppilaille sekä oppimisen kohde että väline eri oppiaineissa, joten niin matematiikan kuin muidenkin oppiaineiden opiskelussa on pedagogisesti haastavaa, miten edetä arkikielestä kohtia käsitteellisen ajattelun kieltä, kun oppilaiden kielelliset resurssit vaihtelevat (esim. Vollmer 2006; Mercer & Littleton 2007). Matematiikan tarkastelu erikielisten oppilaiden näkökulmasta on kiinnostavaa, koska arkiajattelussa luonnollisen kielen roolia, kuten myös oppilaiden oman äidinkielen taitoa voidaan väheksyä. Kuitenkin myös matematiikan opettajalta edellytetään kykyä tunnistaa kielten merkitys oppimisen mahdollistajana (ks. Vollmer 2006; Civil 2009).

Kulttuurisen monimuotoisuuden ja oppilaiden monikielisuuden huomioiminen ja tukeminen suomalaisissa kouluissa kaipaakin vielä kehittämistä. Kansallisen arviointikeskuksen raportin mukaan kehitystyössä tulisi panostaa koulujen ja oppilaitosten henkilöstön osaamisen ja asenteiden kehittämiseen monimuotoisuuden edistämiseksi (Kuukka, Ouakrim-Soivio, Paavola & Tarnanen 2015). Tämän lisäksi koulumenestyksen näkökulmasta sekä ensimmäisen että toisen polven maahanmuuttajataustaisten oppimistulokset ovat keskimäärin valtaväestöä heikompia (ks. Harju-Luuk-

kainen, Nissinen, Sulkunen, Suni, & Vettenranta 2014; Valtiontalouden tarkastusvirasto 2015), joskin erojen on joissain tutkimuksissa havaittu pienentyvän taustamuuttujien tilastollisen vakioinnin jälkeen (Kilpi 2010; Kilpi-Jakonen 2012). Koulutus- ja työelämäpolkujen sujuvoittamiseksi varhaiskasvatuksessa ja perusopetuksessa tehtävä työ on keskeisessä asemassa, sillä maahanmuuttajataustaiset nuoret ovat keskimäärin muita huonommassa asemassa koulutuksessa ja työelämässä, minkä vuoksi heillä on riski syrjäytymiseen (esim. Teräs, Lasonen & Sannio 2010; Rinne & Tuittu 2011). Myös maahanmuuttajaperheiden sosioekonominen ja kulttuurinen pääoma on keskimäärin melko heikko (Harju-Luukkainen ym. 2014).

Maahanmuuttajataustaisista oppilaista ei voida kuitenkaan puhua yhtenä yhtenäisenä ryhmänä. Heidän tilanteensa ja oppimisen tuen tarpeensa voivat vaihdella esimerkiksi sen mukaan, ovatko he ensimmäisen vai toisen sukupolven maahanmuuttajia. Tämän lisäksi merkityksellistä on se, missä vaiheessa opintopolkuaan nuori tulee maahan, onko hänen kotikielensä lähellä tulomaan kieltä sekä mikä on lähtömaan koulutuksen taso ja millaiset standardit siellä ovat (Heath & Kilpi-Jakonen 2012; Kilpi-Jakonen 2012; Harju-Luukkainen ym. 2014). Näin ollen lapset ja nuoret, jotka tulevat Suomeen kouluikäisinä ja aloittavat koulunkäynnin omaa ikäryhmäänsä myöhemmin, voivat kohdata yhtäaikaisesti monia siirtymiä. Nämä siirtymät voivat liittyä maahanmuuttoon, yhteiskuntaan ja kulttuuriin, lapsuuden ja nuoruuden kehityksellisiin siirtymiin sekä toiselle koulutusasteelle siirtymiseen (Alitolppa-Niitamo 2003; Teräs & Kilpi-Jakonen 2013).

Riittävän kielitaidon ja oppimisvalmiuksien saavuttaminen 16 ikävuoteen mennessä voi varsinkin myöhään maahantulleiden nuorten kohdalla olla valmistavasta opetuksesta huolimatta saavuttamaton tavoite. Kuitenkin Kuukan ja Metsämuurosen (2016) mukaan nuorten maahanmuuttajataustaisten nuorten kielitaito peruskoulun päättövaiheessa osoittautui varsin hyväksi. Oppilaista 87 prosenttia saavutti arvioinnissa vähintään hyvän osaamisen tason. Toisaalta hyvästä kielitaidon tasosta huolimatta PISA 2012 -tutkimuksen mukaan noin 50 prosenttia ensimmäisen ja toisen sukupolven maahanmuuttajataustaisista nuorista jäi arviointialueissa (luonnontieteet, lukutaito ja matematiikka) minimitason alapuolelle (Harju-Luukkainen ym. 2014). Toisen sukupolven maahanmuuttajataustaisten oppilaiden voisi olettaa jo integroituneen kouluun. Kaiken kaikkiaan kehittämisen varaa on useilla osa-alueilla, sillä maahanmuuttajataustaiset oppilaat kokevat tulevansa kiusatuksi taustansa vuoksi, heillä on erityistuen tarpeita ja oppimistulokset ovat suomen- ja ruotsinkielisiä heikompia (ks. esim. Sinkkonen, Kyttälä, Karvinen & Aunio 2011; Harju-Luukkainen ym. 2014).

2 Motivaatio ja oppiminen

2.1 Motivaatio ja itsemäärämisteoria

Erilaisista teoreettisista ja käsitteellistä lähestymistavoista huolimatta motivaatiota voitaneen pitää yhtenä keskeisimmistä tekijöistä oppimisessa. Opetuksen näkökulmasta katsottuna opetus ei ole tehokasta, jos oppilaat eivät ole kiinnostuneet opetuksesta ja jos he eivät sitoudu oppimiseen. Oppilaan toiminta kytkeytyy motivaatioon sitä kautta, että tämä on yksi keskeisistä ominaisuuksista, joka näkyy oppilaan aktiivisuutena (Nurmi 2013). Motivaatiota on tutkittu paljon ja sitä on lähestytty lukuisista eri näkökulmista, esimerkiksi sekä oppimispsykologian (ks. esim. Wigfield & Cambria 2010) ja kasvatusfilosofian alueella (esim. Dewey 1938) että eri oppiaineiden, kuten matematiikan (Hannula 2006; Lukin 2013; Pantziaram & Philippou 2015) ja toisen/vieraan kielen oppimisen (esim. Dörnyei 2003; Dörnyei & Ushioda 2011) viitekehyksissä.

Lähestymme tässä artikkelissa motivaatiota itsemäärämisteorian kautta, jota pidetään yhtenä vaikuttavimmista lähestymistavoista oppimispsykologian alueella (Dörnyei 2003). Ryanin ja Decin (2000) mukaan motivaatiota tarkasteltaessa on otettava huomioon sen monitahoinen luonne, sillä ihmisillä voi olla eri määrä motivaatiota (motivaation taso), mutta heillä voi olla myös erityyppisiä motivaatioita (motivaatio-orientaatio). Näistä jälkimmäiseen liittyvät asenteet ja tavoitteet, jotka säätelevät sitä, miksi toimimme. Esimerkiksi oppilas voi olla motivoitunut tekemään kotiläksyjä omasta mielenkiinnosta tai uteliaisuudesta tai vaihtoehtoisesti siksi, että hän haluaa saavuttaa opettajan tai vanhempiensa hyväksynnän (Ryan & Deci 2000: 54). Motivaatio voi olla joko sisäistä tai ulkoista.

Ulkoisesti motivoitunut oppilas on kiinnostunut tekemään asioita, koska muut henkilöt tai hänen sosiaalinen ympäristönsä odottavat sitä häneltä tai asioiden tekemisestä seuraa ulkoisia palkintoja. Sisäisesti motivoituneen oppilaan toimintaa puolestaan ohjaa kiinnostus tai toiminnasta seuraava hyvänolon tunne ja nautinto (Ryan & Deci 2000). Vaikka sisäisen motivaation katsotaan liittyvän laadukkaaseen oppimiseen ja luovuuteen, Ryan ja Deci kuitenkin välttävät perinteistä näiden kahden motivaation vastakkainasettelua ja nostavat esiin sen, että ulkoista motivaatiota voi olla erilaista. Oppilaat voivat työskennellä ulkoisen motivaation ohjaamina vastustaen ja kiinnostumattomina tai yhtä hyvin myös auliisti, jolloin he näkevät aktiviteetin hyödyllisenä tai siinä tekemisen arvoisena. On siis mahdollista, että alun perin ulkoinen syy toimia voi muuttua sisäistetyksi eri tavoin; esimerkiksi halu välttää tekemättömyydestä seuraavaa syyllisyyttä tai alkujaan ulkoisesti määräytyvä tekeminen alkaa tuntua oppilaasta tärkeältä oman taidon kehittymisen ja oman tulevaisuuden kannalta (Ryan & Deci 2000: 61–62). Tämän vuoksi on tärkeää, että motiivit olisivat mahdollisimman paljon sisäistet-

tyjä ja osa itseä, koska silloin ne olisivat enemmän itsemääräytyviä eli autonomia ja niistä seuraisi parempia oppimistuloksia.

2.2 Motivaatio ja monikielinen ja -kulttuurinen koulu oppimisympäristönä

Se, millainen ympäristö koulu on monikieliselle ja eri kulttuurista tulevalle oppilaalle, on merkityksellistä, koska kognitiivisten motivaatiotekijöiden kannalta sosiaalinen ympäristö voi joko luoda sisäistä motivaatiota edistävän tai estävän ilmapiirin (Ryan & Deci 2000: 12). Monikulttuurisella koululla tarkoitetaan, että kaikilla oppilailla on kouluoppimiseen tasavertaiset mahdollisuudet sukupuoleen, sosiaaliluokkaan tai etnisiin ja kulttuurisiin ominaisuuksiin katsomatta (Banks 2010). Myös monikielisyyden tukeminen on keskeistä sekä koulumenestyksen että persoonallisen kasvun ja identiteetin kehittymisen näkökulmasta (esim. Baker 2000; Cummins 2000; Bialystok & Feng 2009). Kielen oppiminen koulussa ja mahdollisuus oman äidinkielen käyttöön yli oppiainerajojen voi tukea oppimista, sillä kielet kytkeytyvät merkitysten ja vuorovaikutuksen kautta kaikkeen oppimiseen (esim. Walqui 2006; Mercer & Littleton 2007). Perusopetuksen opetus suunnitelman perusteiden mukaan monikielisiä oppilaita tulisi rohkaista käyttämään osaamistaan kieliä oppitunneilla ja koulun toiminnassa yleisesti. Myös oman äidinkielenopetuksen tarjoamista suositellaan mahdollisuuksien mukaan. (OPH 2014: 87.)

Kaiken kaikkiaan maahanmuuttajataustaiset oppilaat näyttävät viihtyvän koulussa paremmin ja ovat motivoituneempia kuin kantaväestön oppilaat (Räsänen & Kivirauma 2011; Leinonen 2013; Harju-Luukkainen ym. 2014). Muun muassa Kuuselan ja Etelälähdén (2008) tutkimuksessa suurin ero oppilaiden, joilla oli maahanmuuttajatausta ja joilla ei ollut, löytyi asteikon positiivisesta päästä. Noin 33 prosenttia maahanmuuttajataustaisista oppilaista ilmoitti viihtyvänsä koulussa erittäin hyvin, kun taas vastaava luku oppilaiden osalta, joilla ei ollut maahanmuuttajataustaa, oli 23 prosenttia. Lähtökohdat oppimiselle ovat siis kouluviihtyvyyden näkökulmasta hyvät, sillä maahanmuuttajataustaiset oppilaat pitävät koulua tärkeänä ja kokevat opettajasuhteet myönteisesti. Toisaalta niissä kouluissa, joissa on enemmän maahanmuuttajataustaisia oppilaita, esiintyy myös enemmän kiusaamista, ja oppilaista osa kokee kiusaamista huonon suomen kielen taitonsa tai taustansa vuoksi. (Räsänen & Kivirauma 2011; Leinonen 2013.)

Suomalaisessa koulutusjärjestelmässä on kehittämisen varaa usealla saralla, ja tutkimuksellista tietoa maahanmuuttajataustaisten nuorten oppimisesta ja sen tukemisesta tarvitaan lisää. Toistaiseksi ainoa Suomessa toteutettu laajempi kvantitatiivinen ja kansainvälinen tutkimus, jonka avulla on ollut mahdollista saada vertailukelpoista tietoa maahanmuuttajataustaisten nuorten oppimistuloksista ja niihin liittyvistä taustatekijöistä, on PISA 2012 -arvioinnin yhteydessä kerätty aineisto (ks. Harju-Luukkainen

ym. 2014). Siinä Opetus- ja kulttuuriministeriön toimeksiantona otettiin yliotos 15-vuotiaista maahanmuuttajataustaisista nuorista tilastollisten päätelmien tarkkuuden parantamiseksi. PISA 2012 -tutkimuksen mukaan maahanmuuttajataustaisten nuorten oppimistulokset olivat heikot niin matematiikassa (vuoden 2012 pääarviointialue), lukutaidossa kuin luonnontieteissäkin. Toisaalta heidän motivaationsa (matematiikan) oppimista kohtaan oli poikkeuksellisen korkea, mutta motivaation yhteys oppimistuloksiin oli olematon (Harju-Luukkainen ym. 2014). Näyttäisikin siltä, että oppimistulosten taustalla on muita tekijöitä, jotka vaikuttavat siihen, ettei maahanmuuttajataustaisten nuorten motivaation ja oppimistulosten välillä ole selkeää yhteyttä. Kuitenkin tarkasteltaessa koko tietyn ikäryhmän oppilasjoukkoa Suomessa oppimista edistävillä aseenteilla on vahva korrelaatio oppimistuloksiin (Kärnä ja Rautopuro 2013). Sama pätee myös oppilaisiin, joilla ei ole maahanmuuttajataustaa PISA 2012 -tutkimuksessa, sillä korrelaatio PISA-pistemäärien ja sisäisen sekä ulkoisen motivaation välillä oli vahva.

Vuoden 2012 PISA-taustakyselyssä suomalaisnuorilta mitattiin heidän sisäistä ja ulkoista motivaatiotaan. Tulosten mukaan sekä sisäinen että ulkoinen motivaatio oli maahanmuuttajataustaisilla nuorilla selvästi korkeammalla tasolla kuin kantaväestöön kuuluvilla nuorilla (Harju-Luukkainen ym. 2014: 57–60). Kuten edellä todettiin, näiden nuorten motivaatiolla ei kuitenkaan ollut yhteyttä heidän menestykseensä matematiikassa. Sen sijaan kantaväestöön kuuluvilla nuorilla oli matematiikan osaamisen ja motivaation välillä selkeä korrelaatio. Sisäisen motivaation korrelaatio matematiikan PISA-pistemäärän kanssa oli 0,37, kun ulkoisen motivaation vastaava korrelaatio oli 0,34. Tämä tarkoittaa, että sisäisellä motivaatiolla voidaan selittää 14 prosenttia ja ulkoisella motivaatiolla 12 prosenttia kantaväestöön kuuluvien nuorten matematiikan osaamisen vaihtelusta. Maahanmuuttajataustaisilla nuorilla motivaation korrelaatiot matematiikan pistemäärän kanssa olivat hyvin heikot tai suorastaan olemattomat: toisen sukupolven maahanmuuttajataustaisilla nuorilla ne olivat 0,16 (sisäinen motivaatio) ja 0,13 (ulkoinen motivaatio) ja ensimmäisen sukupolven nuorilla 0,03 ja 0,01 (samassa järjestyksessä). Siten PISA-tutkimuksissa mitattu motivaatio ei selitä maahanmuuttajanuorten osaamisen vaihtelua juuri lainkaan. (Harju-Luukkainen ym. 2014.) Siksi onkin syytä pohtia, miten esimerkiksi koulun opiskelukielen taito, muiden kielten taito, koulukokemukset, matemaattisten taitojen lähtötaso tai kotoutumiseen liittyvät seikat ovat kenties motivaatiota vahvemmin yhteydessä oppimistuloksiin.

Ryanin ja Decin (2000) mukaan yksilön hyvinvointi ja toimiva oppimisympäristö edellyttää, että oppilas kokee, että hänellä on mahdollisuuksia autonomiaan ja omiin valintoihin, läheisyyteen toisten kanssa sekä vaikuttaa ja tuntee pätevyyttä. Jos nämä perustarpeet toteutuvat, on todennäköisempää, että oppilas löytää sisäisen motivaation. Kyvykkyyden kokemuksella ja sisäisellä motivaatiolla voikin olla vahva positiivinen yhteys, mutta tämän toteutumiseksi aktiviteettien tulee olla riittävän haastavia ja tarjottava mahdollisuuksia itseohjautuvalle toiminnalle. Motivaatio ei ole siis vain yksilön

ominaisuus, vaan se rakentuu vuorovaikutustilanteissa ja sen rakentumiseen vaikuttaa se, millaista palautetta oppilaat saavat toimijoina ja osajoina (Lerkkanen, Tarnanen, Salo & Ratinen 2015). Siksi matematiikan opetuksessa on paitsi koulun opiskelukielellä myös oppilaiden omilla kielillä keskeinen asema vertaispohdintaan ja yhteisölliseen tiedonkehittelyyn kannustavan vuorovaikutuksen rakentamisessa, mikä voi näkyä oppilaille tarjoumina mutta myös velvollisuutena osallistua matemaattiseen dialogiin (Walshaw & Anthony 2008; Hähkiöniemi, Kauppinen & Tarnanen 2015).

3 Tutkimusongelmat

Tarkastelemme tässä artikkelissa Suomen PISA 2012 -tutkimusaineiston avulla eri kieliryhmiin kuuluvien nuorten sisäisen ja ulkoisen motivaation yhteyttä heidän matematiikassa saavuttamiinsa pistemääriin. Haluamme erityisesti selvittää maahanmuuttajataustaisten oppilaiden oppimistulosten ja motivaation yhteyttä tilastollisesti edustavalla aineistolla ja luoda yleiskuvaa mahdollisista kieliryhmien välisistä eroista seuraavilla tutkimuskysymyksillä:

1. Onko maahanmuuttajataustaisten oppilaiden sisäisessä ja ulkoisessa motivaatioindeksissä havaittavissa eroja eri kieliryhmien välillä?
2. Onko oppilailla kieliryhmittäin löydettävissä yhteys motivaatioindeksien sekä matematiikan pistemäärien välillä?

4 Aineisto ja menetelmät

PISA on OECD:n (Organisation for Economic and Cultural Development) tutkimusohjelmaan kuuluva kansainvälinen arviointitutkimus, jota on toteutettu vuodesta 2000 lähtien. Arviointi toteutetaan joka kolmas vuosi kolmea pääarviointialuetta (lukutaito, matematiikka ja luonnontieteet) vaihdellen. Suurin osa koetehtävistä kohdistuu kunkin vuoden pääarviointialueeseen. Vuoden 2012 PISA-tutkimuksessa pääarviointialueena oli matematiikka, ja kaksi kolmasosaa oppilaiden vastaamista tehtävistä liittyi matematiikkaan ja yksi kolmasosa muihin arviointialueisiin. Vuonna 2012 PISA-tutkimus toteutettiin lähes 70 maassa tai talousalueella. Siten PISA on yksi maailman suurimmista kansainvälisistä arvioinneista, ja se antaa hyvän vertailupohjan suomalaisten 15-vuotiaiden nuorten osaamiselle.

PISA-arviointitutkimuksella pyritään selvittämään tarkemmin 15-vuotiaiden, peruskoulun päättövaiheessa olevien tai juuri peruskoulun päättäneiden nuorten, osaamista pääarviointi- ja sivuarviointialueissa. Tutkimuksella selvitetään, kuinka nuo-

ret osaavat etsiä, soveltaa ja tuottaa tietoa erilaisten arkisten ongelmatilanteiden ratkaisemiseksi. PISA-arviointitutkimus ei ole sidoksissa siihen osallistuvien maiden opetussuunnitelmiin, vaan siinä kartoitetaan oppilaiden yhtä lailla koulussa kuin koulun ulkopuolella oppimia arjen ongelmatilanteiden ratkaisutaitoja.

Suomen PISA 2012 -tutkimuksessa poimittiin ensimmäistä kertaa yliotos maahanmuuttajataustaisista nuorista. Yliotos tarkoittaa sitä, että tarkasteltavan joukon (tässä maahanmuuttajataustaisten nuorten) osuus otoksessa on suurempi kuin heidän osuutensa oppilasperusjoukossa. Suomen 15-vuotiaiden nuorten perusjoukossa maahanmuuttajataustaisten osuus oli vuonna 2012 noin 3,5 prosenttia. Sen mukaan PISA-otokseen olisi tullut arviolta 150–300 maahanmuuttajataustaista oppilasta. Näin pieni havaintomäärä ei kuitenkaan anna mahdollisuuksia kovin tarkkojen tilastollisten päätelmien tekemiseen, mistä syystä päädyttiin poimimaan yliotos. Yliotostaminen toteutettiin siten, että kaikki koulut, joissa oli vähintään viisi 15-vuotiaista maahanmuuttajataustaista oppilasta, otettiin mukaan otokseen, ja kaikki näiden koulujen maahanmuuttajanuoret pyrittiin testaamaan. Näin Suomen PISA 2012 -aineistoon saatiin kaikkiaan 1 294 maahanmuuttajataustaista nuorta. Tämä on noin 15 prosenttia Suomen havaintoaineistosta, jossa oli yhteensä 8 829 oppilasta 311 koulusta. Vastaava yliotostaminen tehtiin myös ruotsinkielisten koulujen suhteen: kaikki ruotsinkieliset koulut, joissa oli 15-vuotiaita oppilaita, poimittiin mukaan otokseen. Kansallisia ja kansainvälisiä analyysejä tehtäessä ruotsinkielisten ja maahanmuuttajataustaisten oppilaiden osuudet kuitenkin palautetaan painokertoimien avulla vastaamaan heidän todellisia osuuksiaan perusjoukossa (ruotsinkielisillä noin 6 %, maahanmuuttajilla noin 3,5 % aineiston keruuaihana). (Harju-Luukkainen ym. 2014). Siten tässäkin tutkimuksessa esiintyvät luvut ovat vertailukelpoisia muualla raportoitujen kansainvälisten ja kansallisten tulosten kanssa.

PISA-arviointitutkimuksessa kerätään tietoa oppilaille annettujen osaamistehävien lisäksi kahdella taustakyselyllä: oppilaskyselyllä sekä rehtoreille suunnatulla koulukyselyllä. Taustakyselyjen kysymyksistä muodostetaan erilaisia summamuuttujia tai indeksejä, joita voidaan käyttää erilaisissa lisäanalyyseissä esimerkiksi oppilaiden menestymistä selittävinä taustamuuttujina.

PISA-tutkimuksen oppilaskyselyssä oppilaiden sisäistä ja ulkoista motivaatiota arvioitiin kumpaakin neljällä väittämällä. Nuoret ottivat kantaa väittämiin neliportaisella asteikolla, joka vaihteli välillä ”täysin samaa mieltä” ja ”täysin eri mieltä”. Sisäistä motivaatiota mittaavat väittämät olivat seuraavat:

- Nautin matematiikkaa käsittelevien kirjojen lukemisesta.
- Odotan kovasti matematiikan tunteja.
- Opiskelen matematiikkaa, koska nautin siitä.
- Olen kiinnostunut asioista, joita opin matematiikassa.

Ulkoista motivaatiota mittaavat kysymykset olivat puolestaan seuraavat:

- Matematiikkaan kannattaa panostaa, koska siitä on apua työssä, jota toivon tekeväni myöhemmin.
- Minun kannattaa opiskella matematiikkaa, koska se parantaa työmahdollisuuksiani.
- Matematiikka on minulle tärkeä aine, koska tarvitsen sitä jatko-opinnoissani.
- Opin matematiikassa monia asioita, jotka auttavat minua saamaan työtä

Sekä sisäisestä että ulkoisesta motivaatiosta on muodostettu indeksimuuttuja, joka on standardoitu OECD-maiden suhteen siten, että sen keskiarvo on OECD-alueella nolla ja keskihajonta yksi. Indeksien positiiviset arvot kertovat näin ollen OECD-maiden keskitasoa vahvemman motivaatiosta ja negatiiviset arvot puolestaan keskimääräistä heikommasta motivaatiosta.

Tämän artikkelin analyyseissa on käytetty pääasiassa kuvailevia tilastollisia menetelmiä (keskiarvot, prosenttiosuudet). Kuvailevien menetelmien lisäksi kieliryhmien välisten keskiarvoerojen merkittävyyttä arvioitiin 95 prosentin luottamusvälien avulla. PISA-tutkimusasetelman metodologisen monimutkaisuuden (mm. kaksivaiheinen koulu-oppilas-otanta-asetelma eri suurin poimintatodennäköisyyksin, ks. OECD 2009a) vuoksi tavanomaiset tilastolliset menetelmät eivät sellaisenaan sovellu PISA-aineistojen analyysiin, vaan vaaditaan räätälöityjä menetelmiä. PISA:n otanta-asetelman tapauksessa luottamusvälit ovat perinteistä keskiarvotestausta yleiskäyttöisempi työkalu keskiarvoerojen tilastollisen merkittävyyden määrittämisessä. Luottamusvälien laskennassa käytettiin PISA-konsortion tähän tarkoitukseen tuottamia SAS®-tilasto-ohjelmistolle kirjoitettuja makroja, jotka ovat ladattavissa OECD:n PISA-sivuilta (OECD 2009b).

5 Tulokset

Tässä tutkimuksessa pyrittiin vastaamaan seuraaviin kysymyksiin: onko maahanmuuttajataustaisten oppilaiden sisäisessä ja ulkoisessa motivaatioindeksissä havaittavissa eroja eri kieliryhmien välillä sekä onko oppilailla kieliryhmittäin löydettävissä yhteys motivaatioindeksien sekä matematiikan pistemäärien välillä PISA 2012 -arviointitutkimuksessa.

PISA-tutkimuksen taustakyselyssä kysyttiin oppilailta, mitä kieltä he puhuvat kotonaan eniten. Suomalaisoppilaiden määrät eri kieliryhmissä nähdään taulukossa 1. Suomen- ja ruotsinkielisten jälkeen suurimmat kieliryhmät aineistossa ovat venäjää ja somalin kieltä puhuvat. Myös vironkieliset muodostavat kohtuullisen suuren ryhmän. Muut kieliryhmät ovat hyvin pieniä, ja näiden kielten osalta aineistosta saavat tulok-

set jäävät parhaimmillaankin vain suuntaa antaviksi. Luokan ”muu” osuus oli aineistossa odottamattoman suuri. Tämä luokka koostunee kymmenistä kielistä, joita kyselyn suunnitteluvaiheessa ei ollut mahdollista eritellä tarkemmin. Luokkaan kuuluva joukko on hyvin heterogeeninen ja hankalasti hahmotettava, ja siinä on mukana sekä eurooppalaisia että muissa maissa ja maanosissa puhuttavia kieliä ja ylipäänsä hyvin erilaisista kulttuuritaustoista tulevia oppilaita. On luultavaa, että useimpia luokkaan ”muu” sijoittuneista kielistä edustaa aineistossa vain muutama oppilas, mutta valitettavasti luokkaan kuuluu myös joitakin suuria kieliryhmiä (mm. englantia), joista olisi ollut mielenkiintoista saada täsmällistä tietoa. Painotetuista prosenttiosuuksista nähdään, että muuta kuin suomea tai ruotsia kotona puhuvia oppilaita on 15-vuotiaiden suomalaisoppilaiden perusjoukossa vain runsaat kolme prosenttia.

TAULUKKO 1. Oppilaiden lukumäärä ja matematiikan pistemäärien keskiarvo kieliryhmittäin.

Oppilaan kotonaan eniten puhuma kieli	Lukumäärä otoksessa	Painotuksella korjattu %-osuus	Matematiikan pistemäärien keskiarvo
suomi	6253	91,4	523
ruotsi	1201	5,2	521
venäjä	294	0,9	466
somali	179	0,4	391
viro	93	0,3	467
turkki	43	0,1	416
kiina	28	0,1	509
thai	15	0,1	410
muu	559	1,6	441

Taulukosta 1 nähdään matematiikan PISA-pistemäärien keskiarvot eri kieliryhmissä. Taulukkoon on merkitty lihavoituna ne kielet, joita kotonaan puhuvien keskiarvo poikkeaa luottamusvälien perusteella tilastollisesti merkitsevästi ($p < 0,05$) suomea kotonaan puhuvien oppilaiden keskiarvosta. Ainoat ryhmät, joiden tulokset eivät tässä suhteessa poikenneet suomea kotonaan puhuvien oppilaiden tuloksesta, olivat ruotsia ja kiinaa puhuvat. Kiinaa puhuvien ryhmä oli niin pieni, että tilastollisesti merkitsevää eroa on vaikea saada, mutta toisaalta kiinaa puhuvien keskiarvo oli kuitenkin melko lähellä kotimaisia kieliä puhuvien keskiarvoja. Esimerkiksi thai-kieltä puhuvien tulos oli merkitsevästi suomea puhuvien tulosta heikompi, vaikka heitä oli aineistossa vielä vähemmän kuin kiinaa puhuvia.

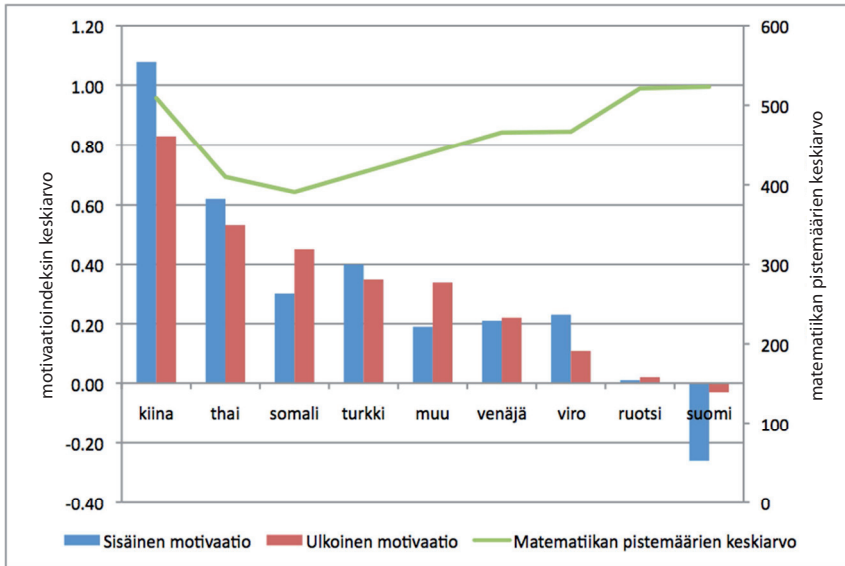
Sisäistä ja ulkoista motivaatiota mittaavien indeksien keskiarvot eri kieliryhmissä on esitetty taulukossa 2. Taulukosta voidaan havaita, että suomenkielisten oppilaiden keskiarvot ovat OECD:n keskitasoa (ulkoinen motivaatio) tai sen alapuolella (sisäinen motivaatio). Suomenkielisten keskiarvosta merkitsevästi poikkeavien kieliryhmien

keskiarvot on lihavoitu. Sisäisen motivaation keskimääräinen taso on kaikissa kieliryhmissä merkitsevästi korkeampi kuin suomenkielisillä. Ulkoisen motivaation osalta puolestaan tämä ero on merkitsevä venäjää, somalia, turkkia ja kiinaa sekä ”muuta kieltä” kotonaan puhuvilla. Aineiston perusteella keskimääräinen motivaatiotaso on korkein aasialaistaustaisilla (etenkin kiinaa puhuvilla) oppilailla ja matalin siis suomea kotonaan pääasiassa puhuvilla oppilailla.

TAULUKKO 2. Oppilaiden sisäinen ja ulkoinen motivaatioindeksi kieliryhmittäin.

Oppilaan kotonaan eniten puhuma kieli	Sisäistä motivaatiota mittaavan indeksin keskiarvo	Ulkoista motivaatiota mittaavan indeksin keskiarvo
suomi	-0,26	-0,03
ruotsi	0,01	0,02
venäjä	0,21	0,22
somali	0,30	0,45
viro	0,23	0,11
turkki	0,40	0,35
kiina	1,08	0,83
thai	0,62	0,53
muu	0,19	0,34

Kuviossa 1 on esitetty sisäisen ja ulkoisen motivaation indeksien keskiarvot (pylväät) sekä matematiikan keskiarvot (keskiarvoprofilili) tarkastelluissa kieliryhmissä. Ryhmät on järjestetty vasemmalta oikealla ulkoisen motivaation mukaiseen järjestykseen. Kuviossa on nähtävissä, että motivaation yhteys matematiikan oppimistuloksiin ei ole kieliryhmien tasolla suoraviivainen. Matematiikan tulokset riippuvat siis maahanmuuttajataustaisilla nuorilla olennaisesti muista tekijöistä kuin (PISA-tutkimuksessa mitatusta) motivaatiosta. Oppilastasolla samanlaisen havainnon ovat aikaisemmin tehneet Harju-Luukkainen ym. (2014). Motivoituneimpien ja vähiten motivoituneiden ensimmäisen sukupolven maahanmuuttajaoppilaiden PISA-matematiikan tulostasossa ei ollut minäkäänlaista eroa, ja toisen sukupolven maahanmuuttajaoppilailla tämä ero oli hyvin pieni (motivoituneimpien eduksi). Sen sijaan nuorilla, joilla ei ollut maahanmuuttajataustaa, motivaatiolla oli selkeä positiivinen yhteys oppimistuloksiin.



KUVIO 1. Oppilaiden sisäisen ja ulkoisen motivaation indeksien keskiarvot sekä matematiikan pistemäärien keskiarvo kieliryhmittäin.

Tutkimustulosten perusteella voidaan päätellä, että oppilaiden kielellinen tausta näyttäisi näiden tutkimustulosten valossa olevan yhteydessä nuorten sisäisen ja ulkoisen motivaation tasoihin. Oppilaiden, jotka puhuivat kotonaan kotimaisia kieliä, motivaatiotaso oli heikko tai parhaimmillaan OECD-maiden keskitasoa. Sen sijaan oppilaiden, jotka puhuivat pääasiassa jotain muuta kuin kotimaista kieltä kotona, keskimääräinen motivaatiotaso (sekä sisäinen että ulkoinen) oli kauttaaltaan korkeampi, usein myös tilastollisesti merkitsevästi, kuin kotimaisia kieliä kotona puhuvilla. Kuitenkaan tämä motivaatio ei ollut maahanmuuttajataustaisilla nuorilla yhteydessä hyviin oppimistuloksiin kuten niillä oppilailla, joilla ei ollut maahanmuuttajataustaa. Tulosten mukaan motivaation ja oppimistulosten välinen yhteys näyttyy kieliriippuvaisena, mutta ei suinkaan suoraviivaisena. Tässä tulee toki huomioda, että pelkkä riippuvuus ei suinkaan tarkoita sitä, että kielitausta olisi motivaatiota muokkaava tekijä, vaan taustalla saattaa olla monia muitakin vaikuttavia tekijöitä, kuten esimerkiksi kulttuuri- tai koulutausta.

6 Pohdinta

Motivaatio on monitahoinen ilmiö, mutta se on yksi keskeisimmistä tekijöistä, joka näkyy oppilaan aktiivisuutena (Ryan & Deci 2000; Nurmi 2013) ja on myös yhteydessä hyviin oppimistuloksiin, sillä motivaatiolla (sisäisellä ja ulkoisella) näyttäisi kokonai-

suutena olevan selkeä yhteys oppilaan oppimistuloksiin (Harju-Luukkainen ym. 2014). Kuitenkin tarkasteltaessa oppilaita, joilla on maahanmuuttajatausta, motivaatio ja sen yhteys oppimistuloksiin näyttäytyy erilaisena. Maahanmuuttajataustaisilla oppilailla oli kokonaisuudessaan erittäin korkea sisäinen ja ulkoinen motivaatio, mutta sen yhteys oppimistuloksiin oli heikko, jopa olematon (Harju-Luukkainen ym. 2014).

Nämä aikaisemmat tutkimustulokset maahanmuuttajataustaisista nuorista toivat tämän tutkimuksen lähtökohtana. Tällä tutkimuksella halusimme selvittää, onko motivaation taso samanlainen eri kieliryhmissä (kotona pääasiassa puhutun kielen mukaan eriteltynä) ja onko motivaation tasolla yhteyttä eri kieliryhmien oppimistuloksiin. Tutkimuksemme tulosten luotettavuutta rajoittaa se tosiasia, että monissa kieliryhmissä oppilaita oli vain vähän. Toinen tutkimuksen yleistettävyyttä heikentävä tekijä on se, että PISA-tutkimuksessa oppilaiden motivaatiota (sekä sisäistä että ulkoista) mitataan kovin suppeasti, kaikkineensa vain kahdeksalla kysymyksellä. Näin ollen onkin syytä kysyä, kuinka osuvasti PISA-tutkimus todella mittaa motivaatiota kokonaisuudessaan. Näistä rajoitteista huolimatta tämä tutkimus antaa lähtökohdan lisätutkimuksille.

Tuloksemme osoittivat, että maahanmuuttajataustaisten oppilaiden eri kieliryhmien sisäisen ja ulkoisen motivaation tasot vaihtelivat, mutta ne olivat kautta linjan korkeammalla tasolla kuin nuorilla, joilla ei ollut maahanmuuttajataustaa. Näin ollen maahanmuuttajataustaiset nuoret olivat erittäin motivoituneita sekä sisäisesti että ulkoisesti kieliryhmästä riippumatta. Heidän keskimääräinen motivaatiotasonsa oli myös korkea muihin OECD-maihin verrattuna. Näin ollen näyttäisikin siltä, että nuorten kielellisellä ja sitä kautta myös kulttuurisella taustalla on jonkinlainen yhteys oppilaiden motivaation tasoon. Kiinaa pääasiassa kotonaan puhuvilla nuorilla oli erittäin korkea sekä sisäisen että ulkoisen motivaation taso ja edelleen korkea matematiikan pistemäärien keskiarvo.

Tämän tutkimuksen tulosten mukaan näyttäisikin siltä, että motivaation (sisäisen ja ulkoisen) yhteys oppilaan oppimistuloksiin ei ole suoraviivainen. Korkea motivaatio ei siis automaattisesti näy nuoren hyvinä oppimistuloksina, vaikka näin voisi olettaa. Selittäviä tekijöitä voi olla useita, kuten nuoren suomen tai ruotsin kielitaito, matematiikan osaamisen lähtötaso, koulukokemukset, lähtömaan ja nykyisen maan kulttuurien läheisyys, olosuhteet kotona tai ylipäättään integroitumiseen liittyvät sosio-emotionaaliset tekijät. Tutkimusnäyttöä on PISA 2012 -arviointitutkimuksen kohdalla kuitenkin siitä, että kulttuuritaustojen läheisyys osoittautui keskeiseksi tekijäksi parhaiten menestyneiden maahanmuuttajataustaisten nuorten tulosten taustalla (Harju-Luukkainen, Nissinen & Tarnanen 2015). Motivaation voidaan olettaa olevan korkealla nuorten tarvehierarkiassa, ja jotta se pääsee vaikuttamaan suotuisalla tavalla, on tiettyjen tekijöiden kuten kielitaidon tai kulttuurituntemuksen oltava riittävällä tasolla (vrt. Onchwari, Onchwari & Keengwe 2008).

Onkin mielenkiintoista pohtia oman äidinkielen taidon merkitystä ja sen hyödyntämistä opetuksessa, kun tarkastellaan kieliryhmittäin motivaation ja oppimistulosten yhteyttä. Oppilaan oman äidinkielen hyödyntäminen yli oppiainerajojen sekä koulun sensitiivisyys monikulttuurisuutta kohtaan voivat tukea oppilaiden identiteettien kehittymistä ja näin tukea myös oppimista, sillä vahvan etnisen identiteetin on osoitettu muissa maissa tehdyissä tutkimuksissa olevan yhteydessä oppimistuloksiin (esim. Altschul, Oyserman & Bybee 2006; Zirkel 2008). Se, että tutkimuksemme nuoret osoittivat korkeaa ulkoista ja sisäistä motivaatiota, antaa viitteitä siitä, että koulun sosiaalinen ympäristö on ehkä pystynyt luomaan motivaatiota edistävän ilmapiirin (Ryan & Deci 2000), ellei aineistomme maahanmuuttajanuorten vastauksissa heijastu erityisen vahvasti tarve antaa kyselyssä suotuisia vastauksia.

Sinällään korkean motivaation voisi olettaa viestivän edellä mainitun sosiaalisen ympäristön lisäksi siitä, että monikielisten ja -kulttuuristen identiteettien kehittymiselle on ollut tilaa (vrt. Hélot & Young 2002; Martin-Jones 2007). Toisaalta heikot oppimistulokset ja niiden heikko tai lähes olematon yhteys motivaatioon ohjaa pohtimaan riittävämmän kielitaidon lisäksi sitä, miten koulu onnistuu tukemaan maahanmuuttajataustaisten oppilaiden oppimista kulttuurisesta ja opiskelutaidollisesta näkökulmasta. Koulutuksen arviointineuvoston raportin mukaanhan koulujen henkilöstö kaipaa tukea asenteiden ja oman monikulttuurisuusosaamisensa kehittämiseen (Kuukka ym. 2015). Tätä peräänkuuluttaa myös Teräs (2015), jonka mukaan tasavertainen keskustelu opetuksen käytännöistä eri kulttuuri- ja kielitaustaisten opiskelijoiden kanssa antaa mahdollisuuden rakentaa uudenlaista opetuskulttuuria. Olisikin suotavaa, että kouluissa löytyisi aikaa kulttuurisesta moninaisuudesta tiedostumiseen ja keskustelemiseen.

On ilmeistä, että varsinkin ensimmäisen sukupolven maahanmuuttajataustaisten nuorten suomen tai ruotsin kielitaidolla voi olla haasteellista opiskella yläkoulun eri oppiaineiden oppimääriä, varsinkin kun kyse on akateemisista taidoista ja kunkin oppiaineen tiedonalan kielen omaksumisesta. Lisäksi etenkin matematiikan opiskelussa voidaan vähätellä kielitaidon merkitystä, vaikka oppilaiden kirjallinen ja suullinen osallistuminen matemaattiseen keskusteluun on sekä oppimisen mahdollistaja että oppimistavoite. Kielellä, olipa se sitten koulun opiskelukieli tai oppilaiden oma äidinkieli, on myös matematiikan oppimisessa keskeinen merkitys (esim. Walqui 2006; Mercer & Littleton 2007).

Koulu, jossa kaikilla oppilalla on tasavertaiset mahdollisuudet oppimiseen sukupuoleen, sosiaaliluokkaan tai etnisiin ja kulttuurisiin ominaisuuksiin katsomatta, näyttäisi kaipaavan vielä kehittämistoimia (vrt. Banks 2010). Yhteenvetona voimme todeta, että tuloksemme näyttäisivät vahvistavan kielitietoisien ja kulttuuritietoisien aineenopetukseen tarvetta yli oppiainerajojen. Kyse ei siis ole välttämättä vain suomi tai ruotsi toisena kielenä -opetuksen määrästä vaan myös siitä, että kunkin oppiaineen opettaja

tunnistaa olevansa oman tiedonalansa lisäksi kielen opettaja kulttuuritietoista opetusta unohtamatta.

Kirjallisuus

- Alitolppa-Niitamo, A. 2003. Liminaalista jäsenyyteen? Somalinkielisten nuorten siirtymien haasteita. Teoksessa P. Harinen (toim.) *Kamppailuja jäsenyyksistä. Etnisyys, kulttuuri ja kansalaisuus nuorten arjessa*. Helsinki: Nuorisotutkimusverkosto, 17–32.
- Altschul, I., D. Oyserman & D. Bybee 2006. Racial-ethnic identity in mid-adolescence: content and change as predictors of academic achievement. *Child Development*, 77 (5), 1155–1169.
- Baker, C. 2000. *The care and education of young bilinguals: an introduction for professionals*. Clevedon, England: Multilingual Matters.
- Banks, J. A. 2010. Multicultural education: characteristics and goals. Teoksessa J. A. Banks & C. A. McGee Banks (toim.) *Multicultural education. Issues and perspectives*. (7. painos.) Hoboken, USA: John Wiley & Sons, Inc, 3–32.
- Bialystok, E. & X. Feng 2009. Language proficiency and executive control in proactive interference: evidence from monolingual and bilingual children and adults. *Brain and Language*, 109 (2–3), 93–100.
- Civil, M. 2009. *A survey of research on the mathematics teaching and learning of immigrant students* [online]. Proceedings of CERME 6 [Luettu 22.11.2016]. <http://ife.ens-lyon.fr/publications/edition-electronique/cerme6/wg8-01-civil.pdf>.
- Cummins, J. 2000. *Language, power and pedagogy. Bilingual children in the crossfire*. Clevedon: Multilingual Matters.
- Dewey, J. 1938. *Experience and education*. New York: The Macmillan company.
- Dörnyei, Z. 2003. Attitudes, orientations, and motivations in language learning: advances in theory, research, and applications. Teoksessa Z. Dörnyei (toim.) *Attitudes, orientations and motivations in language learning*. Oxford: Blackwell, 3–32.
- Dörnyei, Z. & E. Ushioda 2011. *Teaching and researching motivation*. Harlow: Longman.
- Hannula, M. S. 2006. Motivation in mathematics: goals reflected in emotions. *Educational Studies in Mathematics*, 63 (2), 165–178.
- Harju-Luukkainen, H., K. Nissinen, S. Sulkunen, M. Suni & J. Vettenranta 2014. *Avaimet osaamiseen ja tulevaisuuteen. Selvitys maahanmuuttajataustaisten nuorten osaamisesta ja siihen liittyvistä taustatekijöistä PISA 2012 -tutkimuksessa*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Harju-Luukkainen, H., K. Nissinen & M. Tarnanen 2015. Matematiikka ja maahanmuuttajataustaiset nuoret. Teoksessa J. Välijärvi & P. Kupari (toim.) *Millä eväillä uuteen nousuun? PISA 2012 tutkimustuloksia*. Opetus- ja kulttuuriministeriön julkaisuja 2015: 6. Helsinki: Opetus- ja kulttuuriministeriö, 108–123.
- Heath, A. & E. Kilpi-Jakonen 2012. *Immigrant children's age at arrival and assessment results*. OECD Education Working Papers No. 75. PARIS: OECD.
- Hélot, C. & A. Young 2002. Bilingualism and language education in France primary schools. Why and how should migrant languages be valued? *International Journal of Bilingual Education and Bilingualism*, 5 (2), 96–112.
- Hähkiöniemi, M., M. Kauppinen & M. Tarnanen 2015. Luokanopettajaopiskelijoiden kielitietoisuus matematiikan päättelyketjujen tulkinnessa. Teoksessa M. Kauppinen, M. Rautainen & M. Tarnanen (toim.) *Rajaton tulevaisuus: kohti kokonaisvaltaista oppimista*.

- Ainedidaktiikan symposium Jyväskylässä 13.–14.2.2014*. Ainedidaktisia tutkimuksia (8). Suomen ainedidaktinen tutkimusseura ry, 81–95.
- Kilpi, E. 2010. *The education of children of immigrants in Finland*. Väitöskirja. University of Oxford.
- Kilpi-Jakonen, E. 2012. Does Finnish educational equality extend to children of immigrants? Examining national origin, gender, and the relative importance of parental resources. *Nordic Journal of Migration Research*, 2 (2), 167–181.
- Kuukka, K., N. Ouakrim-Soivio, H. Paavola & M. Tarnanen 2015. Kulttuurisen monimuotoisuuden edistäminen ja kehittämistarpeet. Teoksessa T. Pirinen (toim.) *Maahanmuuttajataustaiset oppijat suomalaisessa koulutusjärjestelmässä – koulutuksen saavutettavuuden ja opiskelun aikaisen tuen arviointi*. Helsinki: Kansallinen koulutuksen arviointikeskus, 181–199.
- Kuusela, J. & A. Etelälahti 2008. Maahanmuuttajaoppilaiden oppimistulokset, asenteet ja jatkokoulutusvalinnat perusopetuksen päättövaiheessa. Teoksessa J. Kuusela, A. Etelälahti, Å. Hagman, R. Hievanen, K. Karppanen, L. Nissilä, U. Rönning & M. Siniharju (toim.) *Maahanmuuttajaoppilaat ja koulutus – tutkimus oppimistuloksista, koulutusvalinnoista ja työllistämisestä*. Helsinki: Opetushallitus, 97–134.
- Leinonen, E. 2013. *Perheen sukupolvisuhteet ja kotoutuminen nuorten näkökulmasta*. Helsinki: Helsingin kaupungin tietokeskus.
- Lerikkanen, M.-K., M. Tarnanen, O.-P. Salo & I. Ratinen 2015. Opetussuunnitelmaperusteiden oppimiskäsityksen mukainen opetus: esimerkkejä kielistä, maantiedosta ja ympäristöopista. Teoksessa O.-P. Salo & M. Kontoniemi (toim.) *Kohti uutta: 100 vuotta koulun kehittämistä Jyväskylän normaalikoulussa*. Jyväskylä: Jyväskylän normaalikoulu, Jyväskylän yliopisto, 115–132.
- Lukin, T. 2013. *Motivaatio matematiikan opiskelussa – seurantatutkimus motivaatiotekijöistä ja niiden välisistä yhteyksistä yläkoulun aikana*. Joensuu: Itä-Suomen yliopisto.
- Martin-Jones, M. 2007. Bilingualism, education and the regulation of access to language resources. Teoksessa M. Heller (toim.) *Bilingualism. A social approach*. New York: Palgrave, 161–182.
- Mercer, N. & K. Littleton 2007. *Dialogue and the development of children's thinking*. London: Routledge.
- Nurmi, J.-E. 2013. Motivaation merkitys oppimisessa. *Kasvatus*, 44 (5), 548–554.
- OECD 2009a. *PISA Data Analysis Manual (SAS Second Edition)*. Paris: OECD Publishing.
- OECD 2009b. *SAS® Macro*. Saatavilla osoitteessa <http://www.oecd.org/pisa/pisaproducts/42628103.zip> [luettu 28.11.2016].
- Onchwari, G., J. Onchwari & J. Keengwe 2008. Teaching the immigrant child: application of child development theories. *Early Childhood Education Journal*, 36 (3), 267–273.
- Opetushallitus 2014. *Perusopetuksen opetussuunnitelman perusteet*. Helsinki: Opetushallitus.
- Opetushallitus 2016. *Maahanmuuttajien koulutuksen tilastot*. Saatavilla osoitteessa http://www.oph.fi/tietopalvelut/tilastotiedot/muita_koulutustilastoja/maahanmuuttajien_koulutus [luettu 22.1.2016].
- Pantziaram, M. & G. N. Philippou 2015. Students' motivation in the mathematics classroom. Revealing causes and consequences. *International Journal of Science and Mathematics Education*, 13 (2), 385–411.
- Rinne, R. & A. Tuittu 2011. Peruskoulu maahanmuuttaja- ja suomalaisvanhempien silmin. Teoksessa K. Klemelä, A. Tuittu, A. Virta & R. Rinne (toim.) *Vieraina koulussa? Monikulttuurinen koulu oppilaiden, vanhempien, opettajien ja rehtoreiden kokemana*. Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja. Turku: Turun yliopisto, 97–157.
- Ryan, R. M. & E. L. Deci 2000. Intrinsic and extrinsic motivations: classic definitions and new directions. *Contemporary Educational Psychology*, 25 (1), 54–67.

- Räsänen, M. & J. Kivirauma 2011. Oppilaana monikulttuurisessa koulussa. Teoksessa K. Klemelä, A. Tuittu, A. Virta & R. Rinne (toim.) *Vieraina koulussa? Monikulttuurinen koulu oppilaiden, vanhempien, opettajien ja rehtoreiden kokemana*. Kasvatustieteiden tiedekunnan julkaisuja A: 211. Turku: Turun yliopisto, 39–95.
- Sinkkonen, H.-M., M. Kyttälä, O. Karvinen & P. Aunio 2011. Maahanmuuttajalasten erityisluokkasiirrot – syynä todelliset oppimisvaikeudet vai heikko kielitaito? *NMI-Bulletin*, 1, 14–25.
- Teräs, M. & E. Kilpi-Jakonen 2013. Maahanmuuttajien lapset ja koulutus. Teoksessa T. Martikainen, P. Saukkonen & M. Säävälä (toim.) *Muuttajat. Kansainvälinen muuttoliike ja suomalainen yhteiskunta*. Helsinki: Gaudeamus Helsinki University Press, 84–202.
- Teräs, M., J. Lasonen & A. Sannio 2010. Maahanmuuttajien lasten siirtymät koulutukseen ja työelämään. Teoksessa T. Martikainen & L. Haikkola (toim.) *Maahanmuutto ja sukupolvet*. Helsinki: SKS, 85–109.
- Teräs, M. 2015. *Kehittävä kulttuurienvälisyys ammatillisen koulutuksen haasteena ja mahdollisuutena*. Ammattikasvatuksen aikakauskirjan erikoisnumero. Helsinki: OKKA-säätiö, 41–54.
- Tilastokeskus 2014. *Uutiset 2014*. Saatavilla osoitteessa http://www.stat.fi/ajk/tiedotteet/2014/uutinen_010_2014-03-25.html [luettu 8.2.2016].
- Valtiontalouden tarkastusvirasto 2015. *Tuloksellisuustarkastuskertomus. Maahanmuuttaja-oppilaat ja perusopetuksen tuloksellisuus*. Valtiontalouden tarkastusviraston tarkastuskertomukset 12/2015.
- Vollmer, H. J. 2006. Intergovernmental conference languages of schooling: towards a framework for Europe. Strasbourg: Language Policy Division.
- Walqui, A. 2006. Scaffolding instruction for english language learners: a conceptual framework. *The International Journal of Bilingual Education and Bilingualism*, 9 (2), 159–180.
- Walshaw, M. & G. Anthony 2008. The teacher's role in classroom discourse: a review of recent research into mathematics classrooms. *Review of Educational Research*, 78 (3), 516–551.
- Wigfield, A. & J. Cambria 2010. Students' achievement values, goal orientations, and interest: definitions, development, and relations to achievement outcomes. *Developmental Review*, 30 (1), 1–35.
- Zirkel, S. 2008. The influence of multicultural educational practices on student outcomes and intergroup relations. *Teachers College Record*, 110 (6), 1147–1181.

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 184–211.

Pirjo Pollari

University of Jyväskylä

Daunting, reliable, important or “trivial nitpicking?” Upper secondary students’ expectations and experiences of the English test in the Matriculation Examination

The Matriculation Examination, the school-leaving exam taken towards the end of upper secondary education, is the only high-stakes examination in the Finnish school system. As the exam may have a strong impact on the students’ further education opportunities, it evokes various feelings and thoughts in students. Yet, there is little research on these reactions. This article, based on a mixed-methods approach, sheds light on students’ expectation and experiences of the English test in the Matriculation Examination. A total of 142 second- and third-year students from one upper secondary school shared their views on the possible washback effect and test anxiety caused by the exam. Also, the students expressed their ideas and experiences of the validity, reliability and fairness of the test. Although the test did not seem to cause excessive washback, it caused significant stress and anxiety. Furthermore, students seemed rather critical of its validity and reliability.

Keywords: Matriculation Examination, students’ experiences, test anxiety, reliability, validity, washback

1 Introduction

In the Finnish school system, the Matriculation Examination is the only examination that can be considered a national, high-stakes examination (Atjonen 2015; Mehtäläinen & Välijärvi 2013). As school assessment is otherwise teacher-designed and quite low-stakes, the Matriculation Examination stands in marked contrast with it. Thus, it is no wonder the Matriculation Examination evokes various emotions, expectations and experiences in Finnish upper secondary school students. As students' experiences have rarely been studied their responses remain mainly anecdotal or based on hearsay.

This article aims to shed light on students' expectations for and experiences of the English test in the Matriculation Examination (ME) in one Finnish upper secondary school. The findings are based on a web-based questionnaire that 142 second- and third-year upper secondary students answered in March 2014. The study relies on mixed methods as both quantitative and qualitative data and methodology were used.

Firstly, the article will discuss high-stakes assessment and its characteristics as well as the Matriculation Examination. Then, the present study, its methodology and findings will be introduced. Finally, I will discuss the limitations and practical implications of this small study.

2 Theoretical background

2.1 High-stakes assessment

External, large-scale examinations, such as school-leaving examinations in various countries, are often labelled as high-stakes examinations. However, according to Heubert and Hauser (1999), for instance, what makes assessment high-stakes is not the assessment itself, nor its contents or form, but primarily the way its results are used and what their impacts are on the student, or on other stakeholders. Thus, in the educational setting, high-stakes tests normally refer to tests whose outcome has "high-stakes consequences for students – that is, when an individual student's score determines not just who needs help, but whether a student is allowed to take a certain program or class, or will be promoted to the next grade, or will graduate from high school" (Heubert & Hauser 1999: 14). High-stakes are therefore closely linked with pressure (Nichols, Glass & Berliner 2006).

The proponents of high-stakes testing have argued that today's high-stakes tests are of state-of-art quality: they are, for instance, "highly reliable; free from bias; relevant and age-appropriate" (Cizek 2005: 41). Hence, because of their outstanding validity and reliability, they can have a positive washback effect: when teachers prepare

their students for testing, they will be "teaching to the standards", which leads to better learning (Cizek 2005: 42). Furthermore, the high stakes attached to the test outcomes are believed to motivate students to study harder in order to gain rewards (e.g. admission for further education) and to avoid punishing consequences such as retention or denial of graduation (see e.g. Heubert & Hauser 1999; Kornhaber & Orfield 2001; Nichols et al. 2006). Along the same lines, high-stakes test scores have increasingly been used for other accountability purposes, such as evaluating an individual teacher's effectiveness or a school's performance (Cizek 2005), which they were not necessarily designed for (e.g. Jones, Jones & Hargrove 2003; Stobart 2008). As the rewards or threats are closely linked with money, job security and other significant factors (Amrein & Berliner 2002), they are believed to act as highly effective incentives and thus improve educational quality and effectiveness (e.g. Cizek, 2005).

Those who are critical of high-stakes testing say that instead of improving teaching and learning, the washback effect leads to teaching to test (e.g. Cheng, Watanabe & Curtis 2004; Madaus & Clarke 2001; Stobart 2008). The pressure of accountability means that schools and teachers want to make sure their students do well in exams and start to prepare them for the exams: as teachers devote more time for test revision and practice tests, it narrows both the content and methodology of teaching and learning (Kornhaber & Orfield 2001; Stobart 2008; see also Alderson & Hamp-Lyons 1996). High-stakes tests are also believed to make learning shallower as often students' primary purpose is to pass the test, not to learn the topics or skills per se (Harlen 2012). Furthermore, most tests focus only on the assessment of the outcomes of learning rather than the process of learning. These factors affect the learning strategies chosen by students when studying. All this may contribute to superficial rote learning instead of real conceptual understanding (e.g. Harlen 2012; Volante 2004).

As high-stakes tests are often one single test with highly pressurised time and place constraints, they may also cause considerable stress and test anxiety (e.g. Aydın 2009). According to research, female students in particular seem to suffer from test anxiety, which may weaken their test performance (Cassady 2010; Hembree 1988). Underperforming in the test, in turn, can affect students' motivation, self-efficacy and self-esteem as learners (Harlen & Deakin Crick 2003).

A low-stakes test has no highly significant consequences for the student (Heubert & Hauser 1999). The defining factor not being the test itself but the use and perceived consequences of the test results, what may be a high-stakes test for one student may not necessarily be so for another.

2.2 The quality of assessment

The quality of any assessment or test is attributed to several characteristics, such as validity, reliability, fairness, impact and practicality or cost-efficiency of the assessment (see e.g. Bachman 1990; Race, Brown & Smith 2005). With high-stakes testing, these characteristics are all the more crucial; as Bachman puts it (1990, 56): “The more important the decision, the greater the cost of making an error”.

Validity is traditionally “taken to mean how well what is assessed corresponds with the behaviours or learning outcomes that are intended to be assessed” (Harlen 2010: 36; see also Bachman 1990). Validity, however, is a broad concept and includes various types of validity. *Content validity* is about “the relevance of the test content to the content of a particular behavioural domain of interest and about the representativeness with which item or task content covers that domain” (Messick 1993: 17). *Consequential validity* (Messick 1993), then again, refers to the impact of the assessment (Harlen 2010) and *construct validity* to what is assessed (Harlen 2010; see also Messick 1993, 1996). Black and Wiliam (2012: 244) discuss the notions of construct under-presentation and construct-irrelevant variance (see also Messick 1996), defining them as follows:

“Construct under-presentation therefore occurs when an assessment fails to assess things it should. The opposite threat to valid interpretation – when an assessment assesses things it should not – is called construct-irrelevant variance.”

Often the variation in student scores that is caused by random factors is discussed under the heading of reliability, “reliability being the consistency or accuracy of the results” (Harlen 2010: 36). According to Black and Wiliam (2012), three main sources of construct-irrelevant variance are generally addressed when discussing threats to reliability. One of them is rater reliability, i.e. whether different raters give the same score to the same piece of student work, also known as inter-rater reliability. Intra-rater reliability, i.e. whether the same rater gives the same score to the same answer consistently, is also a significant issue when considering the consistency of scores (see e.g. Bachman 1990; Harlen 2010). The second source of construct-irrelevant variance is the variance in student performance from one day to another: in other words, the student may perform better or worse on different occasions and at different times. The third source is differences in student performance caused by the particular choice of questions or items in the test (see also Bachman 1990). In sum, Black (1998: 54) characterises reliability as follows: “Reliability depends on whether the results are reproducible with different markers, grading procedures, test occasions, and different sets of questions”.

In addition to validity and reliability, several authors also include factors such as transparency and fairness. Tests or any forms of assessments should not have nasty

surprises and they should be in line with the intended learning outcomes. Moreover, "students should not be playing the game 'guess what's in our assessors' minds'" (Race et al. 2005: 2). Also, assessment practices should not discriminate or favour any individuals or groups of students. One way of ensuring fairness and equity is a balanced array of different types of exercises.

2.3 The Matriculation Examination

2.3.1 The Matriculation Examination and its history

The Chinese civil service examinations (c. 600–1905), which in some form date back to the times BCE, are credited as the first large-scale, high-stakes examination system in the world (e.g. Elman 2000). However, the standardised, high-stakes tests have dominated educational assessment mainly since the births of the IQ test and the multiple-choice test in the early 20th century (Hanson 1993; Nichols et al. 2006). The current proliferation of high-stakes testing in the USA and Britain, for instance, dates back to the 1980s (e.g. Amrein & Berliner 2002; Black 1998; Kornhaber & Orfield 2001).

The Finnish tradition of testing is quite different from those of the English-speaking countries. The only external, high-stakes examination that we have in the Finnish school context is the Matriculation Examination (Atjonen 2015; Mehtäläinen & Välijärvi 2013). As the word matriculation suggests, its roots lie in an oral entrance examination for Turku Academy. The first modern Matriculation Examination was arranged in 1852. Organised by the Matriculation Examination Board, the new examination was based on upper secondary school syllabi (Kaarninen & Kaarninen 2002; Lindström 1998). Thus far, the Matriculation Examination had still been a university entrance examination. However, in 1919 the Matriculation Examination became the final examination of the upper secondary school, and passing it ceased to mean automatic matriculation to the university. All parts of the examination, both written and oral, were to be organised at schools themselves at the very same time and under strict regulations (Kaarninen & Kaarninen 2002; Lindström 1998).

As with many other high-stakes examinations, the results of the Matriculation Examination were used for assessing the quality of the school until 1918 (Lindström 1998). So, the recent media interest to rank upper secondary schools on the basis of the Matriculation Examination results is not a new phenomenon in Finland. Neither is the washback effect of the Matriculation Examination: according to Lindström (1998) many teachers and principals criticised the Matriculation Examination for narrowing the curricula and teaching methodology into teaching to the test over a hundred years ago.

The Matriculation Examination has also undergone some changes more recently. For instance, a listening comprehension part was added to major foreign or second language tests in the 1970s. Since 1994, it has been possible to divide the examination over three consecutive exam periods, instead of taking the whole exam in just one term. Separate tests for each of the natural sciences and humanistic subjects, instead of an all-encompassing test including all subjects, were introduced in 2006. Currently, the Matriculation Examination is undergoing a process of digitalisation: all of its tests should be computerised by 2019. (For further information, see the Matriculation Examination Board.)

2.3.2 Earlier research on the Matriculation Examination

Research on student assessment in general is rather scarce in Finland, and so is research on the Matriculation Examination. There is some research that focuses on the comparability and reliability of the Matriculation Examination grades (Mehtäläinen & Välijärvi 2013), the history of the Matriculation Examination (Kaarninen & Kaarninen 2002; Lindström 1998) as well as its status (Vuorio-Lehti 2006, 2007). Furthermore, Anckar (2011) has investigated the processes and strategies that students used when answering multiple-choice questions in one French listening comprehension test of the Matriculation Examination. Her findings showed that items with flaws, such as too 'tricky' questions or options as well as items with excessive textual information load or difficulty, represented threats to the reliability of item scores.

Two or three studies have touched upon students' own expectations or experiences of the exam. First, Syrjälä (1989) studied students' and teachers' views and experiences on student assessment as part of studying and teaching. One question in the questionnaire that was part of this small-scale assessment experiment dealt with the Matriculation Examination: over 60% of the respondents, who all were third-year students, found the Matriculation Examination useful while 35% did not.

Some years later, Välijärvi and Tuomi (1995) investigated upper secondary school as a learning environment. Their sample totalled 2,850 first- and second-year students: 75% of them said that their teachers emphasised the importance of the Matriculation Examination either very often (43%) or fairly often (32%). Välijärvi and Tuomi (1995: 49) concluded that the "shadow of the Matriculation Examination is cast, according to students' observations, quite strongly on the everyday work of upper secondary school". Considering that the respondents were all first- or second-year students, this conclusion seems well warranted: at that time the Matriculation Examination was taken at one time only, which generally was during the spring term of their third year, so the respondents had a rather long time left before taking the exam.

In 2009, the evaluation of pedagogy in Finnish upper secondary education (Väljörvi, Huotari, Iivonen, Kulp, Lehtonen, Rönholm, Knubb-Manninen, Mehtäläinen & Ohranen 2009) surveyed 8,500 third-year upper secondary students. One item in their questionnaire dealt with the Matriculation Examination: "The teachers teach only for the Matriculation Examination". Thirty-five percent of the respondents agreed with the statement while 45% disagreed with it.

3 The present study

3.1 Aims and research questions

The present article is part of a larger study, the purpose of which was to discover what the students at our school thought of assessment in their upper secondary English studies. One topic area of the study was the Matriculation Examination, which is the focus of this article.

The research questions of this article are:

1. Does the Matriculation Examination cause an excessive washback effect? In other words, do students feel that English teachers 'teach to the test' in the upper secondary school? Do students themselves feel that they study for the test alone?
2. Does the English test in the Matriculation Examination evoke test anxiety?
3. Do students consider the Matriculation Examination test a more valid and reliable way of showing their English skills than teacher-based assessment?

3.2 Educational setting

Practically all participating students had started studying English in Year 3 in primary school. Thus far, they had studied EFL for nearly nine or ten years, totalling around 700 or 800 lessons.

Finnish upper secondary school studies are divided into courses, each with 38 lessons. In 2014, there were six compulsory and two specialisation courses of English (Advanced syllabus) and their general guidelines and syllabi were defined by the *National core curriculum for upper secondary schools 2003*. In addition, each school could also offer school-based courses. Although each English course has a theme, they comprise all areas of both oral and written language skills. Hence, course assessment does not focus on any one area, such as grammar, writing or speaking only, but should include them all. Each course is assessed as an independent entity with a numerical

grade (4–10, 10 being the best). In addition to the grade, the student could also be given more detailed assessment and feedback either in writing or orally. (For further information, see *National core curriculum for upper secondary schools 2003*).

All course assessment is teacher-based assessment. The only high-stakes test is the Matriculation Examination, which the students take towards the end of their upper secondary studies. Although the English test is not a compulsory part of the Examination, nearly all students take it, both nationally (Mehtäläinen & Välijärvi 2013) and in this school.

3.3 Participants

The second- and third-year students of our upper secondary school were invited to participate in this study in March 2014. Out of 199 students, 146 answered the questionnaire (response rate 73.4%), and 142 of them answered all the questions regarding the Matriculation Examination. Out of those 142 students, 77 were second-year students, who answered the questionnaire during one of their English lessons. Third-year, i.e. final-year, students answered in their own time (65 respondents) as most of them, preparing to take several subtests of the Matriculation Examination (ME) that spring, did not have lessons at school any more.

Altogether 63 students (44.4% of all respondents) had already taken the English ME test, or part of it. They all were third-year students. Fifty-five of these students had passed the test, but seven of them were retaking the test that spring in the hope of improving their grade. In addition, eight students said that they were in the process of taking the test for the first time that spring. As the students answered the questionnaire in their own time sometime in March, some of these eight students had already completed the whole English test, some had only taken the listening comprehension part (in February). Nevertheless, all these eight students are included in the group of students who had taken the test.

Seventy-nine students (55.6%) had not yet taken any part of the English ME test: among them, there were two third-year students, but all the rest were second-year students. However, they, too, had probably had some personal experience of the exam format, in particular of the listening comprehension part, as sections of them had most likely been used in some of their most recent English courses.

Eighty-five respondents were female (59.9%), 57 male (40.1%). The average of the students' self-reported previous English grade was 8.58 (min. 6, max.10). So far in upper secondary school, they had studied, on average, 6.7 English courses (range: 4–11) and had 3.7 different English teachers (range 2–7). Although the results cannot be generalised, they give quite an accurate picture of the situation in our school at

the time of the study as the respondents represent the total student population well, regarding both gender and their grades.

3.4 Data collection and analysis methodology

The data was gathered through a web-based questionnaire with altogether more than 100 items and questions. The questionnaire covered several topic areas, for instance students' goal orientation, the assessment methods used in English courses and their usefulness, students' personal experiences of and views on the accuracy, fairness, guidance and agency of assessment, as well as the Matriculation Examination.

The data explored in this article come primarily from the Matriculation Examination section of the questionnaire and its Likert-scale items (see Appendix 1) as well as from the goal-orientation questions (see Appendix 2). Those/these data were analysed quantitatively using descriptive statistics. Students' gender, previous (self-reported) grade as well as the fact whether they had taken (part of) the English ME test or not were used as independent variables. Independent samples T-tests were conducted to test the statistical significance of the differences of means of gender and the test-taking. Pearson correlation coefficients were calculated to analyse the correlations between variables.

There were also two open-ended questions dealing with the Matriculation Examination in the questionnaire. Their answers offered qualitative data which were analysed through content analysis (e.g. Patton 2002). First, the content analysis started as inductive analysis "discovering patterns, themes, and categories in one's data" (Patton 2002: 453). However, as the emerging categories and themes, in particular with open-ended answers to Question 9, seemed to match the quality criteria for assessment presented in literature, the second round of content analysis turned into deductive content analysis (e.g. Patton 2002). In other words, at that stage the data were re-categorised according to already existing quality characteristics of validity, reliability and fairness.

4 Findings

My original hypothesis was that whether the students had already taken the test, or not, would somehow affect their answers. Therefore, the results show the descriptive statistics of all respondents first, but also those of the sub-categories of the students who had already taken the test and those who had not as well as female and male students.

4.1 Does the Matriculation Examination cause an excessive washback effect?

The critics of high-stakes testing have been concerned that high-stakes testing narrows teaching and learning. Therefore, the first research question of this study focused on a negative washback effect: Do students feel that English teachers ‘teach to the test’ in the upper secondary school? Do teachers teach to the test only? Do students themselves feel that they study for the Matriculation Examination alone?

In general, 70 percent of the respondents did not think that their teachers taught to the Matriculation Examination only (see Figure 1.). However, the number of students who said that teachers did indeed teach to the test only was greater among the students who had either already passed the exam or were in the middle of taking it.

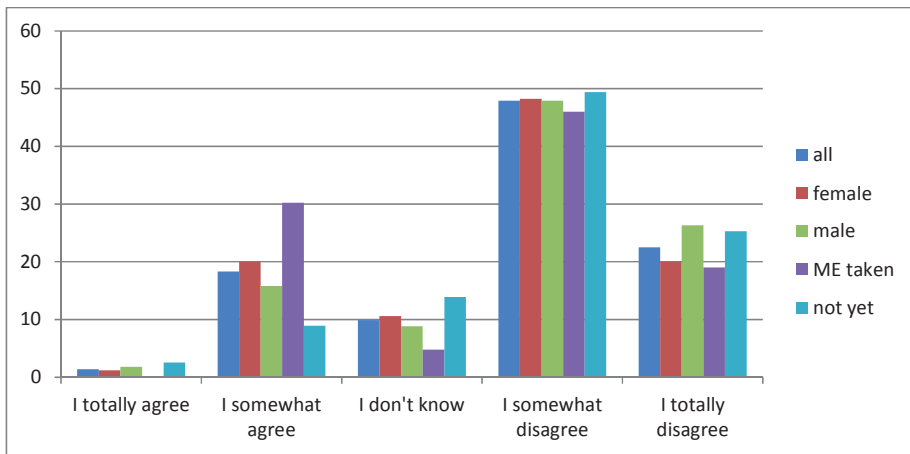


FIGURE 1. Item Teachers teach for the Matriculation Examination only and its responses.

Furthermore, 40% of the students who had not taken the test yet said that their teachers had guided and instructed them too little for the ME test (see Figure 2). Only a good 10 percent of the students who had taken the test shared the same view and almost 80% considered the guidance for the ME test adequate. The difference between those who had taken the test ($m=1.98$) and those who had not ($m=2.91$) was statistically very significant ($p=.000$; $r=.391^{**}$). Female students ($m=2.74$) seemed to consider the guidance for the ME test somewhat less adequate than male students ($m=2.26$, $t=2.453$, $df=140$, $p=.015$; $r=-.203^*$).

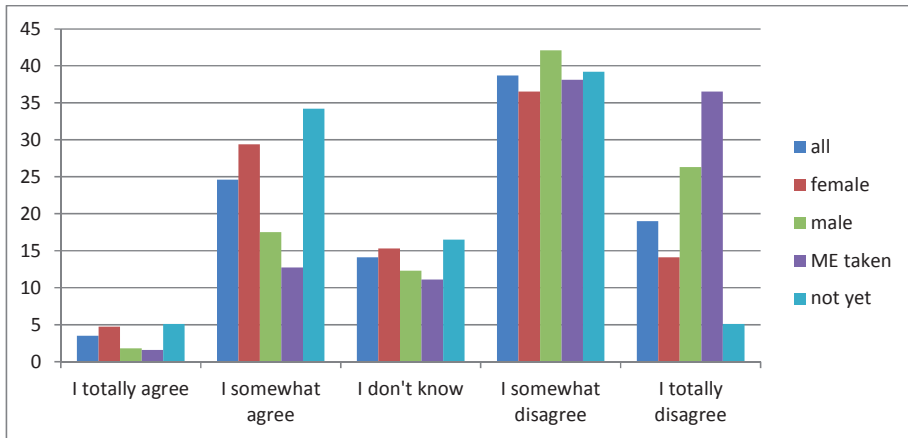


FIGURE 2. My teachers have instructed me too little for the Matriculation Examination.

At the beginning of the questionnaire, the students were asked how much some goals had influenced their studies in the upper secondary school (see Appendix 2). Over 85% of all the respondents said that a good success in the Matriculation Examination had been a goal that had affected their studies either very much or quite a lot. The Matriculation Examination thus seemed to be an important goal – and even more important than a good upper secondary school certificate (see Figure 3).

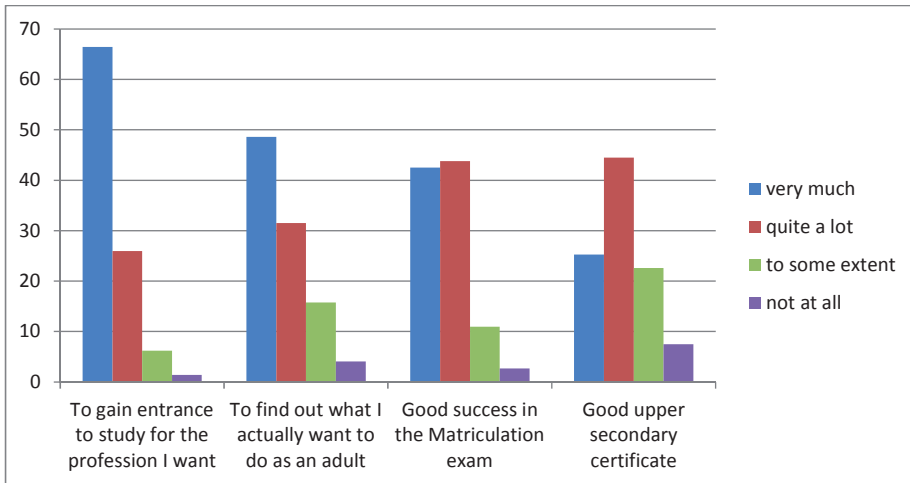


FIGURE 3. To what extent have the following goals guided your upper secondary studies?

Yet again, when asked why they studied English, the results changed. Only about 30% of the respondents said that a good grade in the Matriculation Examination was the most important goal of their upper secondary English studies whereas approximately 55% of the respondents disagreed (see Figure 4). Quite unanimously, the respondents agreed that they studied English primarily for their own future and not for the Matriculation Examination (see Figure 5).

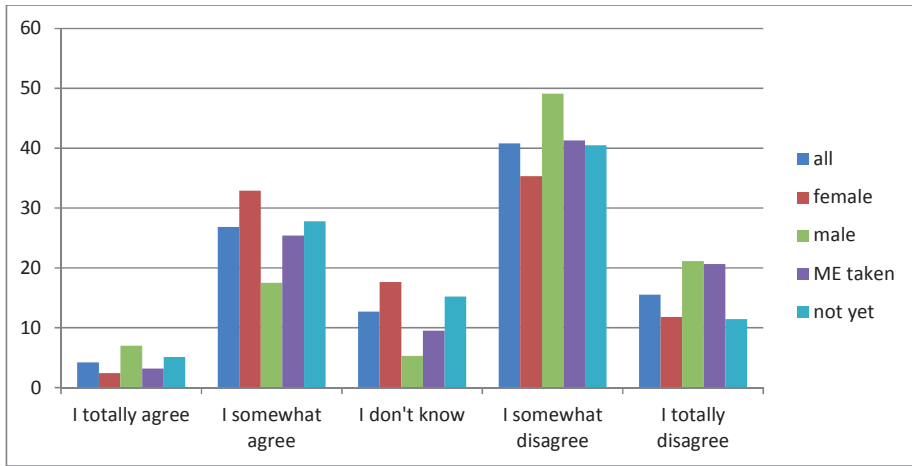


FIGURE 4. The most important goal for me in my English studies is a good grade in the Matriculation Examination.

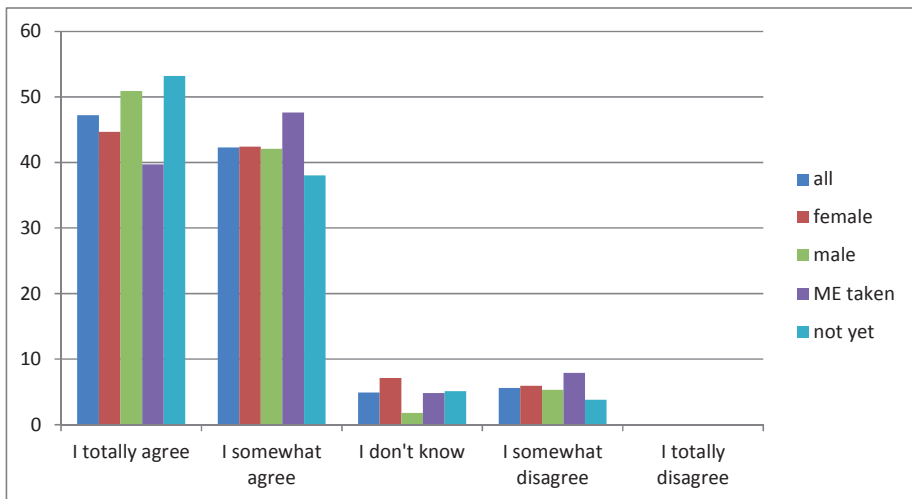


FIGURE 5. I study English for life and for my future, not for the Matriculation Examination.

So, the Matriculation Examination was an important goal that strongly influenced upper secondary studies. However, the primary goal for students' English studies was their own future, not the Matriculation Examination. According to the students in this study, although teachers did not seem to teach to the test at least during the first two years of the upper secondary studies, teaching to the test seemed to increase when the exam approached. Students' earlier success in the English studies, i.e. their previous English grade, did not correlate with any of these items discussed above.

4.2 Does the Matriculation Examination evoke test anxiety?

The second research question dealt with test anxiety, another concern that the critics of high-stakes testing have raised. The results of this study seem clear: the Matriculation Examination evoked some fear or test anxiety in about 60% of the respondents. However, the fear or anxiety seemed to grow a bit milder with the passing of the test, as can be seen in Figure 6. Female students were clearly more susceptible to ME anxiety, and among them, the anxiety was significantly higher ($m=3.84$) than among male students ($m=2.70$, $t=5.108$, $df=101.933$, $p=.000$; $r=-.411^{**}$). In fact, half of the male students did not seem to suffer from any Matriculation Examination anxiety. Students' previous grades did not correlate with anxiety ($r=-.125$).

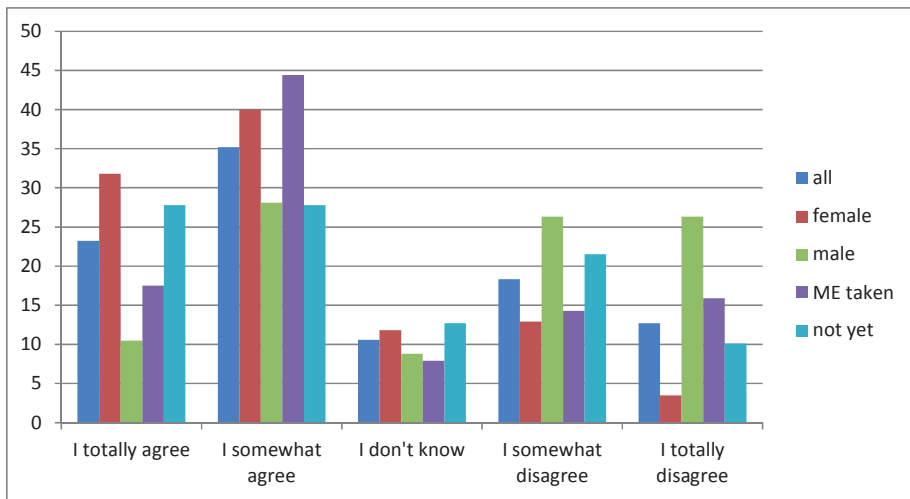


FIGURE 6. The Matriculation Examination scares me.

Approximately one student in four also mentioned either stress or anxiety in their open-ended answers to Question 9: "What do you think of the Matriculation Examination in Advanced English? What kinds of thoughts/emotions does the examination evoke?"

Eight of the students who mentioned stress or anxiety had already taken the test. Their stress or anxiety was mostly linked with the test-taking situation or with the high-stakes of the exam:

The test situation itself is unnerving and exhausting, so the test performance does not always correspond with the real performance.

In listening comprehensions, in particular, stress decreases scores and thus doesn't give a totally reliable picture of the skills.

I'm scared of the ME. test because it affects further studies so much.

Perhaps surprisingly, the students who had not taken the test yet mentioned anxiety or apprehension more often: twenty students (out of 61) mentioned that they were anxious because of the test. Their anxiety or fear ranged from slightly anxious excitement to strong fear that had affected their study plans:

Haven't done it yet. Mostly anxiety and fear, because I'm scared that I will totally fail in the test even though my English skills are quite good in my opinion.

I fear that ME test the most in the upper secondary school. I chose Advanced Maths so that I won't have to take the Advanced English exam. I'm beside myself with fear because I don't believe I'll pass it with dignity.

On the other hand, ten students were confident of their skills and not worried or anxious about the test:

I'll pass it even if I had my eyes shut and hands tied behind my back.

It's quite normal, doesn't evoke any feelings, really.

In sum, expectations, perhaps based on other students' anecdotes, seemed somehow stronger, either more anxiety-ridden or more relaxed and confident, than the actual experiences. As with the quantitative answers, attitude and also anxiety seemed to grow more realistic and perhaps milder when lived through.

As expected, it was quite difficult for me. I had quite a lot of pressure in the test, but I managed well, considering my skills.

Yet, in sum, approximately 60% of all the respondents, and over 70% of female respondents, said that the English test of the Matriculation Examination frightened them at least to some extent.

4.3 Students' views on the validity, reliability and fairness of the ME English test

The proponents of high-stakes tests say that high-stakes testing, designed by assessment experts, is more valid and reliable as an assessment tool than, say, teacher-based assessment (e.g. Cizek 2005). The last research question of this article was to see if the students agreed with this notion.

The questionnaire had two items that addressed the validity and reliability of the Matriculation Examination English test. One of them read as follows: *In the Matriculation Examination, I can reliably show how good my English skills are.* Nearly 60% of the students who had already taken the test did not agree with the claim (see Figure 7). However, approximately a third agreed with the statement. Male students ($m=3.18$) in general seemed to trust the reliability of the Matriculation Examination more than female students ($m=2.39$, $t=-3.943$, $df=140$, $p=.000$, $r=.316^{**}$). There were no statistically significant differences between the students who had already taken the test and those who had not. Students' previous English grade did not correlate with this item ($r=.098$).

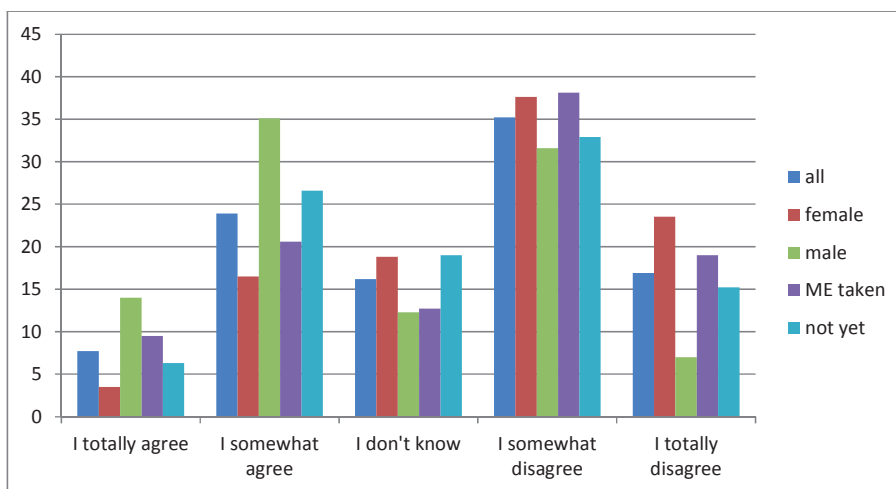


FIGURE 7. In the Matriculation Examination, I can reliably show how good my English skills are.

The second item compared the accuracy of teacher assessment with that of the Matriculation Examination: *The assessment given by the teacher gives a more accurate picture of my skills than the ME test.* Once again, nearly 60% of those who had taken the test thought that the course assessments gave a more accurate assessment of their skills than the Matriculation Examination test (see Figure 8). Only approximately 13% of them disagreed with that claim, leaving 30% undecided. Somewhat surprisingly,

although female students seemed to consider the Matriculation Examination a clearly less reliable format to demonstrate their skills than male students, there were no statistically significant differences between male ($m=3.47$) and female students ($m=3.58$) in this item; nor did the previous grade correlate with this item ($r=.121$).

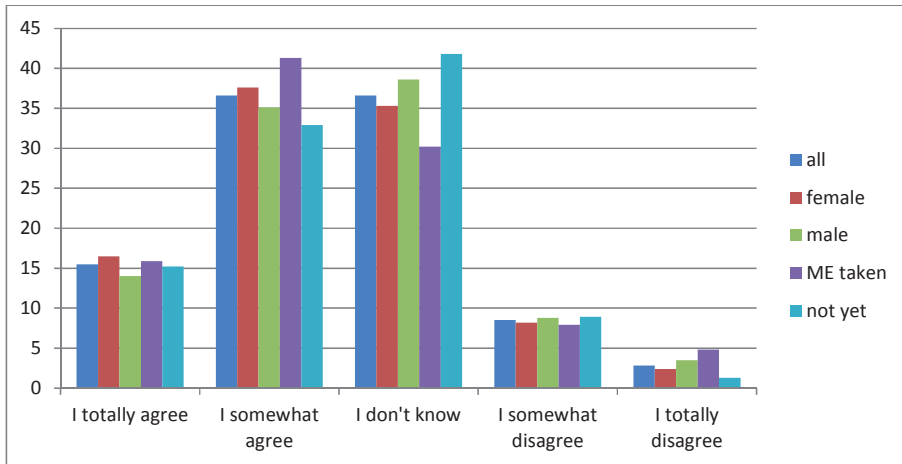


FIGURE 8. The assessment given by the teachers gives a more accurate picture of my skills than the ME test.

In addition, students' open-ended answers illuminated the students' experiences of and expectations of the reliability and validity of the Matriculation Examination test. The students readily volunteered answers: 58 out of the 63 students who had already taken (part of) the test answered the following question: (Q9) *What do you think of the Matriculation Examination in Advanced English? What kinds of thoughts/emotions does the examination evoke?* In the following account, I will concentrate on their answers, because of their first-hand experience. However, I will also briefly mention the expectations of those students who had not taken the test yet. All these answers are categorised according to the main quality requirements of assessment, i.e. validity, reliability and fairness.

4.3.1 Validity: does the test measure what it is supposed to measure?

Out of those 58 students who had taken the test and volunteered open-ended answers, three complimented the test whole-heartedly:

Good, versatile test. Seems that they know their business in the Matriculation Board.

In addition, eight students regarded the test as good, but also offered some criticism or suggested some improvements:

Listening comprehensions are quite difficult but otherwise it is suitable. Essays have sometimes rather bad [topic] options as you should have specific knowledge or experience on things.

In my opinion, the test was good but to my mind an oral test should be part of the package because oral communication is important.

Nonetheless, 35 students (out of 58) questioned the validity of the Matriculation test in one way or another. First of all, the test did not assess students' oral skills in any way, which was criticised in 11 answers:

The test is deficient in the sense that it doesn't measure the student's ability to communicate orally in English.

Oral component is missing. Yet, it's one of the main elements of language skills.

The students also commented on the difficulty or 'excessive difficulty' of the test (21 mentions), which surpassed the difficulty level of the English courses (4 mentions):

Quite challenging, but some structures are really challenging and the teaching during the courses doesn't match their difficulty.

You can't do well by just attending the English courses offered at school - - The vocabulary and reading comprehensions are more difficult than in the English courses.

Students mostly criticised the test for testing too detailed grammatical knowledge (12 mentions) or vocabulary (8) which were "not important or relevant for good language skills" or real life:

We learn languages so that we could encounter new people and get to know different cultures. - - This is something the Matriculation Board doesn't seem to understand when they include excessively difficult lottery exercises that test the grammatical knowledge of the exceptions to the exceptions.

Vocabulary was impossible for an average student wishing for a good grade.

The difficulty level is rising all the time and the vocabulary needn't be quite so scientific.

However, four students understood the difficulty of the test:

The English ME test is frighteningly difficult but I guess that separates the best from the rest.

Four students mentioned consequential validity – i.e. the impact – of the Matriculation Examination test on their further studies in their answers:

Despite the unfairness of the test, in the eyes of further studies institutions, the applicant's English skills are directly comparable with the letter that stands in the Matriculation Examination certificate.

Overall, the students who had taken the test seemed quite critical of its validity. How about the students who had not taken the test yet? Altogether 61 out of 79 students volunteered answers to Question 9. Their answers were not as detailed as the answers of those who had personal experience, and quite a few students also expressed that their answers were based on expectations and other students' stories, not their own experiences. Nonetheless, similar validity issues emerged:

It's a bit too hard for an ordinary Finn because even native speakers have problems with it at times.

It contains too much of all sorts of nitpicking that isn't really that much relevant in the development of practical English language skills.

Altogether, eight students criticised exercises for focusing on too detailed knowledge and four students for the lack of an oral part in the test. The difficulty of the test was mentioned ten times and its irrelevance for real life languages skills an additional five times.

All in all, many students seemed quite critical of the validity of the Matriculation test. Its content validity was not regarded as particularly good because speaking was not tested. Furthermore, too detailed knowledge of grammatical exceptions or rare vocabulary was considered irrelevant for real-life language skills. The difficulty level was also seen as too demanding when compared to the syllabi of Advanced English courses.

4.3.2 Reliability: is the ME test a reliable and accurate way to show one's skills?

Once again, many students who had already taken the test (25 out of 63) mentioned various threats to reliability, i.e. many sources of construct-irrelevant variance (Black & Wiliam, 2012). The greatest threats, according to them, were trick questions and red herrings (17 students). Many students compared answering these questions to the draw of the lottery numbers.

The questions and answers lead you astray, to answer wrong...even if you understand the text/what you hear, the options in the answers trick you to answer wrong, and that is not right.

The listening comprehension test includes too many so called trick questions and thus doesn't really measure your language skills

Those bloody multiple choice trolls irk me every time, but of course similar situation may come up in real life too.

In addition to those 17 students, three students also mentioned luck as a possible factor affecting results.

Because there's only one exam, the result depends very much on test exercises, and doesn't necessarily give the right overall picture of the student's skills.

Another threat to reliability was the test-situation itself with its time constraints and pressure. Four of these seven students specified listening comprehension tests.

The stressful situation affects your results too and all your skills won't necessarily come out as well as possible.

I think the listening comprehensions are unfair because they try to bluff the student deliberately and the pauses are so short that you don't have time to read the questions then. So, the results don't give the right picture of your skills then.

Nevertheless, and perhaps slightly surprisingly, none of the 58 students mentioned any concerns about reliability in the sense of inter- or intra-rater reliability.

Out of the students who had not yet taken the test, ten mentioned trick questions. Chance or luck with the topics of the test or with the test's difficulty was mentioned in three answers.

Exercises made weird and tricky on purpose and not a test that is made on the basis of the real language skill needs.

Scared, because the difficulty level varies so much between years.

Furthermore, five students also concluded that the test did not necessarily measure or capture one's real English skills.

Although I feel that I'm pretty good at English, I'm scared that the test will go badly and everybody will get the wrong image of my skills.

4.3.3 Fairness: Is the scoring and grading of the Matriculation Examination fair?

The other open-ended question dealt with the Matriculation Examination grade and its accuracy and fairness: *If you have already taken the Matriculation Exam in English, did you get the grade you deserved in your opinion? Why/why not?(Q5)*

Within the Finnish Matriculation Examination system, if a student has passed any ME sub-test, he or she can retake it once in an attempt to improve the grade. The better grade of these two attempts will be the official grade. Out of those 48 students who had already completed the English ME test and were not going to re-sit it, 45 students answered the question. Twenty-five of them said that the grade had been what they had deserved; for one it was more than she had expected.

Yes, the grade corresponds with my skills and is in line with my course grades.

Yes, I went there to get a certain grade, and I got it in the end.

I got a far higher grade than I thought so I was happy with the result.

An additional five students said that even if they were not quite satisfied with their grades, they thought they had deserved it for one reason or another:

I would have wished for a better grade, but in my opinion I deserved that grade because I just could not do better then.

Eleven students, however, did not consider the grade to be what they would have merited. In their opinion, the grade was not in line with their course grades or with their real skills. Furthermore, some students criticised both the excessive difficulty as well as the focus and format of the ME test – in other words, the same issues discussed earlier with validity or reliability:

The vocabulary in the test was really challenging, and it went badly. In my opinion, I can use English much better than what the grade suggests.

Multiple choices have often questions and options that are somehow bad: several right answers, no completely correct option or a question that can be interpreted in several ways. In open-ended questions you can't guess/deduce what sorts of things they want in the answer. I understand everything but can't always get my answer 'right'.

All seven students who were going to retake the exam answered the question. Five of them said they had not got the grade they felt they deserved:

In my opinion, no, because I got better results from the tests we did as prep tests than from the real one and that bugs me.

No, I didn't. In my opinion, the grade doesn't reflect my skills because the listening test was a very unnerving experience for me and therefore, anxiety probably ruined my performance. After that when the written part came I was as if I had lost all my hope since I knew I couldn't reach the grade I wanted by any means.

No because it doesn't match my course grades.

Two of them, however, regarded the grade as deserved:

Yes, I put too little effort into it.

To summarise, most students seemed to think that they had been quite fairly scored and graded in the ME English test, and that the grade they got was mostly deserved for that particular test. However, they did not seem to think that the Matriculation Examination test itself was a most valid or reliable way to show their skills.

5 Discussion and conclusions

Several opponents of high-stakes tests blame them for a negative washback effect that narrows the curriculum into teaching to the test. Therefore, one of the aims of this article was to find out whether the Matriculation Examination caused a washback effect in this school, and whether students considered the potential washback effect excessive. In the 1995 study by Välijärvi and Tuomi that seemed to be the case. In this study, the results were quite the contrary: 40% of the second-year students felt that their teachers had instructed them too little for the future Matriculation Examination.

In another study by Välijärvi et al. (2009), a good third of the respondents, who all were third-year students, said that their teachers taught only for the Matriculation Examination; nearly half of the respondents, however, disagreed. Although two thirds of all the respondents in this study did not think that their teachers taught only for the Matriculation Examination, 30% of the third-year students thought that they actually did. In that respect, the result is somewhat in line with that of Välijärvi et al. (2009). In sum, the Matriculation Examination seems to have quite a strong washback effect during the final upper secondary courses, but not earlier. Thus, the washback effect cannot perhaps be considered excessive. However, although the Matriculation Examination should be based on the upper secondary school curriculum and its syllabi, the examination is not part of the upper secondary curriculum per se.

Yet, the Matriculation Examination is still a highly important goal for the students, and, therefore, probably also for their teachers. First and foremost, however, the students regarded English as a life skill; almost all the respondents said that they studied English for their future, not for the Matriculation Examination.

Although the Matriculation Examination does not influence the upper secondary studies as much as it may have done in the past, it still 'casts a shadow' on students' daily work in the form of apprehension, stress and anxiety, with nearly two-thirds of the

respondents saying that the exam scared or frightened them. Female students seemed to be more anxious than male students. The reasons the students mentioned for anxiety were, for instance, the pressurised test-taking situation as well as the consequences of the exam for their further studies. Hence, this study corroborates the findings of earlier studies that high-stakes testing causes test anxiety and that female students are more vulnerable to it (e.g. Cassady 2010; Hembree 1988).

Is anxiety a necessarily evil, in other words, is the examination so important and excellent that it is worth the anxiety it seems to cause? The third and final research question of this article focused on students' experiences of and expectations for the validity, reliability and fairness of the Matriculation Examination test as a test of their English skills. Out of those students who had already taken the test, over half thought that the Matriculation Examination test was not a reliable way to show their skills and considered teacher-based assessments a more accurate assessment of their skills. Not everybody agreed with them, though, and students did not seem totally convinced that teacher-based assessment would necessarily be much better.

Yet, many students seemed quite critical of the validity of the Matriculation test in their open-ended answers. Its content validity, or content relevance and coverage, was not regarded as particularly good because speaking was not tested. Furthermore, too detailed knowledge related to grammatical exceptions or rare vocabulary was considered irrelevant for real-life communication skills. The difficulty level was also seen as too demanding when compared to the goals and syllabi of Advanced English courses.

The reliability of the test was not considered very high, either. Students who had already taken the test mentioned various sources of construct-irrelevant variance (Black & Wiliam 2012; Messick 1996), in other words, several threats to reliability. Deliberately tricky questions "that lead you astray, to answer wrong" were considered the greatest threat to reliability (see also Anckar 2011). Quite a few students compared answering tricky multiple-choice questions to pure guessing (see also Anckar 2011). The pressurised test-taking situation and luck were also regarded as threats to the reliability of the test. Hence, the students did not seem convinced that they could show their English skills very reliably in the Matriculation Examination test. Yet, although not necessarily happy with the test and its format, the students seemed to consider the scoring and grading of their test papers quite fair.

This study has many limitations. First of all, it was limited to one school only, and thus the findings cannot be generalised. Furthermore, the academic achievement of the student population in this school is well above national average, also in the Matriculation Examination. Thus, these data do not include many views or experiences of students who struggle with their upper secondary school studies or who risk failing the ME English test. Although the previous English grade did not correlate with any of

the findings in this study, the experiences of and expectations for the ME English test might look different in larger and more varied student populations. Furthermore, had there been more questions dealing with the Matriculation Examination test, or different questions, the findings might have changed. Also, the data analysis methods employed in this article were quite basic. Thus, other data collection and analysis methods would most probably have yielded additional, or different, information. However, this small study sheds some light on students' experiences of the English ME test and also brings forth many interesting questions that still remain unanswered. Thus, further research on students' experiences of the Matriculation Examination with more varied student groups as well as data collection and analysis methodology is clearly needed.

Even though public discussion on the possible abolition of the Matriculation Examination sometimes surfaces, the exam enjoys a high status in Finland (Vuorio-Lehti 2007). Students were not asked directly whether they considered the exam needed or not in this study but my guess, a pure hunch, is that most of the students who have passed the exam would not like to abolish the examination. It seems to be a rite of passage that is part of the school-leaving tradition (Vuorio-Lehti, 2006).

Nevertheless, the students in this study voiced several concerns over the English test which are worth careful attention. Firstly, assessing speaking should somehow be part of the examination. The Matriculation Examination Board has announced that oral production will, sometime in the future, be included in the test. Secondly, the test format should perhaps be reconsidered. As cost-efficient and seemingly reliable (at least in the sense of rater reliability) as the multiple-choice questions are, is there over-reliance on them in the foreign/second language tests? At the moment, approximately half of the total test score, and most of the reading and listening comprehension score, is based on multiple-choice questions. Furthermore, although the difficulty and trickiness of the items may create variance in test results conveniently, is this variance necessarily fair? Moreover, is that variance not too much based on construct-irrelevant variance (Black & Wiliam, 2012; Messick, 1996)? Also, because of the pressurised test-taking situation (as it is the case with the listening comprehension part, in particular) is the test equally fair for all students – including those who suffer from test anxiety?

What should be done? The idea of using the Matriculation Examination results even more extensively for the admission to further education, as suggested, would raise the stakes of the examination considerably. That would also increase the pressure. That, in turn, might have detrimental effects on teaching and learning, as several studies have shown elsewhere. The shadow of the Matriculation Examination would certainly grow longer, and probably darker, again. How would all that accord with the new *National core curriculum for general upper secondary schools 2015* that emphasises versatile assessment methodology, assessment *for* learning, promoting and encouraging students' learning, as well as self-assessment, for instance? No matter how excellent a

test, one single test should never have too much power over a student's future. And as the students in this study have pointed out, there is room for much improvement in the present Matriculation Examination and its English test.

References

- Alderson, J. C. & L. Hamp-Lyons 1996. TOEFL preparation courses: a study of washback. *Language Testing*, 13 (3), 280–297.
- Amrein, A. & D. Berliner 2002. High-stakes testing & student learning. *Education Policy Analysis Archives*, 10 (18). DOI: <http://dx.doi.org/10.14507/epaa.v10n18.2002>
- Anckar, J. 2011. *Assessing foreign language listening comprehension by means of the multiple-choice format: processes and products*. Jyväskylä: University of Jyväskylä.
- Atjonen, P. 2015. *Kehittävä arviointi kasvatusalalla*. Joensuu: Kirjokansi.
- Aydin, S. 2009. Test anxiety among foreign language learners: a review of literature. *Journal of language and linguistic studies*, 5 (1), 127–137.
- Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Black, P. J. 1998. *Testing, friend or foe? The theory and practice of assessment and testing*. London: Palmer Press.
- Black, P. & D. Wiliam 2012. The reliability of assessments. In J. Gardner (ed.) *Assessment and Learning*. (2nd ed.) London: Sage, 243–263.
- Cassady, J. C. 2010. Test anxiety: contemporary theories and implications for learning. In J. C. Cassady (ed.) *Anxiety in schools: the causes, consequences, and solutions for academic anxieties*. New York: Peter Lang, 5–26.
- Cheng, L., Y. J. Watanabe & A. Curtis 2004. *Washback in language testing: research contexts and methods*. Mahwah N.J: Lawrence Erlbaum.
- Cizek, G. J. 2005. High-stakes testing: contexts, characteristics, critiques, and consequences. In R. Phelps (ed.) *Defending Standardized Testing*. Mahwah, NJ: Lawrence Erlbaum, 23–54.
- Elman, B. A. 2000. *A cultural history of civil examinations in late imperial China*. Berkeley and Los Angeles: University of California Press.
- Hanson, F. A. 1993. *Testing testing: social consequences of the examined life*. Berkeley: University of California Press.
- Harlen, W. 2010. What is quality teacher assessment? In J. Gardner, W. Harlen, L. Hayward & G. Stobart (eds.) *Developing teacher assessment*. Maidenhead: Open University Press, McGraw-Hill Education, 29–52.
- Harlen, W. 2012. The role of assessment in developing motivation for learning. In J. Gardner (ed.) *Assessment and learning*. (2nd ed.) London: Sage, 171–183.
- Harlen, W. & R. Deakin Crick 2003. Testing and motivation for learning. *Assessment in Education*, 10 (2), 169–208.
- Hembree, R. 1988. Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58 (1), 47–77.
- Heubert, J. P. & R. M. Hauser 1999. *High stakes: testing for tracking, promotion, and graduation*. Washington D.C.: National Academy Press.
- Jones, G., B. Jones & T. Hargrove 2003. *The unintended consequences of high-stakes testing*. Lanham, USA: Rowman & Littlefield Publishers.
- Kaarninen, M. & P. Kaarninen 2002. *Sivistyksen portti: Ylioppilastutkinnon historia*. Helsinki: Otava.

- Kornhaber, M. L. & G. Orfield 2001. High-stakes testing policies: examining their assumptions and consequences. In G. Orfield & M. L. Kornhaber (eds) *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press, 1–18.
- Lindström, A. 1998. *Ylioppilastutkinnon muotoutuminen autonomian aikana*. Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.
- Madaus, G. & M. Clarke 2001. The adverse impact of high-stakes testing on minority students: evidence from one hundred years of test data. In G. Orfield & M. L. Kornhaber (eds.) *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press, 85–106.
- Matriculation Examination Board. <https://www.ylioppilastutkinto.fi/fi/>. See also https://www.ylioppilastutkinto.fi/images/sivuston_tiedostot/Ajankohtaista/HS_28052016.pdf [retrieved 20.9.2016].
- Mehtäläinen, J. & J. Välijärvi 2013. *Ylioppilaskokeiden arvosanojen vertailtavuus eri aineissa vuosina 2007–2011*. Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.
- Messick, S. 1993. Validity. In R. L. Linn (ed.) *Educational measurement*. (3rd ed.) Phoenix (AZ): Oryx, 13–103.
- Messick, S. 1996. Validity and washback in language testing. *ETS Research Report Series* (1), i–18. *National core curriculum for upper secondary schools 2003*. (English translation printed in 2004.) Helsinki: Finnish National Board of Education.
- National core curriculum for general upper secondary schools 2015*. (English translation printed in 2016). Helsinki: Finnish National Board of Education.
- Nichols, S., G. Glass & D. Berliner 2006. High-stakes testing and student achievement: does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14 (1). DOI: <http://dx.doi.org/10.14507/epaa.v14n1.2006>.
- Patton, M. Q. 2002. *Qualitative research & evaluation methods*. (3rd ed.) Thousand Oaks, CA: Sage.
- Race, P., S. Brown & B. Smith 2005. *500 tips on assessment*. (2 ed.) London: RoutledgeFalmer.
- Stobart, G. 2008. Testing times: the uses and abuses of assessment. New York, N.Y.: Routledge.
- Syrjälä, L. 1989. *Oppilasarviointi osana lukio-opiskelua ja opetusta: oppilaiden ja opettajien näkemyksiä ja kokemuksia Alppilan lukiossa*. Oulu: Oulun yliopisto.
- Volante, L. 2004. Teaching to the test: what every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, (35). Available at <http://files.eric.ed.gov/fulltext/EJ848235.pdf>.
- Vuorio-Lehti, M. 2006. *Valkolakin viesti: ylioppilastutkintokeskustelu Suomessa toisen maailmansodan jälkeen*. Turku: Turun yliopisto.
- Vuorio-Lehti, M. 2007. Valkolakin hohde: keskustelua ylioppilastutkinnon merkityksestä Suomessa toisen maailmansodan jälkeen. *Kasvatus & Aika*, 1 (1), 19–33.
- Välijärvi, J., N. Huotari, P. Iivonen, M. Kulp, T. Lehtonen, H. Rönholm, G. Knubb-Manninen, J. Mehtäläinen & S. Ohranen 2009. *Lukiopedagogiikka*. Jyväskylä: Koulutuksen arviointineuvosto. Koulutuksen arviointineuvoston julkaisuja, 40.
- Välijärvi, J. & P. Tuomi 1995. *Lukio nuorten valintojen ja oppimisen ympäristönä*. Jyväskylä: Kasvatustieteiden tutkimuslaitos.

APPENDIX 1.

The nine Likert-scale items dealing with the Matriculation Exam with their percentages, means and standard deviations.

	Täysin samaa mieltä <i>I strongly agree</i>	Jokseenkin samaa mieltä <i>I agree</i>	En osaa sanoa <i>I don't know</i>	Jokseenkin eri mieltä <i>I disagree</i>	Täysin eri mieltä <i>I strongly disagree</i>	m	sd
Opiskelen englantia elämää ja tulevaisuuttani enkä yo-kirjoituksia varten. <i>I study English for life and my future, not for the Matriculation Examination.</i>	47.2	42.3	4.9	5.6		4.31	.809
Yo-kirjoitukset pelottavat minua. <i>The Matriculation Examination scares me.</i>	23.2	35.2	10.6	18.3	12.7	3.38	1.357
Opettajien antama arviointi antaa oikeamman kuvan osaamisestani kuin yo-koe. <i>The assessment given by the teachers gives a more accurate picture of my skills than the ME test.</i>	15.5	36.6	36.6	8.5	2.8	3.54	.950
Yo-kirjoitusten arviointi ei vastaa opettajien arviointikäytänteitä. <i>The assessment and grading of the Matriculation Examination doesn't correspond with those of the teachers.</i>	7.7	36.6	38.0	16.9	0.7	3.34	.874
Voin yo-kirjoituksissa luotettavasti osoittaa, kuinka hyvin englantia osaan. <i>In the Matriculation Examination, I can reliably show how good my English skills are.</i>	7.7	23.9	16.2	35.2	16.9	2.70	1.225

Lukiassa pitäisi käyttää vain samoja arviointitapoja kuin yo-kirjoituksissakin. <i>In upper secondary school, only the same assessment methods that are used in the Matriculation Examination should be used.</i>	4.9	26.8	10.6	41.5	16.2	2.63	1.183
Englannin opintojeni tärkein tavoite minulle on hyvä arvosana yo-kirjoituksissa. <i>The most important goal for me in my English studies is a good grade in the Matriculation Examination.</i>	4.2	26.8	12.7	40.8	15.5	2.63	1.158
Opettajani ovat opastaneet minua liian vähän yo-kirjoituksia varten. <i>My teachers have instructed me too little for the Matriculation Examination.</i>	3.5	24.6	14.1	38.7	19.0	2.55	1.158
Opettajat opettavat vain ylioppilaskirjoituksia varten. <i>Teachers teach for the Matriculation Examination only.</i>	1.4	18.3	9.9	47.9	22.5	2.28	1.054

APPENDIX 2.

The goal-orientation items:

Missä määrin seuraavat tavoitteet ovat ohjanneet lukio-opiskeluasi? (%)

To what extent have the following goals guided your upper secondary studies? (in percentages)

	Erittäin paljon <i>Very much</i>	Melko paljon <i>Quite a lot</i>	Jonkin verran <i>To some extent</i>	Ei lainkaan <i>Not at all</i>
Päästä lukion jälkeen opiskelemaan tavoittelemaani ammattiin. <i>To gain entrance to study for the profession I want.</i>	66.4	26.0	6.2	1.4
Selvittää itselleni, mitä isona oikeastaan haluan tehdä. <i>To find out what I actually want to do as an adult.</i>	48.6	31.5	15.8	4.1
Hyvä menestyminen ylioppilaskirjoituksissa. <i>Good success in the Matriculation Examination.</i>	42.5	43.8	11.0	2.7
Hyvä päättötodistus. <i>Good upper secondary certificate.</i>	25.3	44.5	22.6	7.5
Oppia suunnittelemaan opintojani ja tulevaisuuttani. <i>To learn to plan my studies and future.</i>	24.7	47.9	24.7	2.7
Saada hyvä yleissivistys. <i>To get a good all-round education.</i>	24.0	50.7	23.3	2.0
Opiskella mahdollisimman paljon kiinnostavia kursseja. <i>To study as many interesting courses as possible.</i>	21.9	45.2	27.4	5.5
Oppia tuntemaan itseni, vahvuuteni ja heikkouteni. <i>To learn to know myself, my strengths and weaknesses.</i>	18.5	43.2	30.1	8.2
Opetella itse ottamaan vastuuta asioista. <i>To learn to take responsibility.</i>	15.8	45.2	32.9	6.2
Oppia tekemään päätöksiä ja valintoja. <i>To learn to make decisions and choices.</i>	11.0	44.5	37.0	7.5
Oppia ilmaisemaan itseäni. <i>To learn to express myself.</i>	8.9	43.2	40.4	7.5
Oppia tulemaan toimeen erilaisissa ryhmissä ja erilaisten ihmisten kanssa. <i>To learn to get along in different groups and with different people.</i>	8.9	34.9	49.3	6.8
Mennä samoille kursseille kuin kaverinikin. <i>To go to the same courses as my friends.</i>	2.7	15.1	43.8	38.4

Huhta, A. & R. Hildén (toim.) 2016. Kielitaidon arviointitutkimus 2000-luvun Suomessa. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016 / n:o 9. 212–225.

Sofia Stolt

Helsingfors universitet

Kollegial bedömning i summativa bedömarkommentarer på studentexamensuppsatser i modersmål och litteratur

The purpose of this article is to illuminate evaluative remarks and an institutional interaction in a social context. The overall focus is on qualitatively demonstrating patterns of expression that censors, through attitude remarks, use in their grading of matriculation examination essays in Swedish as mother tongue and literature. The matriculation examination in Finland constitutes an institutionalised writing practice and the censors can be seen as authoritative graders in this context. Theoretically and methodically, the study follows the Appraisal system. Through evaluative remarks in interaction with each other the censors show qualities of the essay or the author of the essay, which they as graders want to bring to light. In a dialogue with the other censors, they reference external sources such as standardised policy documents and the teacher's scoring; consensus in their grading is reached by collegial dialogue.

Keywords: appraisal, evaluation, matriculation exam

Nyckelord: appraisal, kollegial bedömning, studentexamen

1 Inledning

I den här artikeln belyser jag värderande uttryck och en institutionell interaktion i bedömarkommentarer. Övergripande fokus ligger på att kvalitativt visa de uttrycksmönster som censorer genom attitydyttranden använder i sin summativa bedömning av studentexamensuppsatser i modersmål och litteratur, samt hur dessa yttranden fyller en interpersonell funktion i bedömningsprocessen. Artikeln baserar sig på min doktorsavhandling (Stolt 2016) som undersöker hur censorer uttrycker sig i bedömarkommentarer på goda och utmärkta studentexamensuppsatser i modersmål och litteratur. Avhandlingen belyser utgående från censorskommentarer en skriftkultur som inte tidigare undersökts. I denna artikel utgår jag dels från samma material som i min avhandling, men här visar jag bedömarkommentarer på inte endast goda texter, utan även på sådana texter som ligger lägre ner på bedömningsskalan, för att åskådliggöra på vilket sätt censorer använder sig av en kollegial bedömning. Teoretiskt och metodologiskt följer studien appraisalssystemet (Martin & White 2005), vilket gör det möjligt att analysera och undersöka retoriska inslag i texter samt se på relationen mellan skribent och läsare. I den här artikeln har jag använt delsystemet ENGAGEMANG inom appraisal för att visa på hur flera censorer som läsare och skribenter är involverade i bedömningsprocessen och hur de kommer fram till en samstämmighet beträffande det slutliga poängtalet för uppsatsen.

Studentexamen i Finland utgör en institutionaliserad skriftpraktik och censorerna kan ses som auktoritativa bedömare i denna kontext (Stolt 2016). Genom värderande yttranden i interaktion med varandra visar censorerna på egenskaper i studentexamensuppsatsen eller hos skribenten som skrivit uppsatsen, vilka de som bedömare vill fästa uppmärksamhet på. I denna artikel ser jag på hur bedömare utgående från en dialog uttrycker attityder i sin bedömning och min avsikt är att lyfta fram vissa typer av språkhandlingar i denna dialog mellan censorerna. Utgående från kommentarer på uppsatser ur olika vitsordsnivåer belyser jag den kollegiala bedömningen i summativa bedömarkommentarer.

2 Material

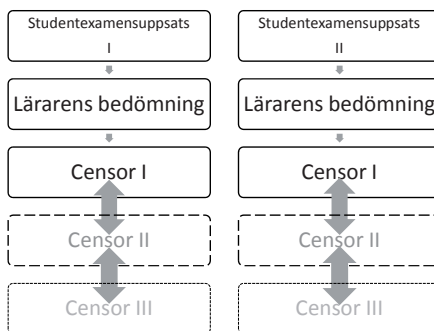
Materialet för denna artikel utgörs av bedömarkommentarer på studentexamensuppsatser i modersmål och litteratur skrivna åren 2004 och 2005, och baserar sig på materialet för min doktorsavhandling (Stolt 2016). Jag gör ett nedslag vid en given tidpunkt och jag närmar mig det med fokus på att belysa tendenser i erfarna bedömares bedömningsvärldar. Elevtexten står inte i fokus och för att garantera anonymiteten hos de parter som förekommer i min undersökning reder jag i beskrivningen av

materialinsamlingen inte exakt hur ut hur urvalet gjorts, utan presenterar det endast i stora drag. De censorer vars kommentarer jag undersöker förekommer i studien som enskilda men anonyma företrädare för sin yrkesmässiga ställning.

Årligen deltar kring 2000 svenskspråkiga abiturienter i studentexamen. I mitt material ingår 360 uppsatser som bedömts med de högsta vitsorden **laudatur** (L) eller **eximia cum laude** (E) och 272 uppsatser som bedömts med **cum laude approbatur** (C) eller med **lubenter** (B), vitsord som ligger lägre ner på vitsordsskalan. Materialet är insamlat från 11 svenskspråkiga gymnasier och utvalt enligt principen att ge en generell bild av Svenskfinland. Poängskalan för de prov som mitt material från våren 2004 och våren 2005 bygger på går från 50 till 100, där allt under 50 poäng är underkänt. I övrigt följer poängskalan en relativ vitsordsfördelning och poänggränsen de aktuella åren var 60 poäng för ett B och 70 poäng för ett C. År 2004 skulle abiturienten ha 93 poäng för att nå ett L, respektive 92 poäng år 2005. (Studentexamensnämnden 2016) Poängskalan är relevant med tanke på beskrivningen av yttrandena i bedömarkommentarerna eftersom censorerna i sin bedömning ofta relaterar till ett tilltänkt vitsord (se Stolt 2016). Bedömningsskalan är också den mest konkreta mätstocken för bedömning av en text och det är av den orsaken viktigt för både bedömare och abiturienter att känna till skalan väl (se Weigle 2011).

3 Censorskommentaren i en institutionell kontext

I och med att censorskommentarerna utgör en aktivitet vars syfte, att summativt bedöma studentexamensuppsatsen och samtidigt fungera som dokumentation på bedömningen, är bestämt på förhand kan censorernas kommentarer på uppsatsen och reaktioner på lärarens poänggivning anses vara institutionella genom att kommunikationen är ämnesmässigt sluten (se Linell 1990). Processen för bedömningen av studentexamensuppsatsen visas i figur 1.



FIGUR 1. Processen för bedömning av studentexamensuppsatsen (utgående från Stolt 2016: 77).

Borgström & Ledin (2014: 144) påpekar att ett viktigt resultat i bedömningsforskning är att lärare inom samma ämne bedömer bäst eftersom dessa har en gemensam förståelse för bedömningsgrunderna, men konstaterar dock att denna tolkningsgemenskap inte garanterar en samstämmig bedömning även om bedömarna är erfarna och kunniga inom ämnet. För detta krävs välutformade bedömningsverktyg som visar på de för prestationen kännetecknande kvalitetsdragen. Svensson (1990: 52) menar att gruppdeltagare med hjälp av språk och sammanhang och genom olika former av diskursiv dominans kan förtydliga sin professionella roll. Den professionella rollen kan i denna studie avspeglas av de yttranden som censorerna gör i sina bedömarkommentarer och som kommer till uttryck genom den appraisalanalys jag gör på mitt material.

Censorerna i Finland är adjungerade medlemmar vid Studentexamensnämnden som verkar under Undervisnings- och kulturministeriet. Genom att betrakta censorssamarbetet ur ett övergripande perspektiv vill jag visa på de skrifthändelser och den skriftpraktik som censorskommentarerna utgör och de interaktioner som censorerna i ett ideationellt rum deltar i (se Grahn, Holmberg & Magnusson 2014). Med ideationellt rum avser jag (se Stolt 2016) en social verksamhet som formar kontexten och ramarna för texten. Textens väg genom det ideationella rummet går via den verksamhet och de erfarenheter som skapas bland censorerna. Den sociala interaktionen tar avstamp i tolkningssammanhang ur språkliga yttranden, inte bara i vardagliga samtal utan även i mer formaliserade samtal inom olika institutionella kontexter. I anslutning till denna studie utgör studentexamen dessa institutionella ramar.

Censorsgruppen utgör en kollegial grupp eftersom den består av censorer som verkar inom samma profession och som innehar en central roll i bedömningen av studentexamensprov. I sin professionella roll kan censorerna i en institutionell kommunikationssituation trots allt föra en mer vardagsspråklig och informell dialog med varandra och vara införstådda i de övriga censorernas uttryck. (Stolt 2016)

4 Appraisalsystemet

För att kunna hantera censorskommentarerna i ett sammanhang där jag tar fasta på relationen mellan censorernas kommentarer i uppsatserna begränsar jag mig här till den del av appraisalteorin som är relevant för denna artikel. Appraisal används i studier där syftet är en analys av uttryck av värderingar och av språket i en social kontext (Martin & White 2005) och kan för denna studie ses som inte endast en teori utan även som en metod för att komma åt värderande yttranden och relationen mellan dessa.

Med hjälp av appraisal (se Martin 2000; Hunston & Thompson 2000; Martin & White 2005; Martin & Rose 2007) är det möjligt att analysera och undersöka retoriska inslag i texter samt se på relationen mellan skribent och läsare. I den här studien

använder jag mig av delsystemet ENGAGEMANG inom appraisal då jag undersöker censorernas kommentarer på studentexamensuppsatser och relationen mellan deras kommentarer. Appraisalteorin hänger starkt samman med evaluering och de attityder och känslor som skribenten i sina yttranden för fram i texten, men även hur läsaren förhåller sig till yttrandet i relation till skribenten (se t.ex. Martin & Rose 2007: 25; Martin & White 2005; Folkeryd 2006: 46). I min studie är det censorerna som innehar rollen som skribenter och den tänkta läsaren är denna skribent själv eller en annan censor. Censorskommentarerna riktar sig inte till abiturienten som skrivit studentexamensuppsatsen eftersom det inte är frågan om en formativ bedömning av uppsatserna. Däremot uttrycker censorerna i sina kommentarer värderingar om texten för att i en dialog med de övriga censorerna komma fram till en slutlig och summativ bedömning av texten.

Eftersom en analys inom appraisal formas av diskursen och kontexten utgår jag i min studie från en läsarposition inom skolkulturen. Utgående från appraisal visar skribenten, i detta fall censorn, genom sina språkval på olika retoriska stilar i en dialog samtidigt som denna medvetet eller omedvetet ger uttryck för sina värderingar. En del av beläggen i mitt material kan dubbelklassificeras och alla belägg uppstår och bör ses i sin kontext eftersom läsarpositioneringen är avgörande för analysen. Ett delfokus i denna artikel ligger på att utgående från yttranden som censorerna gör i sina bedömarkommentarer visa på deras professionalitet som bedömare.

4.1 Appraisal: ENGAGEMANG

Det delsystem inom appraisal, ENGAGEMANG, som jag teoretiskt och metodmässigt utgår från i denna studie, handlar om attitydkällor och varifrån dessa härstammar (White & Rose 2003: 48–49). Den attityd som skribenten för fram kan vara ett eget attitydyttrande, men attityderna kan också grunda sig på en annan källa: ENGAGEMANG visar på en dialog och en intertextualitet i texten. Skribenten visar också hur denna genom den egna rösten förhåller sig till övriga röster i texten (Juvonen 2007: 431). Det centrala med ENGAGEMANG är interaktionen mellan läsare och skribent. Läsaren kan beröras eller inspireras av skribentens yttranden och engageras till att interagera med skribenten genom ett eget yttrande som tar skribentens yttrande i beaktande (se Skjelten 2013: 208).

Nedan tar jag upp ett dialogiskt perspektiv på olika röster i texten. Genom att positionera sig och anpassa sig till en förväntad reaktion kan en censor ta ställning till en annan censors yttrande och därmed öppna för en diskussion eller argumentation. I sin läsning av andra censorskommentarer kan en censor **bekräfta** ett tidigare yttrande, till exempel genom *Javisst!* eller **motsätta** sig det, till exempel genom yttrandet *Absolut inte!* och på så sätt göra ett ställningstagande. Genom ENGAGEMANG kan censorerna

förneka eller förkasta ett erbjudande, men de kan även ifrågasätta, hålla med eller visa sig tveksamma till ett tidigare yttrande. (Stolt 2016: 110)

ENGAGEMANG kan uttryckas enstämigt (monogloss) eller flerstämigt (heterogloss) och det är i de flerstämmiga yttrandena som dialogen kommer till uttryck (Martin & White 2005). Ett flerstämigt ENGAGEMANG kan, om yttrandet aktivt ger rum för alternativa positioneringar och röster, delas in i en **utvidgning av dialogen** eller, om yttrandet ifrågasätter och avvärjer övriga röster, i en **begränsning av dialogen**. I en utvidgning av dialogen förväntas frågor och erbjudanden få ett svar av skribenten. (Stolt 2016: 114–117)

Genom att projicera ett yttrande kan skribenten, i det här fallet censorn, introducera alternativa (explicita och implicita) källor för bedömningen, till exempel hänvisa till bedömningskriterierna (Stolt 2016). Censorn kan antingen genom att citera eller referera skapa en relation till en annan text eller källa. Som bedömare tar censorerna ställning till abiturientens text genom att explicit peka på bedömningskriterier, poängskalor eller tidigare censorskommentarer eller implicit relatera texten till bedömningskriterierna eller sin egen erfarenhet inom bedömning.

4.1.1 ENGAGEMANG: utvidgning av dialogen

Genom en utvidgning av dialogen ger en censor utrymme för andra röster att komma till tals i texten, till exempel genom att hänvisa till provuppgiften eller till läraren: *Materialet helt i bakgrunden (ganska knappt också i häftet); Handlar inte primärt om uppgiften. Otaliga meningsbyggnadsfel; Jag skulle ge 93 (men läraren har gett 90).*¹ Materialet och häftet avser samma sak och syftar på provuppgiften som formulerats i det häfte med uppgifter och bakgrundsmaterial som abiturienten har framfört sig under provsituationen. Genom att ta med externa röster öppnar censorn för en utvidgad dialog i vilken inte enbart de egna yttrandena har möjlighet att komma till uttryck.

En externt förmedlad attityd, en så kallad **attribution**, innebär att skribenten för fram en attityd från en utomstående källa. Attributionerna kan delas in i två undergrupper: **auktoritetsbekräftelse** och **auktoritetsdistansering** (Stolt 2016: 169–172). Med auktoritetsbekräftelse hänvisar skribenten till en auktoritet, t.ex. genom att referera till denna genom att använda ord som *säger, rapporterar, tror, enligt*, för att bestyrka sanningshalten i ett påstående. Skribenten kan till exempel genom sin egen auktoritet bekräfta sanningen i den externt förmedlade attityden. Det auktoritativa yttrandet får sin kraft från traditioner, officiella instanser och accepterade sanningar vilket ger auktoritet åt en person via kunskap och social ställning. Enligt Dysthe (1996:

1 I detta avsnitt är understrykningar i exemplen mina egna för att markera det för diskussionen relevanta nyckelordet.

70) har skolan goda förutsättningar för att fungera som auktoritet då skolan som institution representerar en accepterad sanning och tradition.

Auktoritetsdistansering *Jag skulle ge 93 (men läraren har gett 90)* innebär att skribenten distanserar sig till en extern attityd och med sina egna åsikter framför en alternativ åsikt till den auktoritativa (White 2006: 17). Enligt Martin & White (2005: 113) kan skribenten förhålla sig neutralt eller ointresserat till ett auktoritativt påstående genom att till exempel använda sig av verbet *påstå*. På så sätt tar skribenten avstånd från den auktoritativa åsikten. Även om de externa yttrandena i sig själva kan verka neutrala kan kontexten spegla värderingar som i analysen är viktiga att ta i beaktande.

I en utvidgning av den egna dialogen kan skribenten genom ett **övervägande** låta andra röster komma till tals och visa att hen är öppen för andra tolkningar än den egna *Möjligen; kunde väl gå...* Yttranden av det här slaget erbjuder läsaren en möjlighet att uttrycka sin egen åsikt och sina egna värderingar. En retorisk fråga kan även i vissa fall vara ett övervägande, genom vilket skribenten för fram en åsikt samtidigt som denna ger utrymme för läsaren att reagera på yttrandet (Mikkonen 2010: 68) och exemplifieras i mitt material genom bland andra yttrandet *Kunde detta t.o.m. vara ett L?*

Ett **erbjudande** eller ett förslag, exempelvis *Föreslår 73*, är ett personligt yttrande, men till skillnad från ett påstående, ett erkännande, ett ifrågasättande eller bekräftande yttrande öppnar ett erbjudande för svar av andra röster. Med yttranden av detta slag öppnar censorerna för reaktioner eller uppbackningar av övriga censorer. Genom att jämföra abiturienterna eller abiturientens egna uppsatser med varandra öppnar censorn för en dialog kring externa deltagare, som visserligen inte själva kan komma till tals men som ligger som grund för en möjlig reaktion från övriga censorer: *Språkligt klart bättre än I.*

Exempel på **ifrågasättande** yttranden kommer väldigt tydligt fram i mitt material. Till exempel kan en censor ifrågasätta den poäng som läraren gett uppsatsen *92p. Så högt?* Å andra sidan kan hen vara samstämmig med en föregående censor samtidigt som ett eget ställningstagande tas fram genom ett ifrågasättande yttrande, till exempel *Jag är av samma åsikt. Det är en klok skribent. Men – finns det för många språkliga skönhetsfel för ett L? Föreslår 90.* Genom ett ifrågasättande yttrande utmanas ett auktoritativt yttrande med en fråga, ofta retorisk. Ett ifrågasättande yttrande kan tänkas påminna om **avståndstagande**, **förnekande** och **motsättande** yttranden genom att hänvisa till och konfrontera källan och på så sätt begränsa dialogen. Eftersom ett ifrågasättande yttrande trots allt är formulerat som en fråga, ibland även en retorisk sådan, har jag valt att se yttrandet som en möjlighet för andra röster att komma till tals (Stolt 2016: 164).

4.1.2 ENGAGEMANG: begränsning av dialogen

En **begränsning av dialogen** kan handla om att ta avstånd från ett påstående eller att **uttrycka sin egen ståndpunkt** genom att hävda sin åsikt. Ett **avståndstagande** kan handla om att **förneka** ett tidigare uttalande, (*absolut inte*), att refusera det (*Nej, detta bör nog vara ett B → 67*) eller att genom ett **motsättande yttrande** komma med ett motförslag (*inte L men E*). Genom att förneka ett yttrande kan skribenten föra in läsaren på de egna värderingarna och censorn kan på detta sätt visa en annan syn på tolkningen eller åskådliggöra en bredare kunskap eller erfarenhet på området. En censor som motsätter sig ett yttrande kan rikta sin attityd mot antingen en annan censor (vars kommentar denna läst och motsätter sig) eller en tredje part, t.ex. läraren, och den poäng läraren gett abiturienten. Ett motsättande yttrande fungerar lika som ett förnekande, då censorn visar att denna delar samma värdegrund som en annan censor men vill nyansera och förändra tänket kring yttrandet. (Martin & White 2005: 118–121; Mikkonen 2010: 68–69; Stolt 2016: 172–184)

Genom att uttrycka sin egen ståndpunkt begränsar skribenten alternativ i den pågående dialogen: *Innehållet inte illa. Språket däremot svagt. 60*. En censor kan, helt eller delvis, visa sig **samstämmig** med ett påstående, genom att **bekräfta** eller **erkänna** påståendet, och på så sätt visa att denna är av samma åsikt som en annan censor eller läraren. **Medhåll** och **uttryck för en egen ståndpunkt** är också sätt på vilka censorn kan visa ett eget ställningstagande, till exempel *Enig, 67*. (Martin & White 2005: 121–125; Stolt 2016: 181) Ett uttryck för ett eget ställningstagande är subjektivt och skiljer sig därmed från hänvisningar till andra källor. Skribenten ger sitt yttrande tyngd genom att framhäva sin egen ståndpunkt och begränsar på så sätt vissa dialogiska alternativ genom att ifrågasätta eller konfrontera ett annat yttrande (Mikkonen 2010: 70) och kan i mitt material exemplifieras av *Instämmer mycket välskrivnen men inget extra, tar upp det som alla gör, vilket är helt förståeligt sista sidan har däremot en 'ny' poäng* eller av *snarast 65*. Genom ett eget ställningstagande upplevs skribenten som auktoritativ och kompetent att värdera andras åsikter (Martin & White 2005: 127–128; Stolt 2016: 181).

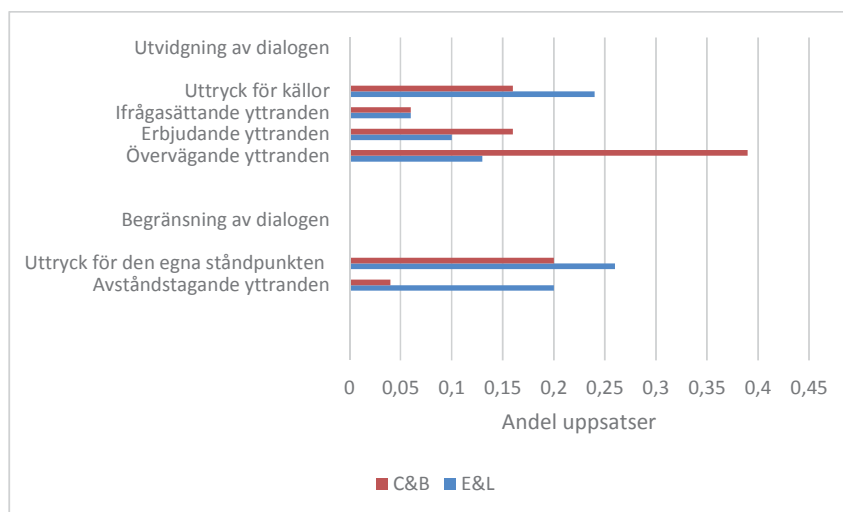
5 Typiska drag i bedömarkommentarer på studentexamensuppsatser

Fokus i denna artikel ligger på hur bedömare resonerar kring bedömningen av uppsatser som ligger på de högsta vitsordsnivåerna (L och E), och hur de resonerar kring uppsatser som bedömts med lägre vitsord (C och B). I detta avsnitt belyser jag typiska drag i bedömarkommentarer på studentexamensuppsatser på olika delar av bedömningskalan. En bedömarkommentar kan bestå av flera olika yttranden och bör ses i sitt sammanhang (Stolt 2016).

Av de 360 uppsatser som ingår i mitt material och som bedömts enligt de högsta poängen på vitsordsskalan (L och E) innehåller 159 (44 %) av uppsatserna en eller flera censorskommentarer där censorerna för en diskussion kring bedömningen. Av dessa uppsatser har 104 (65 %) kommenterats av två censorer eller flera. För de resterande uppsatserna har censorerna varit ense med läraren och gett uppsatsen samma poäng, alternativt en ny poäng utan att ge någon skriftlig kommentar.

90, dvs. 33 procent av de 272 uppsatser i mitt material som bedömts enligt de lägre poängen på vitsordsskalan (C och B) innehåller en eller flera censorskommentarer där censorerna för en diskussion kring bedömningen. Av uppsatserna som kommenterats har censorerna i 12 fall behållit det poäng som läraren föreslagit och i 6 fall höjt lärarens föreslagna poäng. På 69 av uppsatserna har censorerna sänkt poängen gentemot lärarens förslag. För de resterande uppsatserna har censorerna varit ense med läraren och gett uppsatsen samma poäng, alternativt en ny poäng utan att ge någon skriftlig kommentar. Av de 90 studentexamensuppsatser som av läraren bedömts med poäng som skulle motsvara C eller B på vitsordsskalan har 12 (13 %) av uppsatserna kommenterats av två censorer och endast en uppsats av tre censorer.

Nedan visar jag hur yttrandena i bedömarkommentarerna fördelar sig på **utvidgning av den egna dialogen** och på **begränsning av dialogen** på studentexamensuppsatserna på olika delar av vitsordsskalan. I figur 2 visar jag hur yttrandena i bedömarkommentarerna fördelar sig enligt en **utvidgning** respektive en **begränsning av dialogen** inom delsystemet ENGAGEMANG. Figuren visar hur ENGAGEMANG som resurs är betydligt vanligare i bedömarkommentarer på uppsatser i de högre vitsordsklasserna än på uppsatser bedömda med lägre vitsord. I denna vitsordskategori finns 70 yttranden som kan klassas som ENGAGEMANG och kommentarerna i de lägre vitsordsklasserna innehåller i högre grad attitydyttranden inom andra delsystem av appraisal (se Stolt 2016). På uppsatser bedömda med höga vitsord förekommer det 310 yttranden som kan klassas som ENGAGEMANG. I figur 2 nedan visar jag den procentuella fördelningen bland sådana yttranden som åskådliggör en dialog mellan bedömare.



FIGUR 2. Fördelning av yttranden inom ENGAGEMANG enligt vitsordsklasser C & B och E & L.

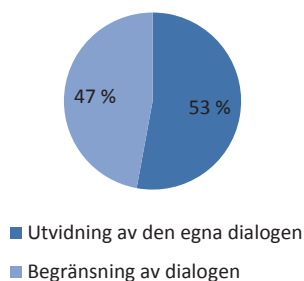
Både inom de lägre och högre vitsordsklasserna är det vanligare att censorn uttrycker en **egen ståndpunkt** än att censorn uttrycker sig avståndstagande (motsättande eller förnekande) till ett tidigare yttrande. På uppsatser bedömda med låga poäng är det märkbart vanligare att censorn ger uttryck för sin egen ståndpunkt, till exempel genom medhåll, bekräftelse eller erkännande, än att censorn yttrar sig avståndstagande till ett tidigare yttrande.

Att censorn i sina bedömarkommentarer kommer med ett **ifrågasättande yttrande** är lika vanligt vid de lägre vitsordsklasserna som vid de högre. Beträffande **övervägande yttranden** är det märkbart vanligare att censorerna uttrycker sig på detta vis på uppsatser lägre ner på vitsordsskalan än på de uppsatser som blivit bedömda med höga poäng. Genom övervägande yttranden ger censorn möjlighet för övriga röster att komma till tals och visar till exempel genom retoriska frågor, att hen är öppen för andra tolkningar än den egna eller för andra förslag till poänggivning. Censorerna visar på detta sätt mer tveksamhet till sin egen bedömning på de uppsatser som bedömts med låga poäng än på de uppsatser som bedömts med höga poäng.

Erbjudande yttranden är vanligare på uppsatser bedömda med lägre poäng än vad de är på uppsatser bedömda med högre poäng. Genom ett erbjudande yttrande föreslår censorn ofta en poäng och ger således en annan bedömare möjlighet att bekräfta eller refusera förslaget. Ett yttrande i vilket censorn hänvisar till yttre källor, så som bedömningsföreskrifter eller poängskalor, är vanligare på de högre vitsordsnivåerna än på de lägre vilket visar att censorerna i sin bedömning av goda uppsatser gärna vill relatera sin bedömning till styrdokumentet.

5.1 Dialogiska yttranden på uppsatser bedömda med höga vitsord

Drygt hälften, 53 % av yttrandena på de uppsatser som ligger högt upp på poängskalan, är uttryckta som en **utvidgning av den egna dialogen**, det vill säga att censorerna ger utrymme för andra röster att komma till tals i den egna texten (se figur 3). På de här uppsatserna förväntar sig censorerna en dialog och en diskussion med de övriga bedömarna. Ett exempel på en utvidgning av den egna dialogen kan vara ett **ifrågasättande** yttrande: *Så högt?.*



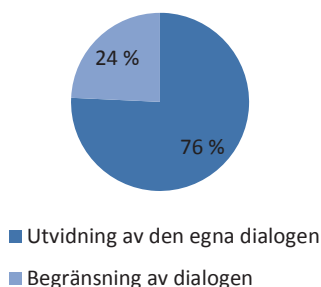
FIGUR 3. Trender i yttranden inom ENGAGEMANG på uppsatser bedömda med E eller L.

De yttranden som utgörs av en **begränsning av dialogen**, till exempel ett avståndstagande och förnekande yttrande *Absolut inte L-nivå*, står för 47 % av yttrandena på uppsatserna högt upp på poängskalan. På uppsatser som bedömts med E eller L är det nästan lika vanligt att censorerna ger plats för en dialog kring bedömningen som att de förlitar sig på sin egen värdering av uppsatsen och inte önskar en diskussion kring bedömningen.

5.2 Dialogiska yttranden på uppsatser bedömda med låga vitsord

Av de uppsatser som ligger lägre på poängskalan är 76 % av yttrandena bedömarkommentarerna uttryckta som en **utvidgning av den egna dialogen**, det vill säga att censorerna ger utrymme för andra röster att komma till tals i den egna texten (se figur 4). Ett exempel på en utvidgning av dialogen är ett **erbjudande** yttrande: *Övervägande referat och ytterst lite egen diskussion. Språket visserligen på C-nivå. Föreslår 65.* Ett annat exempel kan vara ett **överbärande** yttrande: *För många fel för 70? 67, eller Jag skulle vara benägen att ge 63 och inte bekymra mig så mycket om rubriken.*

I 24 % av yttrandena på de uppsatser som ligger lägre ner på bedömningskalan begränsar censorn dialogen, till exempel genom ett bekräftande yttrande: *Enig*, eller genom ett förnekande yttrande *Inte 70* eller ett refuserande yttrande *Nej, detta bör nog vara ett B → 67.*



FIGUR 4. Trender i yttranden inom ENGAGEMANG på uppsatser bedömda med C eller B.

På uppsatser bedömda med låga vitsord är det som figur 4 visar vanligare att censorerna genom sina yttranden öppnar för en dialog med de övriga bedömarna än att de begränsar dialogen. Detta betyder att censorerna gärna diskuterar bedömningen med övriga bedömare och vill ha en bekräftelse för att försäkra sig om en samstämmig värdering av uppsatsen. Även om uppsatserna på denna nivå mer sällan läses och kommenteras av fler än en bedömare än de uppsatser som bedömts med E & L ges det alltså större möjlighet till en dialog.

6 Sammanfattande diskussion

Min analys inom appraisal är främst kvalitativ, men jag har i den här artikeln även redogjort för en del kvantitativa resultat i syftet att kunna ge en helhetsbild av materialet. Det kan ligga lika mycket vikt i det oskrivna som det skrivna då det gäller censorernas kommentarer och det är inte möjligt för mig att greppa allt som censorerna tänker och tycker om studentexamensuppsatserna eftersom jag till exempel inte kommer åt censorernas muntliga diskussioner kring uppsatserna. Jag har gjort en analys av de nerskrivna yttranden vilka jag genom mitt material har haft tillgång till, och visat på den trend av värderande yttranden som dominerar i de nerskrivna bedömarkommentarerna på uppsatser på olika nivåer av vitsordsskalan.

Bland censorerna finns det ingen genomgående praxis att ge skriftliga kommentarer på uppsatsen och i mitt material har 44 % av uppsatserna högst upp på vitsordsskalan och 33 % av uppsatserna lägre ner på skalan fått skriftliga kommentarer. Det intressanta är att endast 13 % av uppsatserna som bedömts med C eller B kommenterats av två eller flera censorer, medan 65 % av uppsatserna högre upp på skalan, de uppsatser som legat på poäng mellan E och L, kommenterats av två eller flera censorer. Det är alltså vanligare att goda uppsatser läses och kommenteras av fler än en bedömare.

På de uppsatser som ligger lägre ner på vitsordsskalan är det vanligare att censorerna genom sina yttranden utvidgar dialogen och ger plats för andra röster att komma till tals, än att de begränsar dialogen. Även om censorn i sina bedömarkommentarer ger utrymme för svar och öppnar för en diskussion, är denna dialog mellan flera censorer väldigt snäv bland de uppsatser som bedömts med låga poäng och censorn får oftast inte det gensvar denna gett möjlighet till utan uppsatsen bedöms utgående från den första censorns poängförslag.

På uppsatser bedömda med höga poäng är det ungefär lika vanligt att censorerna i sina yttranden begränsar som utvidgar dialogen. Ett resultat av analysen visar att censorerna i en dialog med de övriga censorerna hänvisar till externa källor som standardiserade styrdokument och lärarens poänggivning, särskilt på uppsatser med höga poäng. Att konsensus i censorernas bedömning nås av ett kollegialt samspel synliggörs av den andel dialogiska yttranden som censorerna i sina bedömar kommentarer ger uttryck för och på de goda uppsatserna för censorerna ofta en diskussion kring bedömningen. Detta är anmärkningsvärt och visar att de uppsatser som ligger högst upp på vitsordsskalan sambedöms mera än uppsatser som landar på de allra lägsta poängen.

Genom att belysa bedömningen av studentexamensuppsatser tar min artikel upp en skriftkultur som inte tidigare varit föremål för forskning. Skriftkulturen inom studentexamensnämnden består av en institutionell interaktion i en social kontext och censorerna utgör en auktoritativ bedömarkår. Bedömningen av studentexamensuppsatser är summativ men för att komma till ett samförstånd kring bedömningen har censorerna möjlighet att föra en dialog kring bedömningen och de utnyttjar i många fall denna möjlighet till kollegial bedömning, särskilt på de goda uppsatserna. I den här artikeln har jag med hjälp av delsystemet ENGAGEMANG inom appraisal visat på hur flera bedömare är involverade i bedömningsprocessen och hur bedömare genom en institutionell interaktion kommer till en samstämmighet beträffande bedömningen.

Litteratur

- Borgström, E. & P. Ledin 2014. Bedömarvariation. Balansen mellan teknisk och hermeneutisk rationalitet vid bedömning av skrivprov. *Språk och stil. Tidskrift för svensk språkforskning* 24, 133–165.
- Dysthe, O. 1996. *Det flerstämmiga klassrummet. Att skriva och samtala för att lära*. Övers. Björn Nilsson. Lund: Studentlitteratur.
- Folkeryd, J. Wiksten 2006. *Writing with an attitude. Appraisal and student texts in the school subject of Swedish*. Doktorsavhandling. Acta Universitatis Upsaliensis. Studia Linguistica Upsaliensia 5. Uppsala: Institutionen för lingvistik, Uppsala universitet.

- Grahn, I.-L., P. Holmberg & U. Magnusson 2014. Systemisk-funktionell lingvistik. Att analysera språkets betydelsepotential. I: H. Lehti-Eklund, C. Lindholm & C. Sandström (red.) *Folkmålsstudier* 52. Helsingfors: Föreningen för nordisk filologi, 9–30.
- Hunston, S. & G. Thompson 2000. *Evaluation in text. Authorial stance and the construction of discourse*. New York: Oxford University Press.
- Juvonen, R. 2007. Suhtautumisen systematiikka. *Virttäjä*, 111 (3), 431–435.
- Linell, P. 1990. De institutionaliserade samtalens elementära former: om möten mellan professionella och lekmän. *Forskning om utbildning. Tidskrift för debatt och analys* 17 (4), 18–35.
- Martin, J. R. 2000. Beyond exchange: appraisal systems in English. I: S. Hunston & G. Thompson (red.) *Evaluation in text. Authorial stance and the construction of discourse*. New York: Oxford University Press, 142–175.
- Martin, J. R. & D. Rose 2007. *Working with discourse. Meaning beyond the clause*. 2 uppl. London: Continuum.
- Martin, J. R. & P. R. R. White 2005. *The language of evaluation: appraisal in English*. Palgrave Macmillan.
- Mikkonen, I. 2010. "Olen sitä mieltä, että...": Lukiolaisten yleisönosastotekstien rakenne ja argumentointi. *Jyväskylän Studies in Humanities* 135. Jyväskylä: Jyväskylän yliopisto.
- Skjelten, S. M. 2013. *Jakta på kvalitetsforskjellar i elevane sine tekstar*. Doktorsavhandling. Oslo: Det utdanningsvitenskaplige fakultet, Universitetet i Oslo.
- Stolt, S. 2016. *Vägen till vitsordet. En studie av värderande yttranden i bedömarkommentarer på studentexamensuppsatser i modersmål och litteratur*. Doktorsavhandling. Nordica Helsingiensia 46. Helsingfors: Finska, finskugriska och nordiska institutionen, Helsingfors universitet.
- Studentexamensnämnden 2016. *Poänggränser*. [online] Helsingfors: Studentexamensnämnden Tillgänglig: <https://www.ylioppilastutkinto.fi/se/studentexamen/poanggranser>. [hämtad den 6.9.2016].
- Svensson, L. 1990. Kommunikation i professioner. *Forskning om utbildning. Tidskrift för debatt och analys* 90 (4), 18–35.
- Weigle, S. C. 2011. *Assessing writing*. Cambridge: Cambridge University Press.
- White, P. R. R. 2006. Evaluative semantics and ideological positioning in journalistic discourse. I: I. Lassen (red.) *Meditating ideology in text and image. Ten critical studies*. Amsterdam: John Benjamins, 37–68.

Material

Studentexamensnämnden: Modersmålsprov II	21.3.2005
Studentexamensnämnden: Modersmålsprov I	14.3.2005
Studentexamensnämnden: Modersmålsprov II	22.3.2004
Studentexamensnämnden: Modersmålsprov I	15.3.2004