

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 pätevyyden 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 vaihteluista (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). Kieliaiheet 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 vuorovaikutukseen, 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.) Kieliaihteiden tavoitteet jäsenyivät seuraaviin osa-alueisiin: kasvu kulttuuriseen moninaisuuteen ja kielitietoisuuteen, kielenopiskelutaidot, kielitaito (vuorovaikutuksessa ja taitona tulkita ja tuottaa tekstejä). Lisäksi määritellään näihin tavoitealueisiin liittyvät sisällöt ja laaja-alaisen osaamisen teemat. Hyvää osaamista ilmentävän arvosanan 8 tarkoittama taitotaso ilmoitetaan ohjeistamaan perusopetuksen päättöarvosanan antoa. Kuvausasteikko on sovellus edellisellä opetussuunnitelmakierroksella käytetystä eurooppalaiseen viitekehukseen (EVK 2003) pohjautuneesta asteikosta (POPS 2004; Hildén & Takala 2007) ja muista viitekehysasteikoista.

Perusopetukselle rakentuvan lukiokoulutuksen opetussuunnitelman perusteissa vallitsee jonkin verran perinteisempi kielitaitokäsitys. Siellä kielitaito liittyy kiinteimmin 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 kehittymistä sekä syventää kriittistä lukutaitoa ja kielitietoisuutta. Erilaiset lukutaidot kehittyvät 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 osaitona (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 euroopalaisen 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 opiskelijoita, 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 viitekehyksen (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 etenäkään 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 arvioajakunta 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 sähköistäminen. 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 soveltuvissa tehtävissä (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. Formativisen 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 arviointeja 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 tyouralla 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 sanotuna 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 siitäkin 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 käsitteitä 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 tutkinnoissa.

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 (CETrain 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 arvioinnit 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. Batty 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. *Kokeilupersuskoulujen 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.