

Dufva, H., A. Mauranen & K. Mäntylä (toim.) 2025. 50 vuotta soveltavaa kielitiedettä: näkökulmia AFinLA ry:n toimintaan. 50 years of applied linguistics in Finland: Perspectives on AFinLA and its activities. AFinLA-teema / n:o 19, 129–156.

Ari Huhta

Jyväskylän yliopisto

Mirja Tarnanen

Jyväskylän yliopisto

Maisa Martin

Jyväskylän yliopisto

Kielitaidon arviointitutkimus ja arviointien tutkimuspohjainen kehittäminen Suomessa

Nostot

- Kielitaidon arviointi on aktiivisimpia suomalaisen soveltavan kielitieteen tutkimusalueita.
- Muuttuneet käsityksen kielestä, kielitaidosta ja oppimisesta näkyvät myös arvioinnissa.
- Arvioinnin erilaiset tarkoitukset näkyvät myös siinä, miten arviointi on Suomessa kehittynyt.
- AFinLA, kielikeskukset ja Kasvatustieteiden tutkimuslaitos vaikuttivat ratkaisevasti alan kehitykseen.
- Yleiset kielitutkinnot vauhdittivat Jyväskylän yliopiston muotoutumista keskeisimmäksi alan toimijaksi.

Abstract

The article recounts the development of language assessment in Finland since the early 1970s. The field's rise to one of the most active areas of research and development in Finnish applied linguistics can be traced back to the establishment of the Institute for Educational Research, the university-based language centre system, and the AFinLA itself. These provided both work opportunities and publication venues for language educators interested in assessment. The article first describes how changes in conceptualising language and learning have influenced assessments and then details the developments in language assessment in Finland with reference to different assessment uses. Despite being an active area of research, only a small number of individuals actually specialise in language assessment in Finland. Due to decades long work on assessment, and particularly because of the National Certificates examination, University of Jyväskylä has become the most prominent site of language assessment research in the country.

Keywords: language assessment, uses of assessments

Asiasanat: kielitaidon arviointi, arvioinnin käyttötarkoitukset

1 Johdanto

Kieleen, kielenkäyttöön ja kielitaitoon liittyvä arviointi on jokapäiväistä toimintaa. Teemme tietoisia ja tiedostamattomia päätelmiä ihmisistä heidän käyttämänsä kielen perusteella: kieli voi kertoa, mistä päin maata tai maailmaa puhuja on kotoisin ja mahdollisesti hänen yhteiskunnallisesta asemastaan. Arvioimme myös kielenkäytön laatua, esimerkiksi kuinka vakuuttavalta tietty puhe tai teksti vaikuttaa tai miten sujuvaa ja tarkkaa jonkun käyttämä kieli on. Kielitaidon arvioinnista on 1960-luvulta lähtien muodostunut oma tieteenalansa, joka katsotaan osaksi soveltavaa kielitiedettä, mutta samalla toki myös kasvatustiedettä. Emotieteidensä tavoin sillä on lukuisia yhtymäkohtia muihin tieteisiin.

Kun siirrytään arkielämän päivittäisestä, mutta epäsystemaattisesta arvioinnista järjestelmällisempään kielitaidon arviointiin, koulutus ja erityisesti kielenopetus ovat konteksteja, joissa suurin osa kielitaidon arvioinnista tapahtuu. Opetus, oppiminen¹ ja arviointi muodostavat toisiinsa liittyvän kokonaisuuden. Kasvatustieteet ovat siten tärkeä yhteistyökumppani kielitaidon arvioinnille ja lisäävät ymmärrystä esimerkiksi pedagogisesti mielekkäistä arviointitavoista. Kielitiede, psykologia ja sosiologia ovat nekin olennaisia lähitieteitä kielitaidon arvioinnille. Kielitiede auttaa määrittelemään, mitä arvioidaan esimerkiksi sanaston, rakenteiden tai pragmatiikan näkökulmista. Psykologiaa tarvitaan ymmärtämään, mitä kognitiivisia prosesseja liittyy kielen käyttöön esimerkiksi jonkin testin aikana. Sosiologian näkökulmat auttavat puolestaan tarkastelemaan arviointiin liittyviä valtasuhteita ja eettisiä kysymyksiä sekä arvioinnin vaikutuksia yhteiskuntaan. Psykometriikkaa ja tilastotiedettä tarvitaan arviointien johdonmukaisuuden ja arviointi-

1 Jotkut arviointitutkijat erottavat oppimisen ja opiskelun omiksi käsitteikseen (esim. Kansanen 2004), mutta tässä noudatamme esimerkiksi Eurooppalaisen kielten viitekehyksen käyttämää yksinkertaisempaa jaottelua, jossa oppiminen sisältää myös opiskelun.

välineiden teknisen toimivuuden selvittämiseen. Viime vuosina tietotekniikka ja tekoäly ovat nousseet tärkeään rooliin kielitaidonkin arvioinnissa.

Tässä artikkelissa kuvaamme kielitaidon arvioinnin kehitystä Suomessa osana AFinLA:n toimintaa, mutta myös osana yleisempää kansallista ja kansainvälistä soveltavan kielitieteen kehitystä. Päähuomiomme on vieraan ja toisen kielen arvioinnissa, vaikka keskeiset arvioinnin periaatteet (esim. tarkoitukset ja laatupiirteet) koskevat myös ensi- / äidinkielen arviointia. Yksi artikkelimme läpi kulkeva teema on arvioinnin erilaiset tarkoitukset: arviointi ei ole yksi yhtenäinen kokonaisuus, joka voidaan toteuttaa samalla tavalla esimerkiksi luokahuoneessa tai virallisessa kielitutkinnossa. Arvioinnin päämäärien ja arviointitietoon perustuvien päätösten hahmottaminen auttaa myös ymmärtämään muutoksia kielitaidon arvioinnissa.

Kielitaidon arviointi on sen yleisyydestä huolimatta edelleen harvinainen erityisala. Verrattuna vaikkapa kielenopetuksen, -käytön ja -oppimisen tutkijoihin, maassamme ei ole montaa arviointiin erikoistunutta tutkijaa tai kielitaidon arviointia päätyönään tekevää henkilöä. Toki löytyy sellaisia kielikoulutuksen ammattilaisia, joilla on vankka kokemus arvioinnista, vaikka heidän päätyönsä on muualla. Pyrimme artikkelissamme osoittamaan, että kielitaidon arvioinnilla on kuitenkin ollut kokoaan merkittävämpi rooli suomalaisen soveltavan kielentutkimuksen kehittämisessä ja että sillä on ollut annettavaa myös kansainväliselle alan tutkimukselle. Koska arviointi perustuu aina johonkin käsitykseen kielitaidosta ja sen oppimisesta, luomme ensin lyhyen katsauksen siihen, miten käsitykset kielestä ja sen oppimisesta ovat vaikuttaneet arviointiin ja miten muutokset näissä käsityksissä ovat muuttaneet arviointia. Tämän jälkeen kuvaamme kielitaidon arvioinnin alkuvaiheita Suomessa 1970- ja 1980-luvuilla, arvioinnin kansainvälisiä yhteyksiä sekä alan merkittävimpiä suuntauksia ja saavutuksia Suomessa.

2 Miten muutokset näkemyksissä kielestä, oppimisesta ja opetuksesta näkyvät kielitaidon arvioinnissa?

Tutkimusperustainen ja käsitteellisteoreettinen ymmärrys kielestä ja sen oppimisesta on muuttunut merkittävästi 1900-luvun jälkimmäisellä puoliskolla ja sen jälkeen. Näillä (soveltavan) kielitieteen kuten muidenkin tieteenalojen, esimerkiksi kasvatustieteiden kehityskuluilla, kuten myös toimintaympäristöjen muutoksilla, on ollut vaikutuksensa siihen, miten kielitaidon ja sen oppimisen arviointi on muuttunut vuosikymmenien aikana. Tarkastelemme seuraavassa kielitaidon arviointia kolmen käänteen kautta. Näissä käännteissä ei ole kyse vain kielitaidon arvioinnin muuttumisesta, vaan ne heijastelevat laajemmin sekä oppimisen ja sen arvioinnin, että kielitaidon arvioinnin tutkimuksessa tapahtuneita muutoksia.

1950–1960-luvulla behavioristinen lähestymistapa kielenoppimiseen ehdollistumisenä ja toiston korostamisena korosti ulkoisten ärsykkeiden ja vasteen merkitystä (Skinner 1957). Kielitaidon arvioinnissa kielioppi, sanasto ja ääntäminen arvioitiin erillisinä osaluueina. Niin kielitaidon kuin muunkin oppimisen arviointi oli summatiivista eli oppimista

arvioitiin tietyn kurssin, oppimiskokonaisuuden tai lukuvuoden lopuksi, jotta voitiin siirtyä opetuksessa eteenpäin (Nitko & Brookhart 2005; Nguyen 2021). Summatiivinen arviointi perustui tyypillisesti kokeisiin ja testeihin, joiden arviointitehtävät olivat muistamiseen ja tarkkuuteen perustuvia monivalintatyyppisiä tehtäviä. Myös monet nykyiset kansainväliset testit, kuten TOEFL ja IELTS ovat perua tältä ajalta, vaikka ne sisältävätkin nykyisin kommunikatiivisia tehtäviä. Vielä nykyäänkin länsimaisissa luokkahuoneissa yleisin arvioinnin tarkoitus lienee summatiivinen arviointi (Saeed ym. 2018), ja Suomessakin kokeet ovat olleet koko 2000-luvun tyypillinen tapa arvioida kielten osaamista (Luukka ym. 2008; Atjonen ym. 2019). Uudemmat opetussuunnitelmat korostavat formatiivisen arvioinnin merkitystä (esim. OPH 2020b), mikä toki vahvistaa sen roolia kouluissa, mutta tilanteesta ei ole tarkkaa ja yleistettävää, tutkimukseen perustuvaa tietoa.

Merkittävä muutos kielitaidon arvioinnissa tapahtui 1980-luvulta lähtien sosiokulttuurisen teorian (Vygotsky 1978), vuorovaikutuksellisen oppimisen sekä funktionaalisen kielitieteen (Halliday 1978) myötä, joista kahdessa ensimmäisessä korostui kielen oppimisen sosiaalinen ja kulttuurinen ulottuvuus ja jälkimmäisessä kielen käyttö merkityksen rakentamisen välineenä. Myös kommunikatiivinen lähestymistapa toisen kielen oppimiseen (Canale & Swain 1980) korosti kielenkäyttöä autenttisissa tilanteissa ja vuorovaikutuksen merkitystä. Myöhemmin 2000-luvun alkupuolella käyttöpohjainen kielenoppiminen (Tomasello 2005) vahvisti edelleen kielenkäytön merkitystä mutta nosti esiin erityisesti tilannekohtaisen vuorovaikutuksen. Näiden käännteiden myötä myös arvioinnissa keskityttiin yhä enemmän siihen, miten kielenoppijat pystyvät selviytymään aktiviteeteista, jotka simuloivat todellisen elämän kielenkäyttötilanteita. Esimerkiksi kirjoittamisessa siirryttiin aineityypisistä tehtävistä vähitellen todellisen elämän tekstilajeihin, kuten viesteihin.

1980–1990-luvuilla arviointitutkimuksessa tapahtui myös käänne kohti kriteeriviitteistä arviointia, jossa oppilaiden osaamista ei enää arvioitu suhteessa toisiin oppilaisiin vaan arviointikriteereihin (esim. Euroopan neuvosto 2001). Oppiminen alettiin myös ymmärtää enemmän prosessina, minkä seurauksena korostettiin formatiivisen arvioinnin ja palautteen merkitystä oppilaiden oppimisen ohjaamisessa ja tukemisessa (Black & William 2003). Myös arviointivastuun jakamiseksi ja arvioinnin vuorovaikutuksellisuuden vahvistamiseksi korostettiin itse- ja vertaisarvioinnin merkitystä. Niiden katsottiin myös kehittävän oppimista ohjaavien metataitojen, kuten itsesäätelytaitojen ja minäpystyvyyden, vahvistamista (Black & William 1998; Panadero ym. 2017). Kielitaidon arvioinnissa alettiin käyttää ja kehittää tutkimusperustaisesti muun muassa portfolioarviointia sekä erilaisia itsearviointin työkaluja (Järvinen & Kohonen 1995; Little 2005).

2000-luvulla kasvatustieteissä summatiivisen ja formatiivisen arvioinnin rinnalle nousi arviointi oppimisen tuolla puolen (*assessment beyond learning*), joka viittaa akateemisten oppimistulosten arviointia laajemmin sosiaalisiin, emotionaalisiin, kulttuurisiin ja eettisiin ulottuvuuksiin osana oppimista ja arviointia. Tämän suuntauksen mukaan arvioinnin pitäisi pystyä myös tukemaan oppijan toimijuuden ja identiteetin muodostumista, jatkuvaa oppimista ja elinikäisten taitojen kehittymistä (Biesta 2009). Teknologiakin on muuttanut arviointia 2000-luvulla muun muassa siten, että kertaluonteisista kokeista on siirrytty enemmän jatkuvaan arviointiin sekä pelkkien tietojen ja taitojen arvioinnista autenttisten tehtävien

ja laaja-alaisen osaamisen arviointiin sekä adaptiivisuuden hyödyntämiseen teknologisissa arviointiympäristöissä. Kielitaidon arvioinnissa adaptiivisuutta on hyödynnetty esimerkiksi DIALANG-arvioinnissa, joka on verkkopohjainen diagnostinen järjestelmä neljästätoista eurooppalaisessa kielessä. Erityisesti 2020-luvulla tekoälypohjaiset arviointityökalut ovat mahdollistaneet adaptiivisuuden hyödyntämistä eli oppimisen personointia oppijan osaamistason mukaisesti entisestään. Tekoälyn vaikutus näkyy myös puheen tunnistamista ja automaattista kirjoittamisen arviointia hyödyntävissä arviointiympäristöissä, joissa ihmisen arviointitoiminnan tarve vähenee. Esimerkiksi Turun yliopistossa kehitetty ViLLE-oppimisjärjestelmä tarjoaa oppilaalle ja opettajalle tietoa oppimisesta. ViLLE:n avulla opettajat voivat joko itse luoda tehtäviä tai hyödyntää muiden opettajien tekemiä tehtäviä. Tehtävät ovat pääsääntöisesti automaattisesti arvioituja ja tarjoavat opiskelijalle välittömän palautteen. ViLLE sisältää monipuolisesti tehtäviä ohjelmoinnista, matematiikasta, kielistä ja muista oppiaineista, joista opettaja voi valita opetukseen sopivat resurssit.

Tiivistäen teoreettinen ymmärrys kielestä ja sen oppimisesta on siirtynyt mekaanisista ja synnynäisistä malleista kohti vuorovaikutuksellisia, sosiokulttuurisia ja käyttöperustaisia näkemyksiä, mikä on haastanut myös käsityksiä kielitaidon arvioinnista. Yhteisöjen arviointikulttuuri eli se, millaisin käsityksin, arvostuksin ja käytäntein arviointia toteutetaan, muuttuu kuitenkin hitaasti (Xu & Brown 2016). Tämän lisäksi monikielisuuden ja kansainvälisen vuorovaikutuksen kasvavan merkityksen myötä *translanguaging*-näkökulma haastaa perinteistä yksikielistä kielenoppimista ja yksikielistä kielitaidon arviointia (García & Wei 2014).

3 Alan alkutaipaleet Suomessa

Kielitaidon arviointia on tehty yhtä kauan kuin kielenopetustakin. Suomessa vanhinta kielenopetusta edustaa Turun Akatemian ja kirkkojen tarjoama latinan opetus keskiajalta alkaen. Sittemmin opetus laajeni virkamiesten tarvitsemiin kieliin kuten ruotsiin ja myöhemmin venäjään. Koulutuksen vähittäinen laajeneminen 1800-luvun loppupuolella ja erityisesti oppikoulujen ja lukioiden perustaminen merkitsi myös ns. nykykielten eli esimerkiksi saksan ja ranskan opetuksen lisääntymistä. Tällaiseen oppilaitoksissa tehtävään kielenopetukseen kuuluu oleellisena osana sekä formatiivinen arviointi, joka tuottaa opettajalle ja oppilaalle tietoa oppimisen etenemisestä, että summatiivinen arviointi, jossa opettaja arvioi, miten hyvin oppilas on saavuttanut jonkin kokonaisuuden tavoitteet. Myös tutkintotyypillisellä arvioinnilla on maassamme pitkät juuret. Alun perin Turun Akatemian pääsykokeena toimineesta ylioppilastutkinnosta tuli lukion oppimäärään perustuva valtakunnallinen päättökoe jo vuonna 1852. Vuonna 1874 tutkintoa täydennettiin suomen ja ruotsin kokeiden lisäksi myös vieraan kielen kokeella, jossa valittavana oli saksa, ranska, venäjä tai latina. (Kaarninen & Kaarninen 2002.)

Kielitaidon arvioinnin juuret Suomessa ovat siis erilaisten oppilaitosten kielenopetuksessa, joten arvioinnin päätarkoituksetkin liittyivät kiinteästi opetuksen tarpeisiin. Ylioppilastutkinto toimi summatiivisen lukion oppimäärän päättökokeen roolinsa lisäksi

myös ns. portinvartijakokeena (Nguyen 2021), koska sitä käytettiin yliopistojen opiskelijavalinnoissa. 1960- ja 1970-lukujen taitteessa kielitaidon arviointia ryhdyttiin Jyväskylässä sijaitsevassa Kasvatustieteiden tutkimuslaitoksessa (KTL) käyttämään myös aivan uudenlaisiin tarkoituksiin, joissa päähuomio ei ollut yksittäisten oppilaiden osaamisessa vaan opetussuunnitelmien tai kokonaisten koulutusjärjestelmien tulosten mittaamisessa. Ensinnäkin Suomi osallistui vuosina 1970–71 IEA:n (*International Association for the Evaluation of Educational Achievement*) kansainväliseen oppimistulosten tutkimukseen, jossa yhtenä oppiaineena oli englannin kieli (Lewis & Massad 1975). Lisäksi KTL teki 1970-luvun alkupuoliskolla Kouluhallitukselle kansallisia tutkimuksia silloisen kokeiluperuskoulun tuottamasta osaamisesta, koska valmisteilla ollutta peruskoulu-uudistusta vastustettiin muun muassa sillä, että se heikentäisi oppimistuloksia (Kangasniemi 2019).

1970- ja 1980-luvuilla kielitaidon arvioinnin tutkimus ja kehittäminen olivat hajaantuneet eri puolelle Suomea, ja toiminta oli pääosin yksittäisten tutkijoiden varassa. Heistä ehkä paria poikkeusta lukuun ottamatta kukaan ei varsinaisesti erikoistunut arviointiin. Alan toimijat työskentelivät tyypillisesti korkeakoulujen ainelaitoksilla, erityisesti kielten laitoksilla, kieli-instituuteissa tai kielikeskuksissa ja näitä edeltäneissä korkeakoulutasoista kielenopetusta antaneissa yksiköissä. Kaksi organisaatiota toi kuitenkin arvioinnista kiinnostuneita opettajia ja tutkijoita säännöllisesti yhteen sekä erilaisten tapahtumien, kuten seminaarien ja koulutustilaisuuksien että julkaisujen merkeissä. Nämä olivat Suomen soveltavan kielitieteen yhdistys AFinLA ja kielikeskusjärjestelmä.

Sekä AFinLA että kielikeskukset järjestivät tapahtumia ja tuottivat arviointijulkaisuja. Etenkin AFinLA:lla oli merkittävä rooli kielitaidon arviointia koskevassa julkaisutoiminnassa. Jo ennen AFinLA:n julkaisusarjan perustamista yhdistyksen ensimmäinen puheenjohtaja Ola Berggren julkaisi opetusviranomaisten ja Yleisradion yhteisen kieltenopettajille suunnatun jatkokoulutuskokonaisuuden pohjalta kielenopetusta 1970-luvulla käsittelevän kirjan, johon kuului myös arviointiosuus (Berggren 1970). Monet ensimmäisistä AFinLA:n julkaisusarjan osista käsitelivät arviointia, erityisesti ylioppilastutkinnon kielikokeiden radikaalia uudistusta, jossa siirryttiin kielen tarkkuutta korostavista käännöstehtävistä nykyisen kaltaisiin, kielen käyttöä testaaviin lukemis-, kuuntelu- ja kirjoittamistehtäviin. Julkaisusta muodostui Pekka Hirvosen väitöskirja (Hirvonen 1971a, b; 1973a, b; 1974). Myös Jaakkolan (1976) varhainen AFinLA-julkaisu liittyi uusimuotoiseen ylioppilastutkintoon. Enkvistin ja Kohosen (1976) toimittamassa julkaisussa suomalaiset tutkijat tarkastelivat uudentyypistä kielitaidon arviointitapaa eli cloze-testejä, jotka olivat herättäneet runsaasti mielenkiintoa ulkomailla ja joita jotkut pitivät 1970-luvulla ehkä parhaana tapana mitata yleistä kielitaitoa (esim. Oller & Conrad 1971).

Toinen keskeinen toimijaryhmä suomalaisen kielitaidon arvioinnin alkuvaiheissa olivat kielikeskukset. Nykyinen kielikeskusjärjestelmä luotiin vuoden 1974 alussa, jolloin perustettiin Jyväskylän yliopistoon sijoitettu keskusyksikkö sekä alettiin systemaattisesti luoda paikallisia korkeakoulujen kielenopetusta antavia laitoksia (Kokkonen 1974). Keskusyksiköstä käytettiin aluksi yksinkertaista nimitystä Kielikeskus, joka myöhemmin muutettiin muotoon Korkeakoulujen kielikeskus erotukseksi yliopistojen ja korkeakoulujen omista kielikeskuksista.

Korkeakoulujen kielenopetukseen liittyviä arviointeja kehitettiin yhteistyössä Korkeakoulujen kielikeskuksen ja paikallisten kielikeskusten kanssa, joskin edellisellä oli yleensä tässä sille asetetun valtakunnallisen tehtävän mukaisesti päävastuu. Uutisia arvioinnista välitettiin opettajille Kielikeskusuutisissa sekä tutkimuspainotteisemmassa *Finlance*ssa. Julkaisutoiminnan lisäksi Korkeakoulujen kielikeskuksen toimintaan kuului alusta pitäen kielikeskusopettajien täydennyskoulutusta, myös arvioinnista (Kokkonen 1975; Lehtinen 1977). Erityisesti suullisen kielitaidon arviointi oli keskeisiä kiinnostuksen kohteita. Suullisesta kielitaidosta julkaistiin Kielikeskusuutisten erikoisnumero, jossa opettajat eri korkeakouluista raportoivat omista arviointihankkeistaan (Kohonen & Nummenmaa 1976). Osa julkaisuista käsitteli portinvartija-arviointia, esimerkiksi yliopistojen yhteisiä saksan (Klemmt 1974) ja venäjän (de Silva 1977) valintakokeita. Kielikeskusten työntekijät osallistuivat usein myös AFinLA:n toimintaan, joten oli luontevaa, että Kielikeskusuutisissa raportoitiin AFinLA:n asiantuntijaseminaarista, jossa käsiteltiin kielten oppimistulosten arviointia ja johon osallistui eri tahojen edustajia (Räsänen 1980).

1980-luvulla ja etenkin 1990-luvulla suomalainen kielitaidon arviointitutkimus näyttää selvästi kansainvälistyneen, mitä käsitellään tarkemmin artikkelimme seuraavassa luvussa. Kielikeskuksiin liittyvä arviointi ja niistä julkaiseminen jatkui kuitenkin vilkkaana. 1980-luvulla Korkeakoulujen kielikeskus aloitti monivuotisen kielitaidon arviointihankkeen, jossa tuotettiin kurssien alussa järjestettäviä tasokokeita etenkin tekstin ymmärtämiseen (Kurki-Suonio & Huhta 1987). Kehittäminen tapahtui yhdessä kielikeskusten opettajien kanssa, ja ajan mittaan toiminta laajeni myös muihin taitoihin. Tästä työstä raportoitiin lähinnä Kielikeskusuutisissa, jonka artikkelit kertoivat sekä yksittäisistä innovaatioista, kuten videoiden ja itsearviointin hyödyntämisestä suullisen kielitaidon arvioinnissa (Lönnfors 1986) tai aukkosanelusta puheen ymmärtämisen arvioinnissa (Vanderplank 1985) että laajemmista kielikeskuksista koskevista kartoituksista. Tällaisia olivat esimerkiksi kielitaidon arviointikäytänteiden selvitys (Mauranen 1988) ja suullisen kielitaidon kriteereitä ja menetelmiä koskeva kartoitus (Huhta 1988a; Huhta 1990; Huhta & Lönnfors 1990). Korkeakoulujen kielikeskus teki myös laajemman selvityksen kielitaidon arviointitutkimuksista maamme korkeakouluissa (Huhta 1988b).

Jyväskylän yliopistolla oli 1980-luvulle tultaessa yhä keskeisempi rooli kielitaidon arviointien kehittämisessä Suomessa, mikä johtui edellä mainitun Korkeakoulujen kielikeskuksen perustamisen lisäksi myös siitä, että Kasvatustieteiden tutkimuslaitos (nykyinen Koulutuksen tutkimuskeskus, KTL) oli sijoitettu Jyväskylän yliopistoon. KTL:ssä kehitettiin sekä arvioinnin teoriaa (Konttinen 1981) että käytännön arviointitutkimusta suomalaisen koululaitoksen tuottamasta osaamisesta. Tavoitteiden eroista huolimatta KTL:n toiminta lisäsi merkittävästi kansallisen tason osaamista myös kielitaidon arvioinnissa. Keskeisin KTL:n kielitaidon arviointiin erikoistunut tutkija oli Sauli Takala, jonka ura alkoi jo 1960-luvulla ja jatkui pitkälle 2010-luvulle (ks. tarkemmin muistojulkaisu hänen urastaan; Huhta ym. 2019). Takalalla oli keskeisin rooli maassamme nimenomaan kielitaidon arvioinnin tieteellisen tutkimuksen ja tutkimusperustaisen kehittämisen saralla.

Jyväskylän lisäksi merkittävää kielitaidon arviointityötä tehtiin muissakin yliopistoissa. Helsingin yliopisto oli kielikeskuksiin liittyvän arvioinnin ehkä aktiivisin kehittäjä koko

maassa. Arvioinnista julkaisivat useat kielikeskusopettajat, esimerkiksi Lönnfors (1986), Mauranen (1988, 1990), Mason (1990, 1992) ja Vanderplank (1985). Anu Virkkusen (1992) tutkimus, joka liittyi kielikeskusten lukemistesteihin, oli yksi kielitaidon arvioinnin ensimmäisiä väitöskirjoja Suomessa.

Jyväskylän ja Helsingin lisäksi Tampereen yliopisto oli etenkin 1980–1990 -luvuilla merkittävä toimija kielitaidon arvioinnin alalla. Varsinkin Viljo Kohosen työ kielisalkun käytöstä oppimisen, opetuksen ja arvioinnin välineenä oli urauurtavaa (esim. Järvinen & Kohonen 1995), vaikka hän oli aktiivinen alan toimija jo paljon tätä aikaisemmin (Kohonen & Nummenmaa 1976; Kohonen ym. 1985; Kohonen & Pitkänen 1985). Myöhemmin Tampereen yliopisto järjesti myös yhdessä Jyväskylän kanssa 1996 alan merkittävimmän kansainvälisen konferenssin *Language Testing Research Colloquiumin* (LTRC).

4 Kansainvälisyys

Nykymuotoisen kielitaidon arvioinnin katsotaan alkaneen vuonna 1961 USA:ssa, jolloin siellä julkaistiin kaksi alan varhaista perusteosta: Robert Ladon kirja *Language Testing* sekä John B. Carrollin artikkeli *Fundamental Considerations in Testing*. Nämä kansainväliset vaikutteet saavuttivat myös Suomen, kun esimerkiksi Hirvosen julkaisut AFinLA:n julkaisusarjassa 1970-luvun alussa ilmestyivät (Hirvonen 1971a, b; 1973a, b; 1974). Vaikka valtaosa vaikutteista tuli länsimaista, välillä katsottiin kuitenkin myös itään. Esimerkiksi Tilli (1977) raportoi kielitaidon arvioinnista silloisessa Neuvostoliitossa, ja 1991 Jyväskylän yliopisto järjesti kansainvälisen tutkijaseminaarin, johon kutsuttiin esiintyjä länsimaiden lisäksi myös Itä-Euroopasta (Huhta ym. 1993).

Kansainväliset, Suomessa järjestetyt seminaarit olivat 1980- ja 1990-luvuilla yksi tapa pysyä ajan hermolla kielitaidon arvioinnin uusista virtauksista. Seminaareista raportoitiin AFinLA:n teemanumeroissa: Kohonen ym. (1985) loivat katsauksen arvioinnin uusimpiin suuntauksiin, kun taas Kohonen ja Pitkänen (1985) keskittyivät kouluissa tehtäviin arviointeihin. 1990-luvun alussa Jyväskylässä järjestettiin edellä mainitun itä-länsi -tutkijaseminaarin ohella toinenkin kansainvälinen arviointitapahtuma, joka koostui luentojen lisäksi käytännön työpajoista, joissa perehdyttiin esimerkiksi testispesifikaatioiden laatimiseen (Virkkunen 1991). Myös vuosittain Jyväskylän yliopistossa järjestettyjen kielitutkimuksen kesäkoulujen esiintyjien joukossa oli kielitaidon arvioinnin parhaita asiantuntijoita. Seminaarien lisäksi ajankohtaista tietoa arvioinnista levitettiin julkaisuilla, joihin kutsuttiin kirjoittajiksi alan suomalaisia ja kansainvälisiä asiantuntijoita (Kohonen & Korpimies 1981; Korpimies 1984).

Suomalaisten tutkijoiden yhteistyö erityisesti englantilaisen Lancasterin yliopiston kanssa alkoi 1980-luvulla. Varhaisia esimerkkejä tästä on Breenin (1982) laaja erityisalojen kielenopetusta ja -arviointia käsitellyt artikkeli Kielikeskusuutisissa ja Huhdan (1986) raportti vierailusta Lancasterin arviointitutkijoiden luo. Lancasterin tutkijoita osallistui luonnollisesti edellä mainittuihin seminaareihin Suomessa 1990-luvun alussa. Yhteistyö

johti myöhemmin DIALANG-hankkeeseen ja useisiin tutkimusprojekteihin, joita kuvataan tarkemmin tuonempana.

Yksi merkittävä varhainen yhteys suomalaisen ja kansainvälisen arviointitutkimuksen välillä oli Sauli Takalan toiminta IEA:n *Study of Written Composition* -hankkeen toisena kansainvälisenä koordinaattorina 1980-luvun alkuvuosina (esim. Takala 1988). Tämä tutkimusjakso USA:ssa tuotti myös Takalan oman väitöskirjan peruskoulun oppilaiden englannin sanaston hallinnasta. Takalan väitöskirja oli paitsi yksi ensimmäisistä kielitaidon arvioinnin väitöskirjoista Suomessa (Takala 1984) myös hyvä esimerkki yhdestä arvioinnin tavallisesta käyttötarkoituksesta eli keinona hankkia tutkimustietoa osaamisesta (Nguyen 2021).

Kansainväliset tieteelliset konferenssit ovat luonnollinen osa minkä tahansa tieteenalan toimintaa. Suomalaiset kielitaidon arviointitutkijat ovat esitelleet tutkimuksiaan alan suurimmassa kansainvälisessä konferenssissa *Language Testing Research Colloquiumissa* vuodesta 1992 lähtien. Vuonna 1996 konferenssi järjestettiin Suomessa yhteistyössä Tampereen ja Jyväskylän yliopistojen kesken (Huhta ym. 1997). Jyväskylän yliopisto oli yksi eurooppalaisen kielitaidon arviointijärjestön EALTA:n (*European Association for Language Testing and Assessment*) perustajajäsenistä vuonna 2004, ja suomalaiset ovat osallistuneet järjestön toimintaan aktiivisesti. Sauli Takala toimi yhdistyksen puheenjohtajana 2000-luvulla, ja muitakin suomalaisia on ollut mukana yhdistyksen hallituksen ja työryhmien toiminnassa. EALTA:n vuotuinen konferenssi on järjestetty Suomessa kahdesti, Turussa vuonna 2009 ja Helsingissä vuonna 2023. Yleiset kielitutkinnot liittyi eurooppalaisten kielitutkintojen yhteistyöjärjestön ALTE:n (*Association of Language Testers in Europe*) jäseneksi vuonna 1996 (Leblay & Ahola 2024).

5 Keskeisiä hankkeita ja saavutuksia kielitaidon arvioinnin alalla Suomessa

Kielitaidon arviointitoiminta on ollut ehkä yllättävänkin vilkasta maassamme, kun ottaa huomioon sen, että Suomi on suhteellisen pieni sekä väkiluvun että arviointiin erikoistuneiden henkilöiden määrällä mitattuna. Suomen koulutusjärjestelmä ei myöskään ole testiorientoitunut siinä mielessä, että ylioppilastutkintoa ja joitakin äidinkielen diagnostisia testejä lukuun ottamatta kouluissa ei juurikaan käytetä standardoituja kansallisia testejä. Kielet ovat kuitenkin aina olleet tärkeä osa koulujen opetusta, koska suomen ja osin ruotsinkaan kielen osaaminen ei yleensä riitä, kun ollaan tekemisissä muiden maiden kanssa. Ehkä tämä lähtötilanne on yksi selitys monille kielitaidon arviointihankkeille Suomessa. Seuraavassa kuvaamme keskeisiä kehityskulkuja ja hankkeita Suomessa käyttäen hyväksi arvioinnin tarkoitusten jaottelua yksilöiden ja koulutusjärjestelmien arviointiin. Lisäksi mainitsemme joitakin projekteja, joissa arviointi yhdistyy kielen oppimisen tutkimukseen ja joissa arviointi on ollut sekä keino tuottaa tietoa oppimisesta että samalla kehittämisen kohde. On esimerkiksi haluttu kehittää parempia oppimista tukevia arviointitapoja.

5.1 Koulutuksen laadun ja oppimistulosten arviointi

Aloitamme kansallisista ja kansainvälisistä oppimistulosten arvioinneista, koska nämä alkoivat Suomessa jo varhain ja ne ovat edelleen oleellinen osa arvioinnin kokonaisuutta. Tällaisessa arvioinnissa ei tuoteta tietoa tai todistuksia yksittäisille oppijoille, vaan niissä selvitetään tietyn koulutusjärjestelmän tai opetussuunnitelman tuottamaa osaamista. Tutkittavasta joukosta, esimerkiksi kouluista ja niiden oppilaista otetaan sellainen otos, että arviointitulokset voidaan yleistää tilastollisesti koko joukkoon ja myös laskea, miten suuri virhemahdollisuus tuloksiin liittyy. Edellä jo todettiin, että ensimmäiset silloisen Kasvatustieteiden tutkimuslaitoksen (KTL) kansalliset kokeilupерuskoulua koskevat selvitykset ja kansainväliset IEA:n kielitaidon arviointihankkeet 1960-luvulla ja 1970-luvun alussa olivat otosperusteisia oppimistulosten arviointeja.

KTL jatkoi kansallisia arviointitutkimuksia 1970–1990 -luvuilla, esimerkiksi englantia tutkittiin vuosina 1979, 1983 ja 1991, minkä jälkeen vastuu oppimistulosten selvittämisestä siirtyi Opetushallitukselle ja myöhemmin Kansalliselle koulutuksen arviointikeskukselle (Karvi). (Tarkemmin kielten oppimistulosten arvioinneista Suomessa ks. Takala 2004; Huhta 2012.) 2000-luvulla selvitettiin äidinkielen osaamista varsin säännöllisesti, mutta vieraiden ja toisen kotimaisen kielten arviointitutkimuksissa oli paria poikkeusta lukuun ottamatta tauko, kunnes OPH ja Karvi toteuttivat useita kieliä koskevan tutkimuksen vuonna 2013 (ks. Hildén ym. 2013 yhteenveto). 2010–2020 -luvuilla kielten arviointitutkimukset ovat jatkuneet englannissa (Härmälä ym. 2019; Härmälä & Marjanen 2022), ruotsissa (Härmälä & Marjanen 2023), finskassa (Åkerlund ym. 2022, joka oli pitkittäistutkimus) ja suomessa toisena kielenä (Kuukka & Metsämuuronen 2016). Äidinkielen arvioinnit ovat myös jatkuneet (Kauppinen & Marjanen 2020), myös pitkittäistutkimuksina (Harjunen ym. 2019; Ukkola & Metsämuuronen 2023). Suomalaisen viittomakielenkin osaamista on selvitetty (Huhtanen ym. 2016). Useimmat oppimistulosten arvioinnit on suoritettu yhtä aikaa sekä suomen- että ruotsinkielisissä kouluissa, ja molempien tuloksista on raportoitu. Ruotsinkielisen peruskoulun oppimistuloksista on saatavilla myös yhteenveto (Silveström ym. 2024).

Kansainväliset kieliä koskevat oppimistulosten arvioinnit ovat olleet viime vuosikymmeninä varsin tavallisia ja näkyviä OECD:n PISA (*Programme for International Student Achievement*) ohjelman ansiosta. Niissä on tutkittu osallistujamaiden 15-vuotiaiden nuorten lukemistaitoa kunkin maan koulujen opetuskielessä (Suomessa siis suomessa tai ruotsissa riippuen koulusta). Suomi on osallistunut kaikkiin PISA-arviointeihin vuodesta 2000 lähtien. OECD:n ohella myös aiemmin mainittu IEA tekee kansainvälisiä oppimisarviointeja, kuten 4. vuosiluokan oppilaiden lukemistaitoa PIRLS (*Progress in Reading Literacy Survey*) -ohjelmassa, johon Suomi on osallistunut vuodesta 2011 lähtien. IEA on toteuttanut ajoittain myös muita yksittäisiä eri maiden äidinkieltä tai virallista kieltä koskevia tutkimuksia, joista Suomi on osallistunut ainakin 1980-luvun alun kirjoittamistaidon arviointiin (Takala 1988).

Kansainväliset oppimistulosten tutkimukset ovat olleet varsin harvinaisia vieraissa kielissä (ks. Huhdan 2012 katsaus). Kuten aiemmin todettiin, Suomi osallistui IEA:n varhaiseen englannin tutkimukseen, mutta seuraavaa merkittävää tutkimusta on saatu odottaa 2020-luvulle asti. Vihdoin Suomi on mukana PISA 2025 -tutkimuksen vapaaehtoisessa

englannin osaamista mittaavassa osassa, johon osallistuu parikymmentä maata. Tutkimus kattaa tekstin ja puheen ymmärtämisen lisäksi puhumisen, muttei kirjoittamista.

5.2 Summatiivinen arviointi

Tietyn kurssin tai oppimäärän tavoitteiden saavuttamisen arviointi on yksi yleisimmistä arvioinnin tarkoituksista kouluissa. Tyypillisesti tällaista summatiivista tietoa kerätään arvosanojen antamiseksi oppilaille, joskus myös sanallisen arvioinnin pohjaksi, etenkin perusopetuksen alkuvaiheissa. Suomessa summatiivinen arviointi on koulun ja opettajan vastuulla, mutta perustan sille antavat valtakunnalliset opetussuunnitelmat, ja perusopetuksessa erityisesti opetussuunnitelmassa määritellyt kuvaukset arvosanoille 5, 7, 8 ja 9 vuosiluokan kuusi ja yhdeksän päätteeksi tehtävässä arvioinnissa (OPH 2020a, 2023; ks. myös uudistettu opetussuunnitelman arviointiluku: OPH 2020b).

Maamme merkittävin summatiivinen arviointijärjestelmä on lukion päätteeksi järjestettävä ylioppilastutkinto, jonka juuret ulottuvat pitkälle 1800-luvulle, kuten aikaisemmin kuvattiin. Ylioppilastutkinnosta merkittävän osan muodostavat sen äidinkielen, toisen kielen, toisen kotimaisen kielen ja vieraiden kielten kokeet, joihin liittyvää tutkimusta on vuosikymmenten mittaan tehty runsaasti. Tutkinnon kielikokeiden uudistuksella oli merkittävä osuus AFinLA:n alkuaikojen toiminnassa, jolloin Hirvosen julkaisuissa esitettiin kattava pohja kielikokeiden uudistukselle. Sen jälkeen kielikokeita koskeva tutkimus voidaan jakaa kahteen päätyyppiin. Nämä ovat tutkinnon kehittämiseen liittyvä tutkimus ja tutkinnon käyttö välineenä saada tietoa oppijoiden osaamisesta kielenoppimista ja -käyttöä selvittävässä tutkimuksissa (vrt. Nguyen 2021).

Ehkä merkittävin ylioppilastutkinnon kielikokeita kehittämään pyrkivä tutkimus on kohdistunut taitoon, jota niissä ei vielä kukaan testata – puhumiseen. Jo yksi Hirvosen (1973b) julkaisuista hahmotteli, millainen ylioppilastutkinnon puhumiskoe voisi olla, ja 1990-luvulta lähtien on julkaistu useita empiirisiä ja käsitteellisiä arviointijulkaisuja, joiden tarkoitus on ollut pohjustaa tutkinnon suullista osaa. Takalan (1993) toimittama Suullinen kielitaito ja sen arviointi koostui kielitaidon teoreettista perustaa ja puhumisen arvioinnin kriteerejä ja menetelmiä käsittelevistä artikkeleista. Saleva (1997) tarkasteli väitöskirjassaan kielistudiossa nauhoitteilta kuultavien ohjeiden avulla suoritettavia englannin suullisia testitehtäviä. Hildén (2000) puolestaan tutki lukiolaisten selviytymistä erilaisissa kasvokkain suoritettavissa ruotsin suullisissa testitehtävissä. Myös puheen ymmärtämisen kokeita on tutkittu niiden toimivuuden selvittämiseksi (ranska; Anckar 2011) ja video- ja kuvamateriaalin vaikutusta testituloksiin (ruotsi; von Zansen 2019).

Ylioppilastutkinnon aineistojen avulla on tutkittu erilaisia oppijakielen ilmiöitä lukuisissa opinnäytteissä. Hyvä esimerkki tästä on Meriläisen (2010) väitöstutkimus äidinkielen siirtovaikutuksesta suomalaisten englannin kielellä kirjoittamissa teksteissä. Ylioppilastutkintoon liittyvästä tutkimusta on lisätietoa tutkintolautakunnan sivuilla.

5.3 Formatiivinen ja muu oppimista tukeva arviointi

Oppimista tukevan arvioinnin piiriin voidaan katsoa kuuluvan muitakin lähestymistapoja kuin mitä on perinteisesti ajateltu kuuluvan formatiiviseen arviointiin. Näitä ovat esimerkiksi diagnostinen (esim. Alderson 2005) ja dynaaminen (esim. Poehner 2008) arviointi, joiden kehittämiseen suomalaiset kielitaidon arvioijat ovat vaikuttaneet merkittäväksi. (ks. Mäkipää ym. 2023 oppimisesta tukevasta kielitaidon arvioinnista Suomessa.)

Oppimista tukevan arvioinnin päätarkoitus on tuottaa palautetta oppijalle, ja se on tavallisin arvioinnin tarkoitus oppilaitoksissa, koska kaikenlainen opettajan luokassa tekemä oppilaiden havainnointi luetaan tällaiseksi formatiiviseksi arvioinniksi. Tämäkin arviointi on Suomessa opettajan vastuulla, vaikka on toki tavallista, että oppilaatkin osallistuvat tähän itse- ja vertaisarvioinnin kautta (esim. Luukka ym 2008; Tarnanen & Huhta 2011). Formatiivisen arvioinnin merkitystä korostetaan valtakunnallisissa opetussuunnitelmissa (OPH 2014, 2019), mutta usein se kuitenkin jää vähemmälle huomiolle kuin summatiivinen arviointi (Atjonen ym. 2019; Hildén & Härmälä 2015) ja monet oppijat kokevat saavansa vähemmän palautetta kuin he toivoisivat (esim. Mäkipää 2021), erityisesti tytöt (Tarnanen & Huhta 2011). Merkittävä formatiivisen arvioinnin muoto on portfolio- tai salkkuarviointi, jossa oppija koostaa itselleen näytteitä suorituksistaan (esimerkiksi kirjoitelmia tai nauhoitteita omasta puheestaan), joiden avulla hän voi dokumentoida itselle ja opettajalle osaamistaan ja edistymistään. Euroopan neuvosto on luonut ns. eurooppalaisen kielisalkun käytännön välineeksi dokumentoida eri kielten osaamisen lisäksi myös muuta oppijan kielisiin liittyvää kokemusta (esim. Little 2005). Eurooppalainen kielisalkku on ollut jonkin verran käytössä Suomessakin (esim. Järvinen & Kohonen 1995), samoin sitä muistuttava portfolioarviointi (esim. Pollari 2017).

Diagnostinen arviointi on oppimista tukevan arvioinnin muoto, joka on noussut merkittäväksi kielitaidon arviointitutkimuksen alueeksi kansainvälisellä yhteistyöllä kehitetyn DIALANG-järjestelmän ansiosta. DIALANG kehitettiin suuressa EU-rahoitteisessa projektissa vuosina 1997–2004, jonka ensimmäisen vaiheen koordinaattori oli Jyväskylän yliopisto, ja jossa Lancasterin yliopisto oli keskeinen yhteistyökumppani. Järjestelmä toimi internetin välityksellä, ja siinä oli 14 testikieltä, viisi kielitaidon aluetta ja 18 käyttöliittymä- ja palautekieltä. DIALANG:n päätarkoitus oli antaa palautetta kielitaidosta ja neuvoja kielenoppimiseen. (Huhta ym. 2002; Alderson 2005.) DIALANG:lla on ollut runsaasti käyttäjiä, mutta sen tutkimuksellinen vaikutus on ollut vieläkin merkittävämpi jo järjestelmän kehittämisvaiheessa (esim. Luoma & Tarnanen 2003), etenkin koska se synnytti paljon kielen osaamisen tarkempaan diagnosointiin liittyvää tutkimusta, jota kuvataan tuonnempana. (Lukemisen ja kirjoittamisen diagnostisesta arvioinnista ks. Alderson ym. 2015 ja Huhta ym. 2024). Uudempaa oppimista tukevaa arviointia edustaa Vygotskyn (1978) sosiokulttuuriseen teoriaan pohjautuva dynaaminen arviointi, joka on levinnyt psykologiasta ja erityispedagogiikasta myös kielenopetukseen (esim. Poehner 2008). Dynaamisessa arvioinnissa pyritään selvittämään oppijan kykyä suoriutua tehtävistä sekä itsenäisesti – mikä on tavallisin tapa arvioinnissa – että myös tuetusti erilaisten vinkkien ja neuvojen avulla. Sen avulla on siis mahdollista saada laajempi kuva oppijan osaamisesta

ja ohjata opetus oppijan lähikehityksen vyöhykkeelle eli alueelle, jolla oppija suoriutuu tehtävistä vain saamansa tuen avulla. Suomessakin on tätä lähestymistapaa alettu tutkia englannin lukemisen opetuksessa ja arvioinnissa (Leontjev ym. 2024). (Laajan katsauksen dynaamisen ja diagnostisen arvioinnin nykytutkimuksesta tarjoaa Leontjev ym. 2025.)

Diagnostista kielitaidon arviointia on jo pitkään tehty ensikielessä / äidinkielessä, etenkin kun on haluttu selvittää mahdollisia kielellisiä oppimisvaikeuksia. Merkittävän suomalainen esimerkki tästä on lukemaan oppimisen ongelmia selvittänyt Jyväskylän yliopiston noin 15 vuotta kestänyt dysleksiaturkimus (esim. Lohvansuu ym. 2021), jossa kehitettiin verkkovälitteinen Ekapeli, joka diagnosoi suomeksi lukemaan oppimisen alkuvaihetta ja antaa oppijalle hänen osaamiseensa sovitettuja tehtäviä. Peliä on laajennettu moneen muuhunkin kieleen (Richardson & Lyytinen 2014) ja siitä on kehitetty uusi, perusopetuksen alkuvaiheen lukusujuvuutta ja kirjoittamista tukeva Lukukupla-peli (Nieminen ym. 2022).

5.4 Portinvartija-arviointi

Yksi arvioinnin näkyvimmistä ja tärkeimmistä tarkoituksista on kontrolloida pääsyä johonkin tärkeäksi katsottuun etuuteen tai oikeuteen (esim. Bachman & Purpura 2008). Esimerkkejä tällaisesta portinvartija-arvioinnista (engl. *gatekeeping*) ovat pääsykokeet oppilaitoksiin, johonkin ammattiin tai toimintaan pätevöittävät testit tai kansalaisuuden saamiseen liittyvät kielikokeet. Suomessakin on ollut lukuisia kielten pääsykokeita erityisesti yliopistojen kielten laitoksille, joista raportoitiin jo varsin varhain esimerkiksi Kielikeskusuutisissa (Klemmt 1974; de Silva 1977). Nytemmin ylioppilastutkinnon kielikokeet ovat valtaosin korvanneet erilliset kielten pääsykokeet korkeakoulujen opiskelijavalinnoissa, mikä tarkoittaa, että tutkinto toimii paitsi lukion oppimäärän summatiivisena loppukokeen myös portinvartijana myöhempisiin opintoihin. Ammattipätevyyden saamiseen liittyvästä kielitaidon arvioinnista on kyse, kun kääntäjä haluaa saada oikeuden toimia auktorisoituna kääntäjänä ja osallistuu Opetushallituksen järjestämään tutkintoon. Samantyyppisestä arvioinnista oli kyse myös silloin, kun henkilön täytyi osoittaa julkishallinnossa työskentelevältä vaadittava suomen tai ruotsin kielen taito, joka voitiin osoittaa Opetushallituksen ylläpitämän valtionhallinnon kielitutkinnon avulla. Tutkinto poistuu käytöstä vuoden 2026 alussa, jonka jälkeen julkishallinnossa tarvittavan kielitaidon voi osoittaa Yleisillä kielitutkinnoilla. Eri ammatteihin ja työtehtäviin liittyy usein vaatimus kielitaidosta, jonka arviointi vaihtelee huomattavasti riippuen työpaikasta ja työnantajasta. Joiltakin ryhmiltä, kuten sosiaali- ja terveydenhuollon ammattihenkilöiltä vaaditaan tiettyä suomen tai ruotsin kielen osaamistasoa, jonka voi osoittaa Yleisillä kielitutkinnoilla.

Jos ylioppilastutkinnon kielikokeita, jotka ovat myös lukion summatiivisia loppukokeita ei oteta huomioon, niin edellä mainitut Yleiset kielitutkinnot (YKI) ovat maamme merkittävien portinvartijatesti osallistujamääränsä ja monien käyttötarkoituksiensa vuoksi. YKI on näyttökoe, johon voi osallistua riippumatta siitä, miten on kielitaidon hankkinut. Tutkinto on tarkoitettu aikuisille, ja siinä on kolme tasoa (perus-, keski- ja ylin taso), joista osallistuja valitsee yhden. Tarjolla on yhdeksän kieltä: englanti, espanja, italia, pohjois-

saame, ranska, ruotsi, saksa, suomi ja venäjä. Tutkintoon osallistuu nykyään yli 10 000 henkilöä vuodessa, ja sitä käytetään tavallisimmin Suomen kansalaisuuteen vaadittavan suomen tai ruotsin kielen taidon osoittamiseen, mutta usein myös todisteena työssä tai opinnoissa tarvittavasta kielitaidosta. Osallistujamäärältään selvästi suurin on suomen keskitason tutkinto lähes 90 % osuudellaan kaikista osallistujista. Yleisten kielitutkintojen synty onkin yksi merkittävimmistä kielitaidon arvioinnin ammattimaistumista edistäneistä tapahtumista Suomessa, joten kuvaamme seuraavassa sen alkuvaiheita ja vaikutuksia (YKI:stä tarkemmin Martin ym. 1994; Leblay ym. 2014; Leblay & Ahola 2024).

Yleisten kielitutkintojen perustaminen vuonna 1994 johtui monista tekijöistä. Taustalla oli aiemmin kuvattu kielitaidon arvioinnin käytännöllinen ja teoreettinen kehittyminen etenkin Jyväskylän yliopistoon sijoitetuilla Kasvatustieteiden tutkimuslaitoksella ja Korkeakoulujen kielikeskuksessa, mutta myös yliopiston englannin ja suomen kielen laitoksilla. Jyväskylän yliopistoon oli alkanut kertyä ns. kriittistä massaa sekä kielitaidon arvioinnin että kielen oppimisen ja opettamisen tutkimuksessa, esimerkiksi sellaisella uudella alalla kuin suomi toisena kielenä. Lisäksi yliopistossa oli 1980-luvun alusta lähtien voinut opiskella soveltavaa kielitiedettä samannimisellä laitoksella. Yleisten kielitutkintojen syntyä tuki kaksi arviointihanketta, joissa kehitettyjä periaatteita ja käytännön välineitä voitiin hyödyntää YKI-tutkinnossa. Nämä hankkeet koskivat kahden uuden kielitestin, Työelämän kielidiplomin ja suomen kielen tasokokeen kehittämistä.

Työelämän kielidiplomi (TKD) oli vuosina 1989–1993 käytössä ollut kielitutkinto aluksi englannin, mutta pian myös saksan ja myöhemmin ruotsin kielessä. Kyse oli erityisalojen kielitaidon arvioinnista, koska testien sisältö sovitettiin liike-elämän, kaupan ja hallinnon alojen tarpeisiin. Tutkinto kehitettiin yhteistyössä Jyväskylän yliopiston, Markkinointi-instituutin ja Fintran kesken, ja se toimi osoituksena kielitaidosta sekä osallistujalle että hänen työnantajalleen. (Sajavaara 1992; Huhta ym. 1993.) 1990-luvun alun laman johdosta TKD jäi lyhytikäiseksi tutkinnoksi, mutta siinä suunniteltuja tehtävänlaadintaohjeita sekä puhumisen arviointiasteikkoja hyödynnettiin sittemmin YKI:ssä. Hanke vahvisti myös Jyväskylän roolia kansallisen tason kielitaidon arviointien suunnittelijana.

Työelämän kielidiplomin ohella yleisten kielitutkintojen taustalla vaikutti suomen kielen tasokoe, jota järjestettiin muutaman kerran 1980- ja 1990-luvun taitteessa. Kokeelle oli tarvetta, koska vaikka joissakin maissa suomen kielen kokeen saattoi suorittaa paikallisen ylioppilastutkinnon osana, tämä mahdollisuus oli vain harvoilla. Kokeeseen osallistuivat esimerkiksi ulkomailla asuvien suomalaisten perheiden lapset, jotka tarvitsivat todistuksen suomen kielen taidostaan opiskellakseen tai tehdäkseen työtä Suomessa. Myös Suomessa alkoi olla ihmisiä, jotka olivat oppineet suomea arkielämässään, mutta tarvitsivat taidostaan todistuksen esimerkiksi hakiessaan oppilaitoksiin. Tarjollahan oli vain ns. virkamiessuomen koe, joka ei soveltunut useimpiin muihin tarkoituksiin.

Suomen kielen tasokokeen suunnitteli ja laati työryhmä, jossa keskeisimpinä jäseninä olivat Kirsti Siitonen, Maisa Martin ja Mirja Tarnanen, jotka laativat tehtävät ja arvioivat suoritukset. Käytännön työssä oli mukana myös useita opiskelijoita. Enimmillään kokeen suoritti noin sata henkilöä yhdellä testikerralla. Arviointikriteerit pyrkivät kommunikatiivisuuden korostamiseen, ja malleja haettiin ulkomaisista kielikokeista. Kokeella ei missään

vaiheessa ollut virallista asemaa, ja se järjestettiin pääosin vapaaehtoistyönä ja pienen Opetusministeriöltä saadun rahoituksen turvin, mutta ilmeisesti se koettiin hyödylliseksi, koska suorittajia kuitenkin riitti. Kun 1990-luvun alussa alettiin suunnitella yleisiä kielitutkintoja, suomen kielen tasokoe sulautui siihen luontevasti, olihan siinä jo pilotoitu monia YKI:ssä toteutettuja käytänteitä.

Edellä kuvattu arviointiosaamisen kehittyminen 1970–80 -luvuilla loi siis perustaa Yleisille kielitutkinnoille, mutta tutkinnon syntyyn vaikutti useita toisiinsa liittyviä tavoitteita. Näistä keskeisin oli vahvistaa kommunikatiivisen kielitaidon opetusta ja arviointia erityisesti aikuisten kielikoulutuksessa (Leblay & Ahola 2024: 15). Haluttiin myös tarjota mahdollisuus saada todistus osaamisesta mahdollisimman monessa kielessä ja mahdollisimman lähellä osallistujia. Kansainvälisiä kielitutkintoja ei tietyissä kielissä järjestetty ollenkaan Suomessa ja jos järjestettiin niin mahdollisesti vain yhdellä paikkakunnalla. YKI:ä alettiinkin järjestää joka puolella maata. Monikielisen kansallisen tutkinnon tarvetta lisäsi Suomen liittyminen Euroopan unioniin 1990-luvulla, minkä arveltiin lisäävän ihmisten tarvetta osoittaa hankkimaansa kielitaitoa virallisella tutkinnolla (Leblay & Ahola 2024: 28, 152). YKI:n tavoite oli myös luoda laajempaa, kansainvälisen tason osaamista kielitaidon arvioinnissa, mitä tällaisen tutkinnon kehittäminen omin voimin tukisi.

Mainituista lähtökohdista luotiinkin Suomen oloissa varsin uudentyyppinen kielitaidon arviointijärjestelmä, jossa yhdistyivät virallisuus (YKI perustuu lakiin ja asetuksiin), monikielisyys, opetussuunnitelmiin sitoutumattomuus (YKI on näyttötutkinto) ja aikuiset kohderyhmänä. Tutkinnon luomista ryhdyttiin suunnittelemaan vuonna 1992 yhteistyössä Opetushallituksen ja Jyväskylän yliopiston kesken ja ensimmäiset tutkinnot järjestettiin 1994 seitsemässä kielessä (Leblay ym 2014: 9). Alkuvuosina englanti oli tutkinnon suosituin kieli, mutta vuonna 2003 voimaan tulleen kansalaisuuden kielitaitovaatimusta tarkentaneen lain jälkeen suomen kielen tutkinto on noussut osallistujamäärältään ylivoimaisesti suurimmaksi kieleksi (Leblay & Ahola 2024: 56–57).

Yleisiä kielitutkintoja voidaan hyvällä syyllä kutsua tutkimusperusteisiksi tutkinnoiksi, koska niiden laatua on alusta pitäen pyritty varmistamaan tutkimuksen keinoin (esim. Halvari & Tarnanen 1997). Hyviä esimerkkejä tästä ovat useat YKI-tutkinnon tutkijoiden väitöskirjat. On tutkittu kirjoittamisen (Tarnanen 2002) ja puhumisen (Ahola 2022) arviointia sekä arvioijien arviointilinjan mahdollisia muutoksia ajan myötä (Neittaanmäki & Lamprianou 2024a, b). Myös YKI:n tehtävien laatijoiden toimintaa on selvitetty (Ahola 2012). Luoman (2001) väitöskirja testien käsitevalidoinnista käsitteli puolestaan arvioinnin yleisempiä teoreettisia kysymyksiä, jotka koskevat myös YKI:n kaltaisia tutkintoja. Yleisten kielitutkintojen arvioijien toimintaa on tarkasteltu myös laajemmin Rikkinäistä suomea -hankkeessa. Siinä selvitettiin vaikuttaako S2-oppijoiden ensikielen tunnistaminen ja näihin ensikieliin liittyvät stereotyyppiset käsitykset niiden puhujista arviointiin – eli antaako arvioija suorituksesta matalamman tai korkeamman arvion kuin se ansaitsisi, koska hän tunnistaa puhujan kielitaustan (Halonen ym. 2020; Halonen ym. 2024).

Portinvartijatutkinnoilla ja muilla isoilla (kieli)testeillä on monia yhtymäkohtia, jotka johtuvat siitä, että niihin osallistuu suuria määriä kielienoppijoita, joiden osaamista täytyy mitata tarpeeksi luotettavasti ja kattavasti, mutta niin, että se on käytettävissä olevien

resurssien puitteissa mahdollista (esim. Bachman & Purpura 2008). AFinLA:n teemanumero 9 (Huhta & Hildén 2016) käsittelee tällaisiin laajoihin arviointeihin liittyviä kysymyksiä ja tarjoaa kattauksen uudempaa suomalaista tutkimusta isoista kielitesteistä, jotka edustavat sekä portinvartiointia (YKI; inkerinsuomalaisten paluumuuttajien kielitesti), summatiivista arviointia (ylioppilastutkinto) ja oppimistulosten kansallisia ja kansainvälisiä arviointeja (Karvin arvioinnit; PISA-tutkimukset).

5.5 Kielitaidon arviointi tutkimusvälineenä

Arvioinnilla on vielä yksi käyttötarkoitus, joka liittyy kiinteästi tutkimukseen, eli arvioinnin käyttö keinona saada tietoa esimerkiksi kielenoppimisesta (Nguyen 2021). Monet opinnäytteet mutta myös laajemmat tutkimushankkeet käyttävät erilaisia kielitestejä tai muita arviointitapoja mukaan lukien itsearviointia selvittääkseen oppijoiden osaamista ja osaamisen muutoksia. Tässä ei ole mahdollista luetella edes kaikkia suurimpia hankkeita, mutta edellä mainittu diagnostisen arvioinnin nousu merkittäväksi kielitaidon arvioinnin alueeksi merkitsi myös lisääntyneitä kiinnostusta selvittää, millaista kieltä ja osaamista Eurooppalaisen viitekehysten taitotasot oikein edustavat ja miten osaaminen konkreettisesti, esimerkiksi rakenteiden ja sanaston suhteen muuttuu taitotasolta toiselle. Tämä synnytti laajempaa eurooppalaista tutkimusyhteistyötä (Bartning ym. 2010), josta osa tehtiin Suomessa etenkin toisella ja vieraalla kielellä kirjoittamisen edistymistä tutkineissa CEFLING ja TOPLING -hankkeissa (hankkeissa tehtyjä tutkimuksia, ks. esim. Alanen ym. 2010; Martin ym. 2010; Mustonen 2015) sekä lukemiseen ja kirjoittamiseen keskittyneessä Dialuki-hankkeessa, jota Jyväskylän yliopisto teki tiiviissä yhteistyössä Lancasterin yliopiston kanssa (esim. Alderson ym. 2015; Alderson ym. 2016). Uudempaa kielenoppimiseen liittyvää tutkimusta, jossa osaamisen arviointi on keskeinen osa toimintaa, edustavat muun muassa Turun yliopiston puheen sujuvuuden tutkimushankkeet (esim. Lehtilä ym. 2024) sekä useamman yliopiston yhteishankkeet DigiTala (von Zansen & Kallio 2024) ja AASIS. DigiTalassa kehitettiin suomi ja ruotsi toisena kielenä -puhumisen automaattista analysointia ja AASIS-hankkeessa puolestaan interaktiivisen suomi toisena kielenä -puheen automaattista arviointia (Ullakonoja ym., 2025).

6 Eurooppalainen kielten viitekehys kriteeriviitteisen arvioinnin tukena

Kuten aiemmin tässä artikkelissa kuvattiin, yksi muutoksista kielitaidon arvioinnissa on ollut siirtyminen pois oppilaiden vertailuun nojaavasta normiperusteisesta arvioinnista kriteeriperusteiseen arviointiin, josta käytetään myös nimitystä kriteeriviitteinen arviointi. Vertailuun perustuva arviointi nimittäin vääristää tuloksia, kun vertailupohjana käytetään pieniä ryhmiä, kuten yksittäisiä luokkia, koska eritasoisten oppilaiden jakautuminen eri aikoina ja eri ryhmissä on sattumanvaraista. Oppijalle annettava arvosana voi tällöin riip-

pua muiden oppilaiden osaamisesta jopa enemmän kuin hänen omasta osaamisestaan. Kriteereihin perustuvassa arvioinnissa oppijan osaamista arvioidaan nimensä mukaisesti suhteessa johonkin kriteeriin, joka voi olla esimerkiksi määritelmä tavoitteena olevasta osaamisesta. (Takala 1980b; Brown 2012; Huhta & Takala 1999.)

Tunnetuin esimerkki kriteeriviitteisestä arvioinnista on, kun puhumis- tai kirjoitus-suorituksia arvioidaan arviointiasteikolla, jonka tasot on määritelty sanallisesti. Lukiossa tällaista arviointia on alettu tehdä viimeistään silloin, kun uusimuotoinen kielten ylioppilaskoe kirjoitustehtävineen otettiin käyttöön 1970-luvulla (Hirvonen 1973a). Kriteeriviitteinen kielitaidon arviointi oli esillä myös uudessa peruskoulussa, vaikka vaatimustasojen määrittely koettiin vaikeaksi opetussuunnitelmia laadittaessa (Takala 1980a, b). Voidaan sanoa, että vasta Eurooppalainen kielten viitekehys (EVK) (Euroopan neuvosto 2001) merkitsi kriteeriviitteisen arvioinnin laajempaa läpimurtoa. Vasta 2000-luvun alkuvuosina perusasteen ja lukion valtakunnallisissa opetussuunnitelmissa otettiin käyttöön taitotasot keinona määrittellä eri kielten oppimäärien tavoitteita (OPH 2003, 2004). Opetussuunnitelmien taitotasoasteikko perustuu EVK:n kuusiportaiseen asteikkoon, mutta sitä muokattiin sopivammaksi nuorille oppijoille ja sen portaiden määrää lisättiin tasolta toiselle etenemisen seuraamisen helpottamiseksi (Hildén & Takala 2007).

EVK:n taitotasoa hyödynnetään Suomessa laajasti. Perusasteen ja lukion lisäksi niitä tai niistä muokattuja asteikkoja käytetään ammattikorkeakouluissa (Juurakko-Paavola 2005), yliopistoissa ja maahanmuuttajien kotoutumiskoulutuksessa (OPH 2022). Myös useat suomalaiset kielitutkinnot käyttävät EVK-tasoa tai niihin rinnasteisia asteikkoja raportoidessaan osallistujien tuloksia. Näin tekevät Yleiset kielitutkinnot (Leblay & Ahola 2024) ja valtionhallinnon kielitutkinnot (OM 2002). Myös kansalliset kielten oppimistulosten tutkimukset ovat 2010-luvulta lähtien raportoineet tulokset EVK-perusteisia perusasteen tasokuvauksia käyttäen. Näin tekee myös vuoden 2025 PISA-tutkimus, jonka englantia koskevaan osaan Suomi osallistuu. (ks. tarkemmin: Huhta 2020, taitotasojen hyödyntämisestä opetuksessa ja arvioinnissa.)

7 Ajatuksia tulevaisuudesta

Kielitaidon arvioinnin tulevia suuntia maassamme voidaan hahmotella sekä yleisellä tasolla että eritellymmin suhteessa joihinkin arvioinnin keskeisimmistä tarkoituksista, joita kuvattiin tarkemmin edellä. Ehkä näkyvin muutos, joka on jo alkanut, on teknologian hyödyntäminen arviointiprosessin eri vaiheissa, mutta erityisesti arvioinnin suorittamisen välineenä. Tästä hyvä esimerkki on ylioppilastutkinnon digitalisaatio 2010-luvun lopulla. Teknologia ja erityisesti 2020-luvulla tapahtunut tekoälyn nopea esiinmarssi vaikuttaa sekä arviointivälineiden laatimiseen että arviointien toteutukseen ja tulosten ja palautteen tuottamiseen. Vaikutukset eroavat hiukan riippuen arvioinnin tarkoituksesta, mutta yleisesti voidaan sanoa, että teknologia avaa uusia mahdollisuuksia formatiivistyyppiseen arviointiin, joka tukisi entistä paremmin sekä oppimista että opetusta. Kuvaamme tätä hiukan tarkemmin seuraavassa.

Oppimisen arvioinnissa on tulevaisuudessa keskeistä edelleenkin siirtyminen kohti monipuolisempia, oppijalähtöisempiä ja teknologiaa hyödyntäviä arviointitapoja, sillä arviointikulttuurien muutos on jokseenkin hidasta (Xu & Brown 2016). Erityisesti teknologiset ratkaisut voisivat tukea arvioinnin muuttumista yhä enemmän formatiiviseksi eli oppimista ohjaavaksi ja tukevaksi. Tämän lisäksi sekä kieltenopetuksessa että arvioinnissa kommunikatiivisuus (Canale & Swain 1980) ja käyttöpohjaisuus (Tomasello 2005) voisivat, suotuisasta kehityksestään huolimatta, näkyä käytännössä nykyistä enemmän. Arvioinnissa tämä tarkoittaisi esimerkiksi simulaatioiden, roolipelien tai virtuaaliympäristöjen hyödyntämistä. Tämän tyyppinen arviointi edellyttäisi kuitenkin muutosta myös käsityksissä siitä, millaista arviointia pidetään luotettavana (esim. Moss 1994).

Säännöllisen palautteen antaminen perinteisesti voidaan kokea työläänä opettajien keskuudessa, mutta teknologiset ratkaisut, kuten tekoälypohjaiset arviointiympäristöt voisivat mahdollistaa adaptiiviset eli oppijan taidoille sopivat tehtävät, yksilöllisen palautteen antamisen ja oppimisanalytiikan hyödyntämisen oppimista tukevalla tavalla myös isoissa opetusryhmissä. Edistysaskeleet automaattisessa puheentunnistuksessa ja siihen perustuvassa arvioinnissa ja palautteen annossa lisäävät myös suullista kielitaitoa tukevaa arviointia tulevaisuudessa (Evanini & Zechner 2019). Tällaista kehitystyötä on alettu tehdä myös Suomessa (ks. DigiTala ja AASIS -hankkeet; Al-Ghezi ym. 2023; von Zansen & Kallio 2024; Ullakonoja ym. 2025).

Oppimaan oppimisen näkökulmasta metataitojen kehittämisen tarvetta ei voi kuitenkaan ulkoistaa teknologialle. Toisin sanoen oppijoiden oman osaamisen reflektointitaitojen omaksuminen ja oppimistavoitteiden asettaminen on keskeistä tulevaisuudessakin, kuten myös vertaisarvioinnin nykyistä laajempi hyödyntäminen oppijoiden omien arviointitaitojen ja kriittisen ajattelun kehittämiseksi. Monikielisen pedagogiikan ja sen myötä myös monikielisen arvioinnin kehittäminen on niin ikään tulevaisuudessa keskeistä. Kielitaidon arviointi sivuaa läheisesti kulttuuriseen moninaisuuteen liittyvien kompetenssien ja yhteistyötaitojen arvioimista. Näin ollen käsityksen kielitaidosta pitäisi niin ikään muuttua moninaisemmaksi ja holistisemmaksi. (Huhta ym. 2023.)

Oppilaitoksissa tehtävän arvioinnin ohella myös muihin tarkoituksiin käytetyt arvioinnit ovat muutoksessa. Edellä kuvatut edistysaskeleet teknologian ja tekoälyn hyödyntämisessä merkitsevät myös sitä, että tärkeissä summatiivisissa ja portinvartija-arvioinneissa tulee mahdolliseksi arvioida osallistujien puhumista ja kirjoittamista automaattisesti myös suomalaisissa kielitutkinnoissa. Tähän kuuluu kuitenkin vielä jonkin aikaa, sillä arviointialgoritmien kehittäminen vaatii paljon koulutettujen arvioijien arvioimia suorituksia, joiden avulla algoritmit ohjataan toimimaan mahdollisimman samalla tavalla kuin ihmisarvioijat. Suomen ja ruotsin kaltaisten suhteellisten pienten kielten kohdalla tällaista aineistoa on vain rajallisesti käytettävissä (Al-Ghezi ym. 2023). Suunta on kuitenkin selvä, ja todennäköisesti 2030-luvulla nähdään ensimmäiset käyttökelpoiset suomen ja ruotsin kielistä oppijakieltä automaattisesti arvioivat järjestelmät kielitutkinnoissa (tekoälyn soveltamisesta kielitaidon arvioinnissa Suomessa, ks. *Studies in Second Language Assessment* -lehden teemanumero; Huhta & Leontjev 2025). Koska kehitys on pidemmällä englannissa ja muissa laajemmin käytetyissä kielissä (esim. Evanini & Zechner 2019), näiden kielten automaattinen arviointi

voi yleistyä suomalaisissa kielitesteissä jo tätä ennen, esimerkiksi jos jokin tutkinto hankkii lisenssin ulkomailla kehitettyyn arviointijärjestelmään.

Toinen kielitutkintoja eli portinvartiointiin käytettyjä kielitestejä koskeva muutos koskee maahanmuuttoon liittyvää arviointia. Jo yli 20 vuotta Suomen kansalaisuuteen on vaadittu EVK:n tasoa B1 vastaava osaaminen suomessa tai ruotsissa, mikä tyypillisesti osoitetaan YKI-tutkinnon avulla. Vuoden 2026 alusta on tulossa voimaan uusi laki, jonka nojalla myös pysyvää oleskelulupaa varten vaadittaisiin tietty kielitaidon taso, joka täytyy osoittaa virallisella kielitutkinnolla (Hallituksen esitys 62/2025). Lisäksi on mahdollista, että työmarkkinatukeen liitetään kielilisiä, jonka saamiseksi täytyy osoittaa riittävä suomen tai ruotsin taito (kielitaidon ja työllistymisen yhteyksistä ks. myös Tarnanen & Pöyhönen 2011). Nämä muutokset merkitsevät varsin suurta lisäystä tämäntyyppiseen kielitaidon arviointiin maassamme.

Kielitaidon arviointi on ollut Suomessa varsin pieni, mutta kansainvälisestikin aktiivinen tieteenala, johon vain harvat ovat erikoistuneet. Johtuen tietyistä koulutuspoliittisista ratkaisuista, erityisesti Kasvatustieteiden tutkimuslaitoksen ja Korkeakoulujen kielikeskuksen perustamisesta Jyväskylän yliopiston yhteyteen, Jyväskylästä kehittyi kielitaidon arvioinnin tutkimusperusteisen kehittämisen keskeisin toimija Suomessa. Kielitaidon arviointi oli vielä 1970–80 -luvulla hajaantuneempaa eri puolille maata, mutta Yleisten kielitutkintojen perustaminen vuonna 1994 loi tarpeen luoda arviointiin erikoistunut tutkijaryhmä. Nykyään YKI:ssä työskentelee kymmenen kielitaidon arviointiin erikoistunutta tutkijaa ja tilastotieteilijää. Vaikka valtaosa maamme kielitaidon arvioinnin ammattilaisista on keskittynyt yhteen yliopistoon, niin alan asiantuntemusta löytyy muualtakin, vaikka tyypillisesti muualla työskentelevät tekevät päätyönään muuta kieliin ja soveltavaan kielentutkimukseen liittyvää tutkimusta ja opetusta. On kuitenkin nähtävissä, että yleisemmät muutokset kielikoulutuksessa ovat laajentamassa kielitaidon arvioinninkin kenttää yliopistojen välisen yhteistyön kautta. Etenkin teknologian hyödyntäminen arvioinnissa vaatii yhteistyötä kielitaidon arvioijien, kieliteknologioiden ja tekoälyn asiantuntijoiden välillä (esim. edellä mainitut Helsingin, Jyväskylän ja Aalto yliopiston DigiTala ja AASIS -hankkeet sekä Jyväskylän ja Helsingin DD-LANG -hanke; ks. tarkemmin Leontjev ym. 2024). Vastaavanlainen yhteistyö eri alojen tutkijoiden kesken yleistyneenä entisestään tulevaisuudessa.

Suomen soveltavan kielitieteen yhdistys AFinLA on edelleen merkittävä kielitaidon arviointitutkimuksen mahdollistaja Suomessa, vaikka alan suomalaiset tutkijat käyvät luonnollisesti myös kansainvälisissä konferensseissa ja raportoivat työstään monilla kansallisilla ja kansainvälisillä julkaisukanavilla. AFinLA:n vuotuinen syysposium ja sen vuosikirja sekä teemanumerot (esim. Huhta & Hildén 2016; Mäkipää ym. 2023) ovat edelleen tärkeitä paikkoja ja tapoja arviointitutkijoille verkostoitua ja kertoa työstään muille kielikoulutuksen ammattilaisille.

Kirjallisuus

- Ahola, S. 2012. *Yleisten kielitutkintojen laatijoiden käsityksiä kielestä ja tehtävien laadinnasta*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:NBN:fi:ju-201301241110>
- Ahola, S. 2022. *Rimaa hipoen selviää tilanteesta: yleisten kielitutkintojen suomen kielen arvioijien käsityksiä kielitaidon arvioinnista ja suullisesta kielitaidosta*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-9005-3>
- Alanen, R., A. Huhta & M. Tarnanen 2010. Designing and assessing L2 writing tasks across CEFR proficiency levels. Teoksessa I. Bartning, M. Martin & I. Vedder (toim.) *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. EUROSLA Monograph Series, 1, 21–56. <http://eurosla.org/monographs/EM01/EM01home.html>
- Alderson, J.C. 2005. *Diagnosing foreign language proficiency. The interface between learning and assessment*. Lontoo: Continuum.
- Alderson, J.C., E.-L. Haapakangas, A. Huhta, L. Nieminen & R. Ullakonoja 2015. *The diagnosis of reading in a second or foreign language*. Lontoo: Routledge.
- Alderson, J.C., A. Huhta & L. Nieminen 2016. Characteristics of weak and strong readers in a foreign language. *The Modern Language Journal*, 100 (4), 853–879. <https://doi.org/10.1111/modl.12367>
- Al-Ghezi, R., K. Voskoboinik, Y. Getman, A. von Zansen, H. Kallio, M. Kurimo, A. Huhta & R. Hildén 2023. Automatic speaking assessment of spontaneous L2 Finnish and Swedish. *Language Assessment Quarterly*, 20 (4–5), 421–444. <https://doi.org/10.1080/15434303.2023.2292265>
- Ankar, J. 2011. *Assessing foreign language listening comprehension by means of the multiple-choice format: processes and products*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-4410-0>
- Atjonen, P., H. Laivamaa, A. Levonen, S. Orell, M. Saari, K. Sulonen, M. Tamm, P. Kamppi, N. Rumpu, N. Hietala & J. Immonen 2019. "ETTÄ TIETÄÄ MISSÄ ON MENOSSA" *Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa*. Julkaisut 7: 2019. Helsinki: Kansallinen koulutuksen arviointikeskus https://www.oph.fi/sites/default/files/documents/karvi_07191.pdf
- Auktorisoidun kääntäjän tutkinto. Helsinki: Opetushallitus. <https://www.oph.fi/fi/palvelut/auktorisoidun-kaantajan-tutkinto>
- Bachman, L. & J. Purpura 2008. Language assessments: Gate-keepers or door-openers? Teoksessa B. Spolsky & F. Hult (toim.) *Handbook of Educational Linguistics*. Malden, MA: Blackwell, 456–468.
- Bartning, I., M. Martin & I. Vedder (toim.) *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. EUROSLA Monograph Series, 1. <http://eurosla.org/monographs/EM01/EM01home.html>
- Berggren, O. 1970. *Kieltenopetus 1970-luvun koulussa*. Porvoo: WSOY.
- Biesta, G. 2009. Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education)*, 21 (1), 33–46.
- Black, P., & D. Wiliam 1998. Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5 (1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Black, P. & D. Wiliam 2003. 'In praise of educational research': Formative assessment. *British educational research journal*, 29 (5), 623–637. <https://www.jstor.org/stable/1502114>
- Breen, M. 1982. Ten years of ESP courses at Lancaster – what have we learned? *Kielikeskusutisia*, 4/1982, 3–17. <https://urn.fi/URN:NBN:fi:ju-2022090883001>
- Brown, J.D. 2012. *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Canale, M. & M. Swain 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1), 1–47. <https://doi.org/10.1093/applin/1.1.1>
- Carroll, J.B. 1961. Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English proficiency of foreign students*. Washington, D.C: Center for Applied Linguistics, 30–40.

- Enkvist, N. & V. Kohonen (toim.) 1976. *Integratiivisen kielitaidon mittaamisesta cloze-testien avulla: teoriaa ja sovelluksia*. Soveltavan kielitieteen tutkimuksia 13. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4057>
- Euroopan neuvosto 2001. *Common European Framework for Languages: Learning, teaching, assessment*. Strasbourg: Language Policy Unit, Council of Europe. <https://rm.coe.int/1680459f97>
- Evanini, K. & K. Zechner (toim.) 2019. *Automated speaking assessment. Using language technologies to score spontaneous speech*. New York: Routledge.
- García, O. & L. Wei 2014. *Translanguaging: Language, bilingualism and education*. Lontoo: Palgrave Macmillan.
- Halliday, M. 1978. Ideas about language. *Arts: The Journal of the Sydney University Arts Association*, 11, 20–38. <https://openjournals.library.sydney.edu.au/ART/article/view/5494>
- Hallituksen esitys 62/2025. *Hallituksen esitys eduskunnalle laiksi ulkomaalaislain muuttamisesta*. <https://www.finlex.fi/fi/hallituksen-esitykset/2025/62>
- Halonen, M., S. Ahola, T. Hirvelä, R. Neittaanmäki, S. Ohranen & R. Ullakonoja 2024. Kielitaidon arviointi kansalaisuuden portinvartijana. Teoksessa T. Renvik & M. Säävälä (toim.) *Kotoutumisen kokonaiskatsaus 2023: näkökulmana väestösuhteet*. Työ- ja elinkeinoministeriö. TEM oppaat ja muut julkaisut, 2024:1, 77–86. <http://urn.fi/URN:ISBN:978-952-327-703-8>
- Halonen, M., A. Huhta, S. Ahola, T. Hirvelä, R. Neittaanmäki, S. Ohranen & R. Ullakonoja 2020. Ensikielen tunnistamisen merkityksestä suullisen kielitaidon arvioinnissa Yleisissä kielitutkinnoissa. Teoksessa S. Grasz, T. Keisanen, F. Oloff, M. Rauniomaa, I. Rautiainen & M. Siromaa (toim.) *Menetelmällisiä käännteitä soveltavassa kielitutkimuksessa – Methodological turns in applied language studies*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 17. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 56–70. <https://doi.org/10.30661/afinlavk.89453>
- Halvari, A. & M. Tarnanen 1997. Qualitative procedures for comparability – Experiences from a multi-language, multi-level testing system. Teoksessa A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (toim.) *Current developments and alternatives in language assessment. Proceedings of LTRC 96*. Jyväskylä: Jyväskylä yliopisto, 127–136.
- Harjunen, E., J. Marjanen & J. Karlsson 2019. *Äidinkielen pieni pitkittäisarviointi 2014–2017: Perusopetuksen päätöstä lukion päättöön*. Helsinki: Kansallinen koulutuksen arviointikeskus. https://www.karvi.fi/sites/default/files/sites/default/files/documents/KARVI_0419.pdf
- Hildén, R. 2000. *Att tala bra, bättre och bäst: Suomenkielisten abiturienttien ruotsin kielen taito testisuoritusten valossa*. Helsinki: Helsingin yliopisto. <http://hdl.handle.net/10138/229501>
- Hildén, R. & M. Härmälä 2015. *Hyvästä paremmaksi – kehittämisideoita kielten oppimistulosten arviointien osoittamiin haasteisiin*. Helsinki: Kansallinen koulutuksen arviointikeskus & Opetushallitus. <https://www.oph.fi/fi/tilastot-ja-julkaisut/julkaisut/hyvasta-paremmaksi-kehittamisideoita-kielten-oppimistulosten>
- Hildén, R., M. Härmälä, J. Rautopuro, M. Huhtanen, M. Puukko & C. Silverström 2013. *Kielten oppimistulokset perusopetuksen päättövaiheessa 2013*. Helsinki: Kansallinen koulutuksen arviointikeskus & Opetushallitus. <https://www.karvi.fi/fi/julkaisut/kielten-oppimistulokset-9-vuosiluokalla-2013>
- Hildén, R. & S. Takala 2007. Relating descriptors of the Finnish school scale to the CEF overall scales for communicative activities. Teoksessa A. Koskensalo, J. Smeds, P. Kaikkonen & V. Kohonen (toim.) *Foreign languages and multicultural perspectives in the European context – Fremdsprachen und multikulturelle Perspektiven im europäischen Kontext*. Dichtung – Wahrheit – Sprache, Band 9–10. Berliini: Lit Verlag, 291–300.
- Hirvonen, P. 1971a. *Englannin kielen taidon mittaaminen lukion päättyessä: 1. Lähtökohdat*. Soveltavan kielitieteen julkaisuja 1. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4045>
- Hirvonen, P. 1971b. *Englannin kielen taidon mittaaminen lukion päättyessä: 2. Kuuntelukoe*. Soveltavan kielitieteen julkaisuja 4. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/3996>

- Hirvonen, P. 1973a. *Englannin kielen taidon mittaaminen lukion päättyessä: 3. Lukukoe ja Kirjoituskoe*. Soveltavan kielitieteen julkaisuja 7. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4050>
- Hirvonen, P. 1973b. *Englannin kielen taidon mittaaminen lukion päättyessä: 4. Puhumiskoe*. Soveltavan kielitieteen julkaisuja 8. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4051>
- Hirvonen, P. 1974 *Englannin kielen taidon mittaaminen lukion päättyessä: 5. Kielitaitokoe*. Soveltavan kielitieteen julkaisuja 9. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4052>
- Huhta, A. 1986. Testaus: asenteita ja uudistuksia. *Kielikeskusuutisia*, 6/1986, 3–8. <https://urn.fi/URN:NBN:fi:ju-2022090883072>
- Huhta, A. 1988a. Kielitaidon mittaamistutkimusta Suomessa. *Kielikeskusuutisia*, 3/1988, 14–20. <https://urn.fi/URN:NBN:fi:ju-2022090883088>
- Huhta, A. 1988b. Outline of English language oral skills testing in the language centres. *Kielikeskusuutisia*, 10/1988, 11–19. <https://urn.fi/URN:NBN:fi:ju-2022090883097>
- Huhta, A. 1990. Oral assessment criteria used at the Finnish Language Centres. *Kielikeskusuutisia*, 5/1990, 3–13. <https://urn.fi/URN:NBN:fi:ju-2022090883108>
- Huhta, A. 2012. Kansalliset ja kansainväliset kielitaidon arviointitutkimukset – millaista tietoa niissä tuotetaan? *Kieli, koulutus ja yhteiskunta*, 3 (5). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-lokakuu-2012/kansalliset-ja-kansainvaliset-kielitaidon-arviointitutkimukset-millaista-tietoa-niissa-tuotetaan>
- Huhta, A. 2020. Taitotasojen hyödyntämisestä opetuksessa ja arvioinnissa. Teoksessa M. Härmälä & M. Puukko (toim.) *Taitotasaja ja suullista kielitaitoa. Näkökulmia englannin oppimistulosten arvioinnin osoittamiin kehittämiskohteisiin*. Helsinki: Kansallinen koulutuksen arviointikeskus, 25–47. https://www.karvi.fi/sites/default/files/sites/default/files/documents/KARVI_Artikkelikokoelma_1_2020.pdf
- Huhta, A., G. Erickson & N. Figueras (toim.) 2019. *Developments in language education: A memorial volume in honour of Sauli Takala*. Jyväskylä: Jyväskylän yliopisto & European Association for Language Testing and Assessment. <https://www.ealta.eu.org/documents/resources/Developments%20in%20Language%20Education%20A%20Memorial%20Volume%20in%20Honour%20of%20Sauli%20Takala%20-%20June%202026.pdf>
- Huhta, A., C. Harsch, D. Leontjev & L. Nieminen 2024. *The diagnosis of writing in a second or foreign language*. Lontoo: Routledge.
- Huhta, A. & R. Hildén (toim.) 2016. *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 9. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinla/issue/view/4267>
- Huhta, A., V. Kohonen, L. Kurki-Suonio & S. Luoma 1997. *Current developments and alternatives in language assessment. Proceedings of LTRC 96*. Jyväskylä: University of Jyväskylä.
- Huhta, A. & D. Leontjev (toim.) 2025. Artificial intelligence in language assessment in Finland. *Studies in Second Language Assessment (Special Issue)*, 14 (2). <https://arts.unimelb.edu.au/language-testing-research-centre/research/publications/sila>
- Huhta, A., D. Leontjev, M. Honko & R. Hildén 2023. Kielitaidon arviointi koulussa Teoksessa A. Pitkänen-Huhta, K. Mård-Miettinen & T. Nikula (toim.) *Kielikoulutus mukana muutoksessa. Language education engaging in change*. AFinLA teema. Soveltavan kielitieteen tutkimuksia 16. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 242–265. <https://journal.fi/afinla/article/view/126733>
- Huhta, A., S. Luoma, M. Oscarson, K. Sajavaara, S. Takala & A. Teasdale 2002. DIALANG – A diagnostic language assessment system for learners. Teoksessa J.C. Alderson (toim.) *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Case studies*. Strasbourg: Council of Europe, 130–145. <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168069f403>

- Huhta, A. & P. Lönnfors 1990. Oral skills testing: Practices and principles. *Kielikeskusuutisia*, 2/1990, 2–8. <https://urn.fi/URN:NBN:fi:juu-2022090883105>
- Huhta, A., A. Määttä, S. Takala. & M. Tarnanen (toim.) 1997. *Yleiset kielitutkinnot ja kieltenopetus*. Helsinki: Opetushallitus.
- 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*. Jyväskylä: Koulutuksen tutkimuslaitos, Jyväskylän yliopisto, 136–159.
- Huhta, A. & S. Takala 1999. Kielitaidon arviointi. Teoksessa K. Sajavaara & A. Piirainen-Marsh (toim.) *Kielenoppimisen kysymyksiä*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto, 179–228.
- Huhtanen, M., M. Puukko, P. Rainö, N. Sivunen & R. Vivolin-Karén 2016. *Viittomakielen oppimistulokset perusopetuksen 7.–9. vuosiluokilla 2015*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/viittomakielen-oppimistulokset-perusopetuksen-7-9-vuosiluokilla-2015>
- Härmälä, M., M. Huhtanen, M. Puukko & J. Marjanen 2019. *A-englannin oppimistulokset 7. vuosiluokan alussa 2018*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/englannin-oppimistulokset-7-vuosiluokan-alussa-2018>
- Härmälä, M. & J. Marjanen 2022. *Englantia koronapandemian aikaan. A-englannin osaaminen 9. luokan lopussa keväällä 2021*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/englantia-koronapandemian-aikaan-englannin-osaaminen-9-luokan-lopussa-kevaalla-2021>
- Härmälä, M. & J. Marjanen 2023. *A-ruotsin oppimistulokset perusopetuksen päättövaiheessa 2022*. Helsinki: Kansallinen koulutuksen arviointikeskus. https://www.karvi.fi/sites/default/files/sites/default/files/documents/Karvi_2423.pdf
- Jaakkola, H. 1976. *Ylioppilaskokelaiden englannin kielen tutkinnon aineosioiden lauserakenteesta*. Soveltavan kielitieteen julkaisuja 15. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4058>
- Juurakko-Paavola, T. (toim.) 2005. *Ammattikorkeakouluopiskelijoiden kielitaito ja sen arviointikäytännöt*. Hämeenlinna: Hämeen ammattikorkeakoulu.
- Järvinen, A. & V. Kohonen 1995. Promoting professional development in higher education through portfolio assessment. *Assessment & Evaluation in Higher Education*, 20 (1), 25–36. <https://doi-org.ezproxy.jyu.fi/10.1080/0260293950200104>
- Kaarninen, M. & P. Kaarninen 2002. *Sivistyksen portti: Ylioppilastutkinnon historia*. Helsinki: Otava.
- Kangasniemi, E. 2019. Nuoresta opiskelijasta alansa asiantuntijaksi. Teoksessa A. Huhta, G. Erickson & N. Figueras (toim.) *Developments in language education: A memorial volume in honour of Sauli Takala*. Jyväskylä: Jyväskylän yliopisto & European Association for Language Testing and Assessment, 152–157. <https://www.ealta.eu.org/documents/resources/Developments%20in%20Language%20Education%20A%20Memorial%20Volume%20in%20Honour%20of%20Sauli%20Takala%20-%20June%202026.pdf>
- Kansanen, P. 2004. *Opetuksen käsitemaailma*. Jyväskylä: PS-kustannus.
- Kauppinen, M. & J. Marjanen 2020. *Millaista on yhdeksäsluokkalaisten kielellinen osaaminen? – Suomen kielen ja kirjallisuuden oppimistulokset perusopetuksen päättövaiheessa 2019*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/millaista-yhdeksäsluokkalaisten-kielellinen-osaaminen-suomen-kielen-ja-kirjallisuuden-oppimistulokset-perusopetuksen-paattovaiheessa-2019>
- Klemmt, R. 1974. Saksan valintakoeprojekti Jyv-Tam 74. *Kielikeskusuutisia*, 4/1974, 4–9. <https://urn.fi/URN:NBN:fi:juu-2022090882980>
- Kohonen, V., H. von Essen & C. Klein-Braley (toim.) 1985. *Practice and problems in language testing* 8. Soveltavan kielitieteen julkaisuja 40. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA.

- Kohonen, V. & L. Korpimies (toim.) 1981. *Finlance I*. Jyväskylä: Korkeakoulujen kielikeskus, Jyväskylän yliopisto. <https://journal.fi/afinlavk/issue/view/4088>
- Kohonen, V. & L. Nummenmaa 1976. Special issue on teaching and testing communicative competence. *Kielikeskusuutisia*, 4/1976. <https://urn.fi/URN:NBN:fi:ju-2022090883026>
- Kohonen, V. & A. Pitkänen (toim.) 1985. *Language testing in school*. Soveltavan kielitieteen julkaisuja 41. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinlavk/issue/view/4087>
- Kokkonen, H. 1974. Näkemyksiä Kielikeskuksen identiteetistä, asemasta ja toiminnasta korkeakoulutasoisessa kielitaidon opetuksessa. *Kielikeskusuutisia* 1/1974, 4–6. https://jyx.jyu.fi/bitstream/handle/123456789/82977/KKU_74-1.pdf
- Kokkonen, H. 1975. Kielitaidon testauskurssi Jyväskylässä 19.–21.5.1975. *Kielikeskusuutisia*, 4/1975, 8. <https://urn.fi/URN:NBN:fi:ju-2022090882985>
- Konttinen, R. 1981. *Testiteoria*. Helsinki: Gaudeamus.
- Korpimies, L. (toim.) 1984. *Finlance III*. Jyväskylä: Korkeakoulujen kielikeskus, Jyväskylän yliopisto.
- Kurki-Suonio, L. & A. Huhta 1987. Testaussuunnitelmiä. *Kielikeskusuutisia*, 9/1987, 33–36. <https://urn.fi/URN:NBN:fi:ju-2022090883084>
- Kuukka, K. & J. Metsämuuronen 2016. *Suomi toisena kielenä (S2) -oppimäärä oppimistulokset perusopetuksen 9. vuosiluokalla 2015*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/suomi-toisena-kielena-s2-oppimaaara-oppimistulokset-perusopetuksen-9-vuosiluokalla-2015>
- Lado, R. 1961. *Language testing: The construction and use of foreign language tests: A teacher's book*. Bristol: Longmans, Green and Company.
- Language Testing Research Colloquium (LTRC). <https://www.iltaonline.com/general/custom.asp?page=LTRCPrograms>
- Leblay, T. & S. Ahola 2024. *Yleiset kielitutkinnot 30 vuotta*. Helsinki: Opetushallitus. <https://www.oph.fi/sites/default/files/documents/Yleiset%20kielitutkinnot%2030%20vuotta.pdf>
- Leblay, T., T. Lammervo & M. Tarnanen (toim.) 2014. *Yleiset kielitutkinnot 20 vuotta*. Helsinki: Opetushallitus. <https://www.oph.fi/sites/default/files/documents/leblay-t.-lammervo-t.-ja-tarnanen-m.-toim.-yleiset-kielitutkinnot-20-vuotta.-2014.pdf>
- Lehtilä, E., P. Peltonen & P. Lintunen 2024. Listener perceptions of L1 speech in L2 and L3 fluency assessment contexts: a qualitative approach. Teoksessa M. Toivola, P. Lintunen & L. Heikkola (toim.) *Puheen tutkimuksen uusia suuntia – Aineistona vapaasti tuotettu puhe. New directions in speech research – Freely produced speech as data*. Soveltavan kielitieteen tutkimuksia 17. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 191–215. <https://journal.fi/afinla/article/view/136100>
- Lehtinen, M. 1977. Suullisen kielitaidon mittaamisen kurssi. *Kielikeskusuutisia* 4/1977, 8–10. <https://jyx.jyu.fi/handle/123456789/83031>
- Leontjev, D., A. Huhta & K. Koskela 2024. "I could learn more words": Involving stakeholders in defining L2 English reading constructs in diagnostic and dynamic assessment. Teoksessa M. Kivilehto, L. Lahti, T. Pitkänen, E. Pitkäsalo & M. Tervola (toim.) *Tutkimuksellisia siltoja rakentamassa*. AFinLAN vuosikirja 2024. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 186–210. <https://journal.fi/afinlavk/article/view/143435>
- Leontjev, D., M.E. Poehner & A. Huhta (toim.) 2025. *Dynamic and diagnostic language assessment: Learning across frameworks to support L2 education*. Berliini: De Gruyter Mouton.
- Lewis, E. & C. Massad 1975. *The teaching of English as a foreign language in ten countries*. Tukholma: Almqvist & Wiksell International.
- Little, D. 2005. The Common European Framework and the European Language Portfolio: Involving learners and their judgements in the assessment process. *Language Testing*, 22 (3), 321–336. <https://doi.org/10.1191/0265532205lt311oa>

- Lohvansuu, K., M. Torppa, T. Ahonen, K. Eklund, J. Hämäläinen, P. Leppänen & H. Lyytinen 2021. Unveiling the mysteries of dyslexia: Lessons learned from the prospective Jyväskylä Longitudinal Study of Dyslexia. *Brain Sciences*, 11 (4), Article 427. <https://doi.org/10.3390/brainsci11040427>
- Luoma, S. 2001. *What does your test measure? Construct definition in language test development and validation*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:951-39-0897-6>
- Luoma, S. & M. Tarnanen 2003. Creating a self-rating instrument for second language writing: from idea to implementation. *Language Testing*, 20 (4), 440–465. <https://doi.org/10.1191/0265532203lt267>
- Luukka, M.-R., S. Pöyhönen, A. Huhta, P. Taalas, M. Tarnanen & A. Keränen 2008. *Maailma muuttu- mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-4214-4>
- Lönnfors, P. 1986. Active use of video to develop oral production and self-evaluation. *Kielikeskusuutisia*, 1/1986, 1–13. <https://urn.fi/URN:NBN:fi:jyu-2022090883067>
- 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.
- Martin, M., S. Mustonen, N. Reiman & M. Seilonen 2010. On becoming an independent user. Teoksessa I. Bartning, M. Martin & I. Vedder (toim.) *Communicative proficiency and linguistic development, intersections between SLA and language testing research*. EUROSLA Monograph Series 1. European Second Language Association, 57–80. <https://eurosla.org/monographs/EM01/EM01tot.pdf>
- Mason, D. 1990. The seventh oral skills workshop. *Kielikeskusuutisia*, 1/1990, 8–10. <https://urn.fi/URN:NBN:fi:jyu-2022090883104>
- Mason, D. 1992. Developing a new test – is it worth the effort? *Kielikeskusuutisia*, 3/1992, 1–12.
- Mauranen, A. 1988. Kielitaidon mittaamisesta kielikeskuksissa. *Kielikeskusuutisia*, 3/1998, 9–13. <https://urn.fi/URN:NBN:fi:jyu-2022090883088>
- Mauranen, A. 1990. Kielitaidon mittaamisesta kielikeskuksissa. *Finlance VIII*, 1–16. https://jyx.jyu.fi/jyx/Record/jyx_123456789_84469
- Meriläinen, L. 2010. *Language transfer in the written English of Finnish students*. Joensuu: Itä-Suomen yliopisto. <http://urn.fi/URN:ISBN:978-952-61-0230-6>
- Moss, P. 1994. Can there be validity without reliability? *Educational Researcher* 23 (2), 5–12. <https://www.jstor.org/stable/1176218>
- Mustonen, S. 2015. *Käytössä kehittyvä kieli: paikat ja tilat suomi toisena kielenä -oppijoiden teksteissä*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-6213-5>
- Mäkipää, T. 2021. *"I have never received feedback from my teachers." Students' perceptions of assessment and feedback practices in Finnish general upper secondary education*. Helsinki: Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-951-51-7644-8>
- Mäkipää, T., R. Hildén & A. Huhta (toim.) 2023. *Kielenoppimista tukeva arviointi. Assessment for supporting language learning*. AFinLA-e. Soveltavan kielitieteen tutkimuksia 15. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA. <https://journal.fi/afinla/issue/view/9173>
- Neittaanmäki, R. & I. Lamprianou 2024a. Communal factors in rater severity and consistency over time in high-stakes oral assessment. *Language Testing*, 41 (3), 584–605. <https://doi.org/10.1177/02655322241239363>
- Neittaanmäki, R. & I. Lamprianou 2024b. All types of experience are equal, but some are more equal: The effect of different types of experience on rater severity and rater consistency. *Language Testing*, 41 (3), 606–626. <https://doi.org/10.1177/02655322241239362>

- Nieminen, L., U. Richardson, L. Aoia, M. Halttunen, & N. Väkeväinen 2022. Opettajat innoissaan taitoihin mukautuvasta Lukukupla-pelistä. *Luokanopettaja*, 4/2022, 36–37. <https://issuu.com/luokanopettajat>
- Nitko, A. & S. Brookhart 2005. *Education assessment of students*. New Jersey: Pearson Education Ltd.
- Nguyen, P. 2021. Uses of language assessments. Teoksessa C. Chapelle (toim.) *The Encyclopedia of Applied Linguistics*. Hoboken, NJ: Wiley. <https://doi.org/10.1002/9781405198431.wbeal1237.pub2>
- Oller, J. & C. Conrad 1971. The cloze technique and ESL proficiency. *Language Learning*, 21 (2), 183–194. <https://doi.org/10.1111/j.1467-1770.1971.tb00057.x>
- OM 2002. *Valtionhallinnon kielitutkintojärjestelmän uudistaminen*. Helsinki: Opetusministeriö. <http://julkaisut.valtioneuvosto.fi/handle/10024/80843>
- OPH 2003. *Lukion opetussuunnitelman perusteet 2003*. Helsinki: Opetushallitus. https://www.oph.fi/sites/default/files/documents/47345_lukion_opetussuunnitelman_perusteet_2003.pdf
- OPH 2004. *Perusopetuksen opetussuunnitelman perusteet 2004*. Helsinki: Opetushallitus.
- OPH 2014. *Perusopetuksen opetussuunnitelman perusteet 2014*. Helsinki: Opetushallitus. https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf
- OPH 2019. *Lukion opetussuunnitelman perusteet 2019*. Helsinki: Opetushallitus. <https://eperusteet.opintopolku.fi/eperusteet-service/api/dokumentit/9593357>
- OPH 2020a. *Perusopetuksen päättöarvioinnin kriteerit*. Helsinki: Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20p%C3%A4%C3%A4tt%C3%B6arvioinnin%20kriteerit%2031.12.2020_1.pdf
- OPH 2020b. *Oppilaan oppimisen ja osaamisen arviointi perusopetuksessa. Perusopetuksen opetussuunnitelman perusteiden 2014 muutokset*. Helsinki: Opetushallitus. https://www.oph.fi/sites/default/files/documents/perusopetuksen-arviointiluku-10-2-2020_1.pdf
- OPH 2022. *Kotoutumiskoulutuksen opetussuunnitelman perusteet 2022*. Helsinki: Opetushallitus. https://www.oph.fi/sites/default/files/documents/Kotoutumiskoulutuksen_opetussuunnitelman_perusteet_2022_2.pdf
- OPH 2023. *Perusopetuksen arviointikriteerit kuudennen vuosiluokan päätteeksi*. Helsinki: Opetushallitus. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20arvosanakriteerit%20kuudennen%20vuosiluokan%20p%C3%A4%C3%A4tteeksi_0.pdf
- Panadero, E., A. Jonsson & J. Botella 2017. Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22, 74–98. <https://doi.org/10.1016/j.edurev.2017.08.004>
- Poehner, M. 2008. *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. New York: Springer.
- Pollari, P. 2017. *(Dis)empowering assessment? Assessment as experienced by students in their upper secondary school EFL studies*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-7178-6>
- Richardson, U. & H. Lyytinen 2014. The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, 10 (1), 39–60. <https://ht.csr-pub.eu/index.php/ht/article/view/153>
- Räsänen, A. 1980. Raportti AFinLAN asiantuntijaseminaarista: koulujen kielenopetuksen tuloksellisuus ja tulosten mittaaminen. *Kielikeskuutisia*, 4/1980, 6–7. <https://urn.fi/URN:NBN:fi:juu-2022090883048>
- Saeed, M., H. Tahir & I. Latif 2018. Teachers' perceptions about classroom assessment techniques in elementary and secondary schools. *Bulletin of Education and Research*, 40 (1), 115–130. https://pu.edu.pk/images/journal/ier/PDF-FILES/10_40_1_18.pdf
- Sajavaara, K. 1992. Designing tests to match the needs of the workplace. Teoksessa E. Shohamy, A. Walton & C. Morfitt (toim.) *Language assessment for feedback: Testing and other strategies*. Dubuque, IO: Kendall/Hunt Publishing Company, 122–144.

- Saleva, M. 1997. *Now they're talking: Testing oral proficiency in a language laboratory*. Jyväskylä: Jyväskylän yliopisto. https://jyx.jyu.fi/jyx/Record/jyx_123456789_72490
- de Silva, V. 1977. Venäjän kielen yhteinen valintakoe 05.07.1976. *Kielikeskusuutisia* 1/1977, 3–7. <https://urn.fi/URN:NBN:fi:jyu-2022090883028>
- Silveström, C. J. Hellgren, J. Marjanen & J. Metsämuuronen 2024. *Ruotsinkielinen peruskoulu – yleiskatsaus tuloksiin*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/ruotsinkielinen-peruskoulu-yleiskatsaus-tuloksiin>
- Skinner, B. F. 1957. *Verbal behavior*. New York: Appleton-Century-Crofts.
- Takala, S. 1980a. *Vaatimustasojen määrittelemisen opetussuunnitelmia laadittaessa*. Selosteita ja tiedotteita 145. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Takala, S. 1980b. *Kriteerimittamisen käsitteestä ja käytännön sovelluksista*. Selosteita ja tiedotteita 146. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Takala, S. 1984. *Evaluation of students' knowledge of English vocabulary in the Finnish comprehensive school*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Takala, S. 1988. Origins of the International Study of Writing. Teoksessa T. Gorman, A. Purves & R. E. Degenhart (toim.) *The IEA Study of Written Composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press, 3–14.
- Takala, S. (toim.) 1993. *Suullinen kielitaito ja sen arviointi*. Julkaisusarja B: Teoriaa ja käytäntöä 77. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Takala, S. 2004. Englannin kielitaidon tasosta Suomessa. Teoksessa K. Leimu (toim.) *Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa*. Jyväskylä: Koulutuksen tutkimuslaitos, Jyväskylän yliopisto, 255–275.
- Tarnanen, M. 2002. *Arvioija valokeilassa: suomi toisena kielenä -kirjoittamisen arviointia*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-3419-4>
- Tarnanen, M. & A. Huhta 2011. Foreign language assessment practices in the comprehensive school in Finland. Teoksessa D. Tsagari & I. Csepes (toim.) *Classroom-based language assessment*. Frankfurt: Peter Lang, 129–146.
- Tarnanen, M. & S. Pöyhönen 2011. Maahanmuuttajien suomen kielen taidon riittävyys ja työllistymisen mahdollisuudet. *Puhe ja kieli*, 4, 139–152. <https://journal.fi/pk/article/view/4750>
- Tilli, J. 1977. Neuvostoliittolaisten pedagogien käsityksiä oppimistulosten arvioinnista kielen opetuksessa. *Kielikeskusuutisia*, 3/1977, 9–12. <https://urn.fi/URN:NBN:fi:jyu-2022090883030>
- Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Ukkola, A. & J. Metsämuuronen 2023. *Matematiikan ja äidinkielen taidot alkuopetuksen aikana – Perusopetuksen oppimistulosten pitkäjäsenarviointi 2018–2020*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/julkaisut/matematiikan-ja-aidinkielen-aidot-alkuopetuksen-aikana-perusopetuksen-oppimistulosten-pitkittaisarviointi-2018-2020>
- Ullakonoja, R., I. Lähteenmäki, N. Raud, N. Phan, T. Grósz, H. Suuronen, R. Hildén, M. Kurimo, M. Kuronen, A. von Zansen & M. Kautonen (2025) Building the foundations for automatic assessment of verbal and nonverbal aspects of spoken interaction in Finnish as a second language. *Studies in Second Language Assessment (Special Issue)*, 14 (2). <https://arts.unimelb.edu.au/language-testing-research-centre/research/publications/sila>
- Vanderplank, R. 1985. Insights into listening: Partial dictation in placement testing. *Kielikeskusuutisia*, 10/1985, 3–18. <https://urn.fi/URN:NBN:fi:jyu-2022090883066>
- VILLE oppimisympäristö. Turun yliopisto: oppimisanalytiikan tutkimusinstituutti. <https://www.oppimisanalytiikka.fi/ville/>
- Virkkunen, A. 1991. Jyväskylä teki sen taas! *Kielikeskusuutisia*, 10/1991, 2–4. <https://jyx.jyu.fi/handle/123456789/83509>
- Virkkunen, A. 1992. *Affixation as part of an English reading proficiency test at the University of Helsinki*. Jyväskylä: Korkeakoulujen kielikeskus, Jyväskylän yliopisto.

- Vygotsky, L. S. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Xu, Y., & G. Brown 2016. Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>
- von Zansen, A. 2019. *Uudenlaista kuullun ymmärtämistä – Kuvan ja videon merkitys ylioppilastutkinnon kielikokeissa*. Jyväskylä: Jyväskylän yliopisto. <https://urn.fi/URN:ISBN:978-951-39-7961-4>
- von Zansen, A. & H. Kallio 2024. *DigiTala – Moodle-sovellus suullisen kielitaidon automaattiseen arviointiin*. AFinLA-e. Teoksessa M. Kivilehto, L. Lahti, T. Pitkänen, E. Pitkäsalo & M. Tervola (toim.) *Tutkimuksellisia siltoja rakentamassa*. AFinLAn vuosikirja 2024. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AFinLA, 91–116. <https://journal.fi/afinla/article/view/131465>
- von Zansen, A., M. Sneek & R. Hildén (arvioitavana). Lukiolaisten käsitykset ja heidän antamansa palaute suullisen kielitaidon automaattisesta arvioinnista. Teoksessa R. Kantelinen, M. Kautonen & Z. Elgundi (toim.) *Kielipedagogisia näkökulmia elinikäiseen kielenoppimiseen* (tulossa). Suomen ainedidaktinen tutkimusseura.
- Åkerlund, C., J. Marjanen & E. Peltola 2022. *A-finska i årskurs 9 – Utvärdering av lärresultat våren 2021*. Helsinki: Kansallinen koulutuksen arviointikeskus. <https://www.karvi.fi/fi/node/78>