

# Kielitaidon arvioinnin metodologisia vaihtoehtoja

## Johdanto

Kielitaidon arviointia on harjoitettu jossakin muodossa niin kauan kuin ihmiset ovat kommunikoineet keskenään kielen avulla joko äidinkielellään tai jollakin muulla kielellä. Vanhimpia muistiin kirjattuja esimerkkejä on Vanhassa Testamentissa (Tuomarien kirja 12, 5-6) esiintyvä yksinkertainen "ääntämistesti", jossa *shibboleth* sanan ääntämisen perusteella eroteltiin ystävät vihollisista - varsin kohtalokkain seurauksin niille, joiden ääntäminen ei ollut hyväksyttävää. Uudemman ajalla kielitestejä alettiin käyttää yhtenä osana yliopistojen pääsyvaatimuksia mm. Englannissa. Kielitaito (nimenomaan äidinkielen) on eri muodoissa ollut olennainen osa 1900-luvun alkupuolella yleistynyttä älykkyys- ja muuta psykologista testaamista. Kielitaito on ollut jo pitkään osa myös suomalaista ylioppilastutkintoa.

Vieraan kielen taidon arviointi tutkimuksen kohteena ja omana tieteenalanaan on varsin uutta toimintaa. Kuten monella muullakin alalla, eniten tutkimusta on tehty ja julkaistu USA:ssa, jo 1960-luvulta lähtien. Modernin kielitaidon arvioinnin 'isinä' pidetyt Robert Lado ja John B. Carroll julkaisivat molemmat ensimmäiset vaikutusvaltaiset teoksensa 1961 USA:ssa (Lado 1961; Carroll 1961). Nyttemmin kenttä on laajentunut, ja vieraan kielen arviointitutkimusta tehdään paljon myös Euroopassa ja erityisesti viime aikoina myös Itä-Aasian maissa. Tärkeitä virstanpylväitä alan kehityksessä ovat sen tieteellisten järjestöjen perustaminen. ILTA (International Language Testing Association, [www.iltaonline.com](http://www.iltaonline.com)) perustettiin 1994, joskin alan tärkein kansainvälinen konferenssi juontaa juurensa 1970-luvun lopulle. Eurooppaan keskittyvä EALTA (European Association for Language Testing and Assessment; [www.ealta.eu.org](http://www.ealta.eu.org)) perustettiin vuonna 2004.

Tässä artikkelissa kuvaamme ensin, miten kielitaidon arviointi on muuttunut vuosikymmenien mittaan meillä ja muualla, ja mikä on tyypillistä nykytilanteelle. Yleiskuvan jälkeen kerromme, millaiset arviointimenetelmät ovat tavallisimpia isoissa kielitutkinnoissa ja laajoissa arviointitutkimuksissa, joissa on kymmeniä, jopa

satojatuhansia osallistujia vuosittain. Emme kuitenkaan unohda tavallisimpia luokkahuone- ja formatiivisen arvioinnin menetelmiäkään. Yritämme myös arvioida eri menetelmien etuja ja rajoituksia kielitaidon arvioinnin laadun arvioinnissa käytetyn Bachmanin ja Palmerin (1996) analyysikehikon piirteiden avulla. Tämän jälkeen esittelemme joitakin hankkeita, jotka liittyvät jollakin tavalla kielitaidon kansallisten arviointien tekemiseen Suomen oloissa. Osa näistä on arviointitutkimuksia, osa taas puhtaita kehityshankkeita, jotka kuitenkin ovat tuottaneet arviointitutkimuksenkin kannalta relevanttia tietoa.

## Millaista on nykyaikainen kielitaidon arviointi?

Kielitaidon arviointi on muuttunut monissa maissa varsin perusteellisesti siitä, mitä se oli vielä muutama vuosikymmen sitten. Muutos ei ole kuitenkaan tapahtunut joka puolella maailmaa samassa tahdissa eikä yhtä syvällisesti, vaan erilaisia perinteisiä, hitaasti muuttuvia käytänteitä löytyy vielä monista maista. On myös varsin tavallista, että käytänteet muuttuvat yksittäisessä maassa nopeammin joillakin koulutussektorilla kuin toisilla. Muutokset, joita seuraavassa käsitellään koskevat suurta osaa Eurooppaa sekä anglosaksisia maita eli sitä osaa maailmaa, jossa suurin osa kielitaidon arvioinnin tutkimuksestakin tehdään. Erityisen selvästi tässä kuvattu kehitys näkyy suurissa kansainvälisissä kielitutkinnoissa, jotka ovat tärkeitä esikuvia monille kansallisille tutkinnoille ja kokeille, joskus myös kouluissa tapahtuvalle arvioinnille.

Kielitaidon arviointi on muuttunut, koska taustalla oleva käsitys kielestä ja kielitaidosta on muuttunut ja koska tietoisuus arvioinnin haasteista on myös kasvanut. Aina 1960-luvulle asti vallalla ollut strukturalistinen kielitiede näki kielen joukkona erillisiä elementtejä ja taitoja, joita oli perusteltua myös opettaa ja arvioida 'puhtaasti' yksi taito kerrallaan (esim. Lado 1961). Osittain psykologiasta käsin levisi samaan aikaan myös näkemys siitä, että kielitaitoa on järkevää arvioida myös integratiivisemmin, esimerkiksi lukemisen, puhumisen, kirjoittamisen ja kuuntelun kaltaisten makrotaitojen tasolla (Carroll 1961). Nytemmin integroinnissa mennään toisinaan vieläkin pitemmälle ja joissakin testitehtävissä saatetaan ensin kuunnella, sitten lukea ja lopuksi kirjoittaa kuullun ja luetun pohjalta (esim. uudistetussa TOEFL iBT -tutkinnoissa). Toinen kielikäsitteeseen vaikuttanut suunta perustuu mm. sosiologiaan, keskustelun analyysiin ja etnografiaan. Jo 1960-luvun lopulla kielitieteen

teoreetikot argumentoivat, että olennaista kielenkäytössä on vuorovaikutus, asioiden tekeminen kielen avulla ja kielen käyttö erilaisiin tarkoituksiin erilaisissa tilanteissa (Hymes 1972; Halliday 1973).

Canale ja Swain (1980; Canale 1983) julkaisivat vuonna 1980 edelleen paljon siteeratun mallin kommunikatiivisesta kompetenssista, jonka mukaan kielitaidossa on erotettavissa kieliopillinen kompetenssi (sanaston ja rakenteiden hallinta), diskursiivinen kompetenssi (taito muodostaa koherentteja puhuttuja ja kirjoitettuja tekstejä), sosiolingvistinen kompetenssi (taito käyttää kieltä erilaisissa sosiaalisissa tilanteissa) ja strateginen kompetenssi. Viimeksi mainittu ei ole varsinainen kielellinen taito vaan yleisempi valmius, jolla paikataan muiden taitojen puutteita, esimerkiksi silloin kun oppija ei muista tarvitsemaansa vieraskielistä sanaa. Kielitaidon arviointiin ovat vaikuttaneet myös muut teoreettiset mallit, joissa useimmiten esiintyy varsin samantapaisia elementtejä kuin Canalen ja Swainin mallissa (esim. Bachman 1990). Yksi mielenkiintoinen uudehko lisä käsityksiimme kielitaidosta koskee kielellisen suorituksen rakentumista puhujien keskinäisessä vuorovaikutuksessa. Käsitys haastaa perinteisen näkemyksen kielitaidosta yksilön sisäisenä, psykologisena ominaisuutena. Kun arvioimme yksittäisen oppijan suullista kielitaitoa missä tahansa normaalia vuorovaikutustilannetta muistuttavassa tehtävässä, arvioimme samalla viestinnän toisen osapuolen (opettajan, toisen oppijan tai ulkopuolinen testaajan / haastattelijan) osaamista ja käyttäytymistä (ks. esim. McNamara & Roever 2006).

Yksi tapa tiivistää edellä esitetyt muutokset kielikäsitteissä on sanoa, että siinä missä ennen korostettiin tietoa kielestä (säännöt, rakenteet) nykyään painotetaan kykyä käyttää kieltä. Vallalla on ns. kommunikatiivinen kielikäsite ja arvioinnissa puhutaan kommunikatiivisesta, kielen käyttötaidon arvioinnista.

Kielenopetusta ja arviointia Euroopassa ja osaksi sen ulkopuolellakin on viime vuosina muokannut voimakkaasti Euroopan neuvoston vuonna 2001 julkaisema yleiseurooppalainen kielten oppimisen, opetuksen ja arvioinnin viitekehys (EVK tai Viitekehys; Euroopan neuvosto 2001, 2003). Se yhdistää kommunikatiivisen kielikäsitteen ja kriteeriperusteisen arvioinnin. ja on vaikuttanut suuresti monen maan kielikoulutukseen. Hyvä esimerkki tästä on Suomi, jonka peruskoulun ja lukion opetussuunnitelmat (Opetushallitus 2003, 2004) hyödyntävät Viitekehystä monella tapaa. EVK on luonut yhteistä käsitteistöä, joilla kuvata ja keskustella kielikoulutuksen eri puolista. Siinä kuvataan varsin yksityiskohtaisesti, mitä kielenkäytössä ja

oppimisessa tarvitaan (funktioita, tekstejä, tehtäviä, toimintoja; kompetensseja, taitoja). Lisäksi EVK sisältää 6-portaisen asteikon, jolla voidaan kuvata kielitaidon kehittymistä vasta-alkajan taidoista aina kielen erinomaiseen hallintaan. Itse asiassa EVK sisältää kymmeniä tällaisia asteikkoja, joissa kuvataan kielitaidon eri osa-alueita ja kielenkäyttötilanteita. Lisäksi joissakin asteikoissa on useampiakin tasoja. EVK mahdollistaa - ja itse asiassa kannustaa - oppijan kielitaitoprofiilin kuvaamisen. Ei ole pakko yrittää määritellä oppijan yleistaso (esim. B1) vaan on luontevampaa yrittää saada selville, millä alueilla hän on tätä tasoa, millä taas vielä sen alapuolella ja millä taas ehkä jo pitemmällä.

Viitekehys on ollut erityisen vaikutusvaltainen kielitaidon arvioinnin alueella, vaikka se ei oikeastaan ole erityisen sopiva testien laatimisen apuväline (Alderson ym 2006). Sen asteikkoja käytetään kuitenkin nykyään niin laajasti kielitaitovaatimusten määrittelyyn, että käytännössä kaikkien suurten kansainvälisten kielitutkintojen on ollut pakko suhteuttaa tuloksensa EVK:n tasoihin jo pelkästään kaupallisista syistä. Tämä pätee myös Euroopan ulkopuolisiin tutkintoihin kuten amerikkalaiseen TOEFL:iin (Tannebaum & Wylie 2008), koska sitä käytetään myös Euroopassa sekä sellaisissa Aasian maissa, joissa EVK on nykyään varsin tunnettu kielikouluttajien parissa. Myös juuri valmistunut Euroopan komission kansainvälinen kielitaidon arviointitutkimus rakentui vahvasti EVK:n pohjalle, kuten kerrotaan tarkemmin tuonnempana (Euroopan komissio 2012).

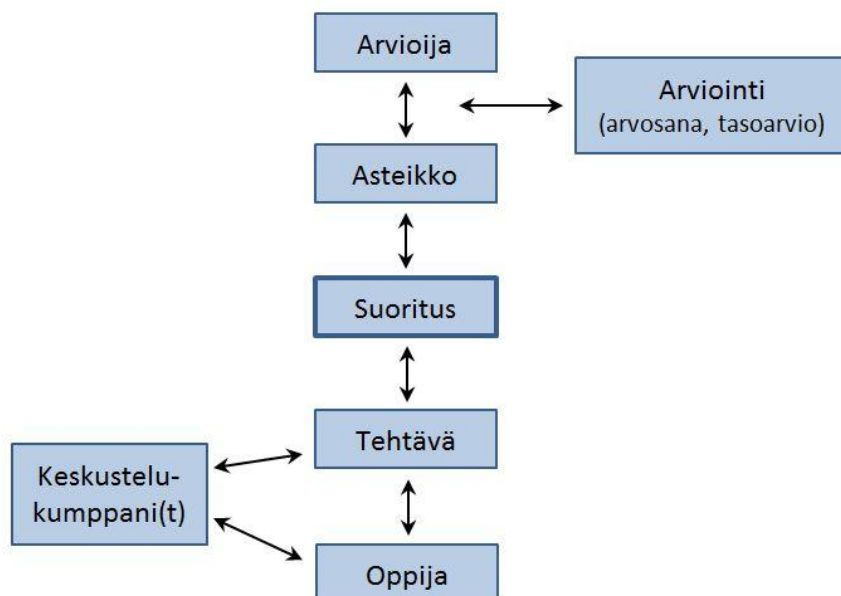
Kansainväliset muutokset kielitaidon arvioinnissa ovat luonnollisesti näkyneet myös Suomessa. Ylioppilastutkinnon kielikokeet uudistettiin varsin perusteellisesti jo 1970-luvulla, kun käännökset korvattiin kielen makrotaitojen testeillä. Tutkinnon tehtävävalikoimaa on myös vuosien mittaan laajennettu. Tutkinrossa ei tosin vielä ole puhumisen osakoetta, eikä kielikokeita ole virallisesti kytketty EVK:hon, vaikka asiaa onkin jo tutkittu (esim. Kaftandieva & Takala 2002). Uudet tuulet kielitaidon arvioinnissa näkyvät myös muissa suomalaisissa kielitutkinnoissa kuten Yleisissä kielitutkinnoissa ja valtionhallinnon kielitutkinnoissa, joissa molemmissa mm. tulokset raportoidaan EVK-perustaisesti.

## Muuttuvat menetelmät

### Mikä ja miksi on muuttunut?

Kielitaitokäsityksen muutos näkyy myös käytännössä arviointimenetelmien tasolla. Aikaisemmin ymmärtämisen arvioinnissa käytettiin pääasiassa monivalintakysymyksiä, mutta nykyään myös avoimempia tehtävätyyppejä, joissa oppija itse tuottaa vastaukset (esim. avokysymyksiä ja tekstin / kuullun sisällön tiivistämistä omin sanoin). Kielen tuottamisen eli puhumisen ja kirjoittamisen arviointi on ylipäänsä lisääntynyt - tarkoitetaanhan kielen käyttötaidolla monesti nimenomaan puhumista. Myös tehtävien kirjo on laajentunut, ja usein tehtävillä pyritään simuloimaan arkielämässä suoritettavia tehtäviä, vaikkapa kirjeen kirjoittamista tai suullisen esityksen pitämistä omasta työstä.

Kielikäsitusten evoluutio on muuttanut myös sitä, mitä ja miten arvioidaan. Samalla ymmärryksemme arvioinnin haasteista on parantunut eikä arviointia enää (yleensä) pidetä ongelmattomana toimintana kuten yleisesti aikaisemmin. Käyttämästämme arviointimenetelmästä johtuva metodiefekti mainittiin jo edellä, samoin kuin sen seurauksena yleistynyt useampien eri menetelmien käyttö. Toinen hyvä esimerkki arvioinnin haasteista on puhumisen ja kirjoittamisen arvioinnin monimutkaisuus (ks. Kuvio 1; McNamara 1996). Vaikka puhumisen ja kirjoittamisen arviointi on 'suorempaa' kuin ymmärtämisen siinä mielessä, että oppijan suoritus on nähtävissä ja/tai kuultavissa, on niidenkin arviointiprosessi silti varsin moniulotteinen. Oppijan suoritus riippuu käytetystä tehtävästä ja ainakin puhumisessa usein myös puheikumppanista. Annettuun arvioon vaikuttavat lisäksi arviointiasteikko ja siinä määritelty osaaminen sekä arvioijan ominaisuudet, kuten ankaruus / sallivuus, kokemus, tulkinta arviointikriteereistä ja kielitaitokäsitys.



KUVIO 1. Puhumisen arvioinnin osatekijät McNamaran (1996) mukaan

Kommunikatiivinen arviointi ja erityisesti tuottamistaitojen yleistynyt testaus on samalla lisännyt ns. kriteeriperusteista arviointia normiperusteisuuden kustannuksella. Kriteereihin perustuvassa arvioinnissa oppijan suorituksia verrataan nimensä mukaisesti johonkin ulkoiseen kriteeriin, esimerkiksi arviointiasteikon kuvauksiin. Oppijoiden arvosanoja ei siis anneta jonkin normin tai normiryhmän (esim. toiset samaan testiin osallistujat) perusteella. Kaikkien tunnetuimpien kansainvälisten kielitutkintojen tulokset raportoidaan nykyään kriteeriperusteisesti.

## Arvioinnin laatupiirteet

Seuraavassa kuvaamme lyhyesti kielitaidon arvioinnin laatuksiteerejä Bachmanin ja Palmerin (1996; 2010) luoman viitekehysten avulla. Sen jälkeen esittelemme yleisimmin käytettyjä arviointimenetelmiä ja lopuksi pyrimme arvioimaan menetelmien vahvuuksia ja ongelmia kyseisten laatuksiteerien kautta.

Tarkasteltaessa laadukkaan arvioinnin ominaispiirteitä, kielitaidon arvioinnissa laajalti tunnustettu teoreettinen malli on Bachmanin ja Palmerin laatima (1996; 2010). Tätä mallia on sittemmin kehitetty ja laajennettu arviointiteoreettisten virtausten

myötä, mutta käytännön arviointityötä jäsentämään soveltuvat edelleen sen perusosaset. Nämä ovat pätevyys (validius), luotettavuus (reliabelius), vuorovaikutuksellisuus (interactivity), autenttisuus (authenticity), käytännöllisyys (practicality) ja vaikuttavuus (impact). Validius koostuu tässä mallissa kaikkien osasten yhteisvaikutuksena eli on laajempi kuin perinteisimmät validiusmääritelmät ("Mitataan sitä, mitä pitääkin mitata"). Reliabelius määräytyy mittauksen virheettömyydestä ja suoritusolojen yhtenäisyydestä. Vuorovaikutuksellisuus ja autenttisuus liittyvät koesuorituksen edustavuuteen. Vuorovaikutuksellisuus viittaa siihen, missä määrin arvioinnin perusteena käytetty suoritus edellyttää samoja tietoja ja taitoja kuin se arkielämän suoritus, jota kokeessa simuloidaan. Siksi puhutaan myös sisäisestä autenttisuudesta. Ulkoinen autenttisuus liittyy arvioitavan suorituksen ja tosielämän suoritusilanteen samankaltaisuuteen. Käytännöllisyydessä on kyse arviointiin käytettävistä voimavaroista, niin henkisistä kuin taloudellisista. Vaikuttavuus viittaa niihin seurauksiin, joita arvioinnilla on yhtäältä arvioinnin kohteille ja heidän sidosryhmilleen ja toisaalta arviointia edeltävään opetukseen. Suomen oloissa kuuluisa esimerkki vaikutusvoimaisen kielikokeen heijastumista opetukseen on ylioppilastutkinnon kielikoetyyppien edustuminen oppimateriaaleissa ja niiden harjoittaminen lukio-opetuksessa.

## Yleisesti käytettyjä kielitaidon arviointimenetelmiä meillä ja muualla

Kielitaidon kirjalliset arviointitavat kohdentuvat luontevimmin kirjallisiin viestintätaitoihin, kuten kirjoittamiseen, mutta kynä ja paperi -testauksella on pyritty selvittämään myös sellaisten kielellisten viestintätoimintojen hallintaa, jotka arkielämässä tapahtuvat puhumalla. Suullisilla testitavoilla arvioidaan sen sijaan tavallisimmin juuri suullisen viestinnän hallintaa.

### Kielitaidon kirjalliset arviointitavat

Kirjallisen tuottamisen arvioinnista suomalaisessa koululaitoksessa on vankat perinteet. Käännöstehtävät, kirjoitelmat ja kieliopillisten rakenteiden täydennystehtävät ovat olleet erottamaton osa sekä koulukokeita että pian 160-vuotiasta ylioppilastutkintoa. Viestinnällisen kielitaitokäsityksen myötä arvioinnin

fokus on siirtynyt jonkin verran oikeakielisyydestä sujuvuuden ja ymmärrettävyyden suuntaan, mutta kovin kääntein tekeviä muutoksia kirjallisen tuottamisen arvioinnissa ei ole tapahtunut. Kirjoittamista on myös kauan käytetty puhumisen mittaamiseen.

Kieliopin ja sanaston hallintaa sekä laajempaa kirjallista tuottamista mittaavat menetelmät ovat kielitaidon arvioinnin perinteisintä välineistöä, joilla on yhä sijansa myös digitalisoituvassa maailmassa. Latinan opetuksen perua on kenties vanhin kielikokeen tehtävätyyppi: sanatarkka käänös kielestä toiseen. Perinteisellä käänöksellä testataan kielimuodon hallintaa, missä suhteessa se on reliabeliusmielessä luotettava ja myös helposti toimeenpantava ja siten käytännöllinen arviointitapa. On selvää, että käänöksen taustalla oleva kielitaitokäsitys on ahtaan muotopainotteinen ja muut kuin kieliopillinen kompetenssi suljetaan tarkoituksellisesti arvioinnin ulkopuolelle. Sama pätee käänöstä hieman modernimpiin aukkotehtäviin (produktiiviset ja monivalinta-aukkotehtävät). Lisärajoitteena voidaan nähdä monivalintatehtävien arvaustodennäköisyys.

Jonkin verran laajempaa kielitaitokäsitystä edustavat eriasteisesti ohjatut kirjallisen tuottamisen muodot: kirjoittaminen annetusta otsikosta tai eritasoisin sisältövihjein ohjattu kirjoittaminen. Näiden kautta voidaan mitata myös kielitaidon pragmaattisia ja sosiolingvistisiä puolia sekä myös yleisempiä kompetensseja. Tiukimmin ohjatut kirjoitustehtävät ovat testitehtävinä lähellä käänöstä niin vahvuuksiltaan kuin rajoituksiltaan. Vapaamuotoisempien kirjoitustuotosten etu on puhtaasti muotopainotteista testausta suurempi ulkoinen ja sisäinen autenttisuus, mikä toisaalta tuo mukanaan arvioijavaikutuksen ja reliabeliusongelmat. Myös kielitaitokäsityksen määrittely ja rajaus eroon yleisemmistä kompetenssialueista kuten yleissivistyksestä tai tietyn aihealueen erityisosaamisesta voivat nousta ongelmaksi.

### Kielitaidon suullisia arviointitapoja

Puheen arviointi on kiehtonut ja askarruttanut kieliväkeä vuosikymmenien ajan. Se ajankohtaistui erityisesti toisen maailmansodan jälkeen sekä poliittisella että kaupallisella saralla. Varhaisimpia puhekokeita oli tosin laadittu Britanniassa jo 1900-luvun alussa, mutta selkeästi laajemmin suullinen kielitaito ja sen systemaattinen testaus huomioitiin 1950-luvulla, jolloin American Council for Testing Foreign Languages ([www.actfl.org](http://www.actfl.org)) sisällytti testivalikoimaansa haastattelumuotoisen puhumista



mittaavan kokeen nimeltä Oral Proficiency Interview (OPI). Siinä erityisen arviointikoulutuksen saanut henkilö haastattelee yhtä kokeilasta kerrallaan aloittaen kysymyksenä kokeilajan oletetun tason alapuolelta ja päätyen sen ylärajan tunnusteluun. OPI on laadukas testityyppi, mikäli käytettävissä on koulutettu arvioija, mutta haastattelijakoulutus vaatii resursseja, yksilöhaastattelujen toteutus on aikaa vievää ja haastattelijan persoonan vaikutus testitulokseen on ilmeinen reliabeliusuhka. OPIsta kehitettiin sittemmin studiotallenteinen muunnos SOPI (Simulated Oral Proficiency Interview), jossa haastattelijana toimii äänitallenne, mikä vakioi haastattelijakäyttäytymisen, mutta kaventaa kokeen autenttisuutta inhimillisen vuorovaikutuksen puuttuessa. OPI lienee edelleen jonkinlainen puheen arvioinnin prototyyppi, vaikka sen rinnalle on viime vuosina noussut laajalti parimuotoinen testaus. Paritestausta puoltavat erityisesti käytännöllisyysnäkökohdat: lyhyemmässä ajassa pystytään testaamaan useampia kokeilaita ja arvioija vapautuu kokonaan tarkkailijan rooliin, mikä nopeuttaa arvosanan antamista. Yksimielisyyttä paritestausten monitahoisista validiushaasteista ei sen sijaan ole saavutettu, vaikka testiparien vuorovaikutusta ja sen seurauksia suorituksen kannalta on tutkittu laajasti (Brown ym. 2005; Fulcher 2003; Luoma 2004). Koulukontekstin oppimistulosten arvioinnissa paritestausta vastaa hyvin kielenopetuksessa vakiintuneita käytänteitä ja on siten ulkoisilta piirteiltään autenttinen.

Puhumisen arviointi on periaatteessa sisältynyt viestinnälliseen kielenopetukseen aina peruskoulun alkuajoista saakka, mutta vasta parikymmentä vuotta sitten suullisen kielitaidon arviointia ryhdyttiin ponnekkaasti edistämään sille omistetuilla erityishankkeilla. (esim. Huttunen, Paakkunainen & Pohjala 1995) Tavoitteena oli kehittää myös yleissivistävän lukion päättövaiheeseen vaikutusvoimainen suullista kielitaitoa mittaava koe. Puhumisen opettamista ja arviointia on kehitetty sitkeästi 1990-luvun alusta lähtien ja myös siihen liittyvä tutkimus on ollut vilkasta. Puhumisen arvioinnin suurin ongelma on ollut arviointitulosten marginaalinen asema verrattuna kirjallisiin kokeisiin niin kouluajana kuin päättövaiheissakin. Tilannetta parantamaan Opetusministeriö antoi 2010 asetuksen, jolla yksi lukion valinnaisista kursseista muutettiin keskeisimmässä oppimäärässä suulliseksi kurssiksi, jonka päätteeksi pidetään Opetushallituksen tuottama yhteinen koe.

Edellä kuvatut testitehtävyydet ovat luonteeltaan summatiiviseen arviointiin kehitettyjä tai sellaiseen aikojen saatossa vakiintuneita. Ilmiönä yhtä vanha tai kerta-

arviointia vanhempikin on kuitenkin se monimuotoinen palaute, jota oppija on saanut opettajaltaan tai muista lähteistä oppimisprosessinsa aikana. Tätä arviointia kutsuttakoon formatiiviseksi arvioinniksi ja sen keskeinen ero summatiiviseen arviointiin verrattuna on, että päätavoitteena on oppimisen edistäminen, ei ainoastaan sen tietynhetkisen tuloksen toteaminen. Käytetyimpiä formatiivisen arvioinnin muotoja koulujen kieltenopetuksessa ovat tuntikysymykset, tuntityöskentelyn seuranta, kestoltaan tai kohdennukseltaan suppeat ennalta ilmoitetut tai ilmoittamattomat kokeet. Formatiivisin menetelmin voidaan kerätä näyttöä myös summatiivisiin tarkoituksiin kuten numeroarvosanan pohjaksi, mutta tunnusomaista formatiiviselle arvioinnille on, että yksittäisen havainnon tai näyttötilanteen vaikuttavuus ei ole yhtä suuri kuin summatiivisten kokeiden. Formatiivista näyttöä voidaan antaa toistuvasti ja eri asioista, mikä laajentaa kielitaitokäsitettä. Systemaattisemman formatiivisen arvioinnin muotoja ovat erilaiset oppimispäiväkirjat ja portfoliot, tärkeimpänä näistä suomalaisessa kontekstissa Euroopan neuvoston edistämä European Language Portfolio (ELP), josta käytetään suomenkielistä nimitystä Eurooppalainen kielisalkku (EKS).

Kielisalkun teknisenä lähtökohtana on oppilaan suoritusten arkistointi ja pohdinta pitkällä aikavälillä. Suomalaisen version esikokeilu alkoi jo vuosituhannen vaihteessa Tampereen yliopistossa (Kohonen 2001) ja jatkui paikallisina kehittämishankkeina eri yliopistojen vaikutusalueilla usean vuoden ajan (Kohonen 2005). Tämä kokeilu- ja tutkimustoiminta tuotti runsaasti suomalaiseen kontekstiin soveltuvaa aineistoa oppilaiden autonomisuuden ja itsearviointitaitojen harjoittamiseen sekä opettajien ammatillisen kehittymisen ja yhteistyön edistämiseen. Vuoden 2012 lopussa Opetushallituksen on tarkoitus julkaista suomalaisten kieltenopettajien ja oppilaiden käyttöön verkossa vapaasti saatavilla oleva suomenkielinen kielisalkkuaineisto, joka sisältää kielipassin, kielenoppimiskertomuksen ja työkansion.

Kielisalkkututkimuksen tulokset olivat rohkaisevia, joskaan eivät tiukan empiristisessä katsannossa yleistyskelpoisia, koska kokeiluihin osallistuminen oli vapaaehtoista ja mukaan valikoitui näin jo entuudestaan hankkeen tavoitteisiin myönteisesti suhtautuvia ja kokeiltaviin työmuotoihin perehtyneitä henkilöitä. Kielisalkulla on silti kiistaton paikkansa autenttisen arvioinnin välineenä ja etenkin henkilökohtaisen kasvun dokumentoijana. Salkkututkimukset meillä ja muualla ovat myös osoittaneet, että oppilaiden metakognition kehittyminen on sidoksissa osallistaviin ja vuorovaikutteisiin työmuotoihin ja tuettuun reflektioon. Oppilaiden

ymmärrys oman kielitaitonsa kehittymisestä karttuu vähin erin, kun taitotasolausumat puretaan yksinkertaisiksi lausumiksi ("Osaan esitellä itseni kielellä x.") Salkkutyyppisen arvioinnin vahvuus on yksilöllisyys ja sitoutus itse asetettuihin tavoitteisiin. Sen heikkous on reliabelius, joten korkeaa yhdenmukaisuutta edellyttävissä massa-arvioinneissa kielisalkku ei ole parhaimmillaan (Hildén 2009), mikä ei kuitenkaan merkitse, ettei tätä formatiivisen arvioinnin muotoa voitaisi systematisoida ja integroida myös summatiivisen arvioinnin suuntaan.

Kooste tavallisimpien kielitaidon arviointitapojen vahvuuksista ja haasteista Bachmanin ja Palmerin (1990, 1996) jäsennyistä

Tehtävätyyppi	Vahvuuksia	Heikkouksia
<b>Kirjallisia arviointitapoja</b>		
käännös	reliabelius käytännöllisyys	autenttisuus muotoon rajoittunut kielitaitokäsitys
monivalinta	reliabelius käytännöllisyys	autenttisuus
produktiivinen aukkotehtävä	reliabelius käytännöllisyys	autenttisuus tavallisesti muotoon rajoittunut kielitaitokäsitys
ohjattu kirjoitelma	reliabelius käytännöllisyys autenttisuus (jos simuloi esim. työn tai vapaa-ajan tehtäviä)	autenttisuus usein muotoon rajoittunut kielitaitokäsitys
kirjoitelma otsikosta	autenttisuus vuorovaikutteisuus (mahdollistaa ei- kielellisten taitojen näytön)	mitattavan käsitteen raja- aus? (kielitaidon/ muiden taitojen ja ominaisuuksien osuus) kielitaidon eri ulottuvuuksien painoarvo?
<b>Suullisia arviointitapoja</b>		
yksilöhaastattelu	ulkoinen ja sisäinen autenttisuus	mitattavan käsitteen raja- aus? (kielitaidon/ muiden taitojen ja ominaisuuksien osuus) kielitaidon eri ulottuvuuksien painoarvo? suppea konteksti (vain yksi puhekumppani), epätasa-arvo (haastattelijavetoinen)
studiokoe	reliabelius testattavan sisällön, funktoiden, jne laajuus	autenttisuus (kasvokkaiskontaktin puuttuminen)

		muotopainotteinen kielitaitokäsite?
parikeskustelu	ulkoinen autenttisuus	reliabelius (puhekumppanin osuus?) käytännöllisyys
jatkuva havainnointi	ulkoinen autenttisuus käytännöllisyys	reliabelius
formatiiviset kokeet	reliabelius käytännöllisyys	vaikuttavuus muotopainotteinen kielitaitokäsite?
kielisalkku	ulkoinen autenttisuus vuorovaikutteisuus laaja kielitaitokäsite sisältää muita taitoja vaikuttavuus?	reliabelius käytännöllisyys?

Kielitaitokäsityksen laajentuessa ja pedagogisten näkemysten monimuotoistuesssa on vähitellen menetetty usko yhteen parhaiten toimivaan testaustapaan. Kaikkien sinänsä käyttökelpoisten menetelmien ongelma on ns. metodiefekti eli se, että testaustavalla on aina oma vaikutuksensa tuloksiin, koska jokainen menetelmä on aina joillekin oppijoille tutumpi, mieluisampi tai muulla tavalla sopivampi kuin toisille (Bachman1990). Yleinen suositus nykyään onkin käyttää useita erilaisia menetelmiä aina, kun se vain on käytännössä mahdollista.

## Uudistuvat menetelmät

Selkeä suuntaus isoissa kielitutkinnoissa on lisääntyvä tietotekniikan käyttö arviointiprosessin eri vaiheissa. Jo pitkään on teknologiaa sovellettu tehtävien laadintaan, ylläpitoon ja tulosten analyysiin, mutta nykyään myös testit yhä useammin suoritetaan tietokoneen välityksellä ja joissakin tapauksissa tietokone jopa arvioi suoritukset automaattisesti.

Yksi edelläkävijä tietokoneen ja internetin käytöstä kielitesteissä oli 1990-luvun lopussa laadittu Dialang, joka on 14-kielinen, verkon kautta suoritettava testistö, joka antaa oppijoille palautetta osaamisen vahvuuksista ja heikkouksista sekä harjoituttaa myös heidän itsearviointitaitojaan (Alderson 2005; Alderson & Huhta 2005). Kyse ei siis ollut kielitutkinnoista vaan diagnostis-formatiivisesta arviointivälineestä, jossa

tietokone pisteyttää oppijoiden suorittamat monivalinta- ja yksinkertaiset täydennys tehtävät ja antaa niistä välitöntä palautetta.

2000-luvulla perusteellisesti uudistettu amerikkalainen TOEFL (Test of English as a Foreign Language) on tietotekniikkaa ja verkkoon hyväksi käytävä kielitutkinto, joten sen nimikin on nykyään TOEFL iBT (TOEFL internet Based Test; ks. tarkemmin Chapelle, Enright & Jamieson 2008). Koko testi, mukaan lukien puhumis- ja kirjoitustehtävät suoritetaan tietokoneella; tehtävät ja vastaukset lähetetään paikallisesta testikeskuksesta internetin kautta USA:han arvioitaviksi. Ymmärtämistestit arvioi tietokone, mutta puhumistestien digitaaliset nauhoitteet arvioivat ihmisarvioijat. Kirjoitustehtävien arviointi on yhdistelmä ihmis- ja konearviointia: kumpikin antaa oman arvionsa samalla asteikolla, ja jos ne eroavat toisistaan, toinen ihmisarvioija ratkaisee erimielisyyden. TOEFL iBT:n verkossa suoritettavissa olevassa puhumisen näytetestissä on koneellinen arviointi, joten kyseisen harjoitustestin suorittaminen on ehkä helpoin tapa tutustua käytännössä automaattiseen vieraan kielen puheen arviointiin.

Pearsonin kustannusyhtiön uusimmat englannin testit menevät automaatiossa vieläkin pitemmälle: lukuun ottamatta ajoittaisia ihmisarvioijien suorittamia tarkistuksia, kaikki suoritukset testeissä arvioi tietokone, myös puhumisessa ja kirjoittamisessa (ks. <http://pearsonpte.com>).

Suoritusten automaattinen arviointi on yksinkertaista monivalinnan kaltaisissa tehtävissä. Myös hyvin rajattujen avovastausten koneellinen arviointi on helppoa, jos vastaukset koostuvat vain sanasta tai parista ja ne ovat erittäin ennustettavia, jotta arviointimekanismille voidaan antaa kattava lista hyväksyttävistä vastauksista (vastauksissa esiintyvät kirjoitusvirheet aiheuttavat tosin sekä käytännöllisiä että periaatteellisia ongelmia arvioinnille). Laajemmat avovastaukset sekä puhe- ja kirjoitussuoritukset vaativat kuitenkin älykkäämpiä arviointiratkaisuja. Tavallisesti suorituksista kootaan laaja vastausaineisto, jonka avulla arviointialgoritmi voidaan 'kouluttaa' arvioimaan halutulla tavalla. Lisäksi arviointi voidaan rajata esimerkiksi sellaisiin puheen piirteisiin, jotka soveltuvat helpommin automaattiseen arviointiin (äänteiden laatu, ääntämisnopeus, sana- ja lausepainotus, puheen rytmitys, vastauksissa odotettavissa olevien sanojen / ilmausten selkeys). Sisältöä ja esimerkiksi kielen sociolingvistisiä piirteitä ei edes pyritä analysoimaan, koska ajatellaan, että em. ääntämiseen ja sujuvuuteen liittyvät piirteet riittävät antamaan vieraan kielen puheesta

Huhta, A. & Hildén, R. 2012. Kielitaidon arvioinnin metodologisia vaihtoehtoja.

tarpeeksi edustavan ja yleistettävän näytteen (ks. esim. Bernstein, Van Moere & Cheng 2010). Niinpä esimerkiksi Pearsonin puhumistesteissä käytetään tehtäviä, joissa oppijan täytyy toistaa kuulemiaan sanoja ja lauseita, vastata lyhyesti kuulemiinsa kysymyksiin, lukea ääneen tehtävävihossa näkemänsä teksti ja kertoa jotakin annetusta aiheesta.

## Arviointitutkimusta 2000-luvulla

Seuraavassa kuvaamme lyhyesti muutamia kansallisia ja kansainvälisiä kielitaidon arviointitutkimuksia sekä myös muita kielten tutkimushankkeita, jotka ovat aiheen kannalta relevantteja. Arviointitutkimuksella tarkoitamme siis kielitaidon arviointia mm. opetussuunnitelman toteutumisen seuraamiseksi ja kansallista ja joskus myös paikallista kielikoulutusta koskevan päätöksenteon tueksi (ks. myös Huhta 2012 katsaus arviointitutkimuksista Kielikoulutuspoliittisen verkoston verkkolehdestä).

Kansainvälisiä kielitaidon arviointitutkimuksia ei ole monta. IEA (International Association for the Evaluation of Educational Achievement) -järjestö toteutti ensimmäisen kansainvälisen kielten arviointitutkimuksen jo 1970-luvun alussa. Englannin kielen tutkimukseen osallistuivat myös suomalaiset abiturientit, tosin vielä varsin vaatimattomalla menestyksellä (Huhta 2012, Takala 2004). 1980-luvulla IEA arvioi äidinkielen kirjoitustaitoja varsin laajassa tutkimuksessa, johon Suomikin osallistui (Leimu 2004). Tutkimus kohdistui peruskoulun 6. ja 9. vuosiluokan oppilaisiin sekä abiturientteihin. Järjestö yritti uusia vieraiden kielten tutkimushankkeen 1990-luvulla, mutta joutui rahoitusvaikeuksien vuoksi luopumaan asiasta. Nyt IEA on suunnittelemassa 2010-luvun puolivälin jälkeen mahdollisesti toteutettavaa laajaa englannin kielen arviointitutkimusta.

Joukko Länsi- ja Etelä-Euroopan maita toteutti 2000-luvun alussa englannin osaamistutkimuksen, joka kohdistui perusopetuksen loppuvaiheeseen (Bonnet 2004). Oppilaat Pohjoismaissa Suomi mukaan lukien sekä Alankomaissa osoittautuivat englannin taidoiltaan varsin samantasoisiksi, mutta selkeästi paremmiksi kuin ranskalaiset ja espanjalaiset oppilaat. Tutkimusvälineiden epävarman laadun vuoksi on kuitenkin vaikea esimerkiksi erotella eri maita toisistaan tämän tarkemmin (Huhta 2012). Tutkimuksen tuloksia ei myöskään raportoitu EVK-tasoina, joten on vaikea sanoa, minkä tasoista englannin osaaminen itse asiassa oli osallistujamaissa.

Merkittävin vieraiden kielten kansainvälinen arviointitutkimus on jonkin aikaa sitten valmistunut Euroopan komission toimesta toteutettu European Survey of Language Competences (ESLC). Tutkimukseen osallistui 14 EU-maata, muttei valitettavasti Suomi, joten meillä ei ole ajantasaista, luotettavaa tietoa oppilaittemme osaamisesta verrattuna muihin maihin. Hankkeessa testattiin kussakin maassa kahden siellä yleisimmin opetetun kielen osaamista lukemisessa, kuuntelussa ja kirjoittamisessa (testatut kielet olivat englanti, espanja, italia, ranska ja saksa). Useimmissa maissa kohdejoukko oli alemman toisen asteen koulutuksen loppuvaiheessa olevat eli Suomen peruskoulun 9. luokkaa vastaavalla tasolla opiskelevat. Tutkimuksessa käytetyt välineet kehitettiin EVK:n pohjalta ja myös sen tulokset raportoitiin EVK:n taitotasosteikon avulla (ks. tarkemmin Euroopan komissio 2012, <http://ec.europa.eu/languages/eslc>.) Tutkimus on tarkoitus toteuttaa tulevaisuudessa noin viiden vuoden välein.

### Kansalliset arviointihankkeet vieraissa kielissä

Opetushallituksen yhtenä lakisääteisenä tehtävänä on seurata perusopetuksen oppimistuloksia ja käyttää niitä opetuksen laadun parantamiseen informaatio-ohjauksen hengessä. Oppimistulosten seuranta-arviointi on osa valtakunnallista koulutuksen arviointia. Seuranta-arviointi perustuu tehtävään, josta Opetushallitus huolehtii perusopetuslain 21 §:n, valtioneuvoston asetuksen 805/2008 ja opetus- ja kulttuuriministeriön päättämän koulutuksen arviointisuunnitelman perusteella.

Tuloksia käytetään ainoastaan tutkimus- ja kehitystarkoituksiin eikä niiden perusteella seurata yksittäisten koulujen tai oppilaiden tuloksia tai aseteta näitä paremmuusjärjestykseen. Arvioinnit ovat otosperustaisia kattaen noin 5% ikäluokasta ja toistuvat noin kerran vuosikymmenessä. Varhaisimmat kielten opetuksen oppimistulosten arvioinnit tehtiin peruskoulun alkuvaiheessa englannin kieltä koskien (Havola-Pitkänen 1974; Havola & Saari 1993). Viimeisimmät vieraiden kielten oppimistulosten arvioinnit ovat kohdentuneet englannin pitkään oppimäärään (Tuokko 2000), ruotsin, ranskan, saksan ja venäjän pitkiin oppimääriin (Tuokko 2002; Väisänen 2002) ja ruotsin keskipitkään oppimäärään (Tuokko 2009). Näissä arvioinneissa

kuvataan ja analysoidaan perusopetuksen 9. luokan oppilaiden kielitaitoa erikseen ymmärtämisen ja tuottamisen osalta ja verrataan oppimistuloksia kaikissa kansallisissa arvioinneissa toistuviin taustamuuttujiin. Näitä ovat sukupuoli, kieliryhmä (kansalliskielet suomi ja ruotsi), erilaiset alueelliset ja sosiaaliset indikaattorit (esim. kuntamuoto,) ja oppilaskohtaiset taustamuuttujat (vanhempien koulutustaso, kouluarvosanat, asennoituminen ja jatko-opintosuunnitelmat). Kielitaitoa mitataan perinteisesti kuullun ja luetun ymmärtämistehtävillä ja kirjoitustehtävillä, joiden lisäksi viimeisimmissä arvioinneissa on ollut mukana myös suullisia tehtäviä.

Keskeisiä tuloksia englannin kielen oppimistulosten saavuttamisesta ovat olleet kielitaidon paraneminen vuosikymmenten saatossa, tyttöjen ylivoimaisuus poikiin nähden, ruotsinkielisten suomenkielisiin ja kaupunkien maaseutuun verraten (Tuokko 2000). Koulujen väliset erot näyttävät olevan kasvussa monien aineiden arviointitulosten valossa. Englannin kielen nuorisokulttuurinen valta-asema on edistänyt kielen osaamista kouluopetuksesta riippumatta.

Suomalaisessa perusopetuksessa harvemmin opiskelluissa kielissä järjestettiin oppimistulosten arvioinnit saksan, ranskan ja venäjän pitkille oppimäärille 2002. Tällöin tutkitut oppilaat olivat opiskelleet vuoden 1994 opetussuunnitelman perusteiden mukaisesti, joten heidän tuloksiaan ei vielä saattanut suoraan verrata vuonna 2004 käyttöön otettuun tapaan ilmaista taitotasotavoite viitekehysperustaisella asteikolla. Arvioinnin yhtenä tavoitteena oli verrata toisiinsa ensimmäisenä ja toisena vieraana kielenä aloitettujen oppimäärien oppimistuloksia. (Väisänen 2003.)

Ranskan kielessä parhaat tulokset saavutettiin luetun ymmärtämisessä ja kirjoittamisessa, heikoimmat kielen rakenteiden hallinnassa. Osaamisessa havaittiin selvä ero A1- ja A2-oppimäärien välillä. Saksan kielessä tehtävistä suoriuduttiin keskimääräisesti paremmin kuin ranskan kielessä, joskin vahvuudet ja heikkoudet esiintyivät samoilla kielitaidon alueilla: parhaiten menestyttiin kuullun ymmärtämisessä ja kirjoittamisessa, heikoimmin rakenteiden hallinnassa. Saksan kielessä ei havaittu eroa ensimmäisenä ja toisena vieraana kielenä aloittaneiden suorituksissa. Venäjän kielessä vahvimmin hallittiin niin ikään kuullun ymmärtäminen ja kirjoittaminen, heikoimmin rakenteet. Suuria eroja ilmeni sen sijaan A1- ja A2-oppimäärien välillä, mikä johtui kielikoulujen oppilaiden ja syntyperäisten oppilaiden vahvasta panoksesta A1-venäjän aineistossa. (Väisänen 2003, 6–7.)



Kaiken kaikkiaan parhaiten hallittu kielellinen viestintätaito harvemmin opiskelluissa kielissä oli kirjallinen tuottaminen. Yhteistä kielille oli myös tyttöjen parempi menestys poikiin nähden ja myös se havaittiin, että tytöt valitsevat ranskaa, saksaa ja venäjää enemmän kuin pojat. Tyttöjen asenteet olivat myönteisempiä kuin poikien. Tuloksista näkyy orastava alueellinen epätasa-arvo, sillä arviointiin osallistuneita ja harvemmin opiskeltuja kieliä opiskelevia oli huomattavasti vähemmän Pohjois- ja Itä-Suomessa ja maaseudulla kuin Etelä-Suomessa ja taajamissa.

Toisen kotimaisen kielen ruotsin oppimistuloksia on mitattu vuosina 2001 pitkän oppimäärän ja 2008 keskipitkän oppimäärän osalta. Vuonna 2001 tekstin ymmärtäminen osoittautui molempien oppimäärien oppilaille keskimäärin helpoimmaksi ja kieliopilliset rakenteet vaikeimmaksi kielitaidon alueeksi. Molempien oppimäärien kokeissa tytöt menestyivät poikia paremmin kaikilla osa-alueilla. Erot olivat pienimmät tekstin ymmärtämisessä ja suurimmat kirjoittamisessa. Vuonna 2008 toteutetun toisen kotimaisen kielen arvioinnin mukaan oppimistulokset ruotsin keskipitkässä oppimäärässä eivät vastanneet opetussuunnitelman perusteissa asetettua tavoitetasoa (Tuokko 2009). Erityisen ongelmalliseksi osoittautuivat kirjoittaminen ja puhuminen, joissa molemmissa vain vajaa viidesosa oppilaista saavutti opetussuunnitelman perusteissa asetetun hyvän osaamisen tavoitteen (Tuokko 2009, 5–6). Merkittävää vaihtelua koulujen välillä esiintyi myös arviointikäytännöissä ja arvosanojen suhteessa arviointitehtävien osoittamaan osaamiseen.

Tuoreimmissa englannin ja ruotsin oppimistulosten arvioinneissa saavutettuja tuloksia on kuvattu eurooppalaisen EVK-taitotasosteikon ja sen suomalaisen opetussuunnitelmasovelluksen avulla (Tuokko 2007). Taitotasoperustainen laadinta- ja raportointikäytäntö jatkuu seuraavassa ja tähän saakka laajimmassa vieraiden kielten oppimistulosten seuranta-arvioinnissa. Se toteutetaan keväällä 2013 ja silloin kartoitetaan ruotsin ja englannin pitkä oppimäärä sekä ranskan, saksan ja venäjän pitkä ja lyhyt oppimäärä. Yhteiskunnallisena tavoitteena on tuottaa tietoa opetushallinnon ja opetuksen järjestäjien käyttöön tulevaisuuden kieltenopetusta varten. Tieteellisenä tavoitteena on kartoittaa, kuvata ja analysoida saavutettuja oppimistuloksia ja selittää niitä opiskelukontekstin ja oppilaan piirteillä. Tutkimusnäkökulmasta pääongelmat ovat:

1. Missä määrin vieraiden kielten ja ruotsin pitkän oppimäärän tavoitteet on saavutettu suhteessa hyvän osaamisen tavoitetasoihin ja kielenkäytön viestinnällisiin sisältöihin?
2. Mitä yhteyksiä saavutetuilla oppimistuloksilla on oppilaiden ja heidän opiskelukontekstinsa piirteisiin?
3. Miten oppilaiden ja heidän opiskelukontekstinsa piirteet selittävät saavutettuja oppimistuloksia?

Vieraiden kielten ja A-ruotsin oppimistulosten arvioinnin 2013 tutkimusongelmat, muuttujat, data ja aineiston analyysimenetelmät

Ongelma	muuttujat	aineisto	analyysimenetelmä ja tunnusluvut
1. Missä määrin vieraiden kielten ja ruotsin pitkän oppimäärän tavoitteet on saavutettu suhteessa hyvän osaamisen tavoitetasoihin ja kielenkäytön viestinnällisiin sisältöihin?	*kuullun ja luetun ymmärtämistaito pistemäärä *suullinen kielitaito *kirjoitustaito *arviointitehtävien ominaisuudet	*kuullun ja luetun ymmärtämistehtävien pistemäärät *puhe- ja kirjoitussuoritus- ten taitotasoarviot *tehtäväkuvaus- lomake	ratkaisuprosentit erotteluindeksit korrelaatiot standard setting (bookmark method)
2. Mitä yhteyksiä saavutetuilla oppimistuloksilla on oppilaiden ja heidän opiskelukontekstinsa piirteisiin?	*rehtorin tieto ja käsitykset opiskelun kontekstitekijöistä (kunnan ja koulun taso) *opettajan tieto ja käsitykset ammatti-identiteetistä, opetustavoista *oppilaan tausta, asenteet ja oppimista edistävät käytänteet koulussa ja sen ulkopuolella	rehtorikysely opettajakysely oppilaskysely	
3. Miten oppilaiden ja heidän opiskelukontekstinsa piirteet selittävät saavutettuja oppimistuloksia?			

Englannin otos käsittää noin 3700 ja ruotsin noin 2000 oppilasta. Englannin otannassa ositteina käytettiin kieltä, maakunta ja kuntamuotoa. Harvemmin opiskelluista kielistä

venäjän molempien oppimäärien ja ranskan pitkä oppimäärän opiskelijat ovat kaikki mukana arvioinnissa. Saksan molempien oppimäärien ja ranskan pitkän oppimäärän opiskelijoista noin puolet otettiin mukaan. Vuoden 2013 arvioinnissa mitataan kaikkia kielellisiä viestintätaitoja oppilaille suunnatuin tehtävin. Lisäksi tietoa kerätään kyselyin, joihin vastaavat rehtorit, ryhmien opettajat ja oppilaat. Rehtorikyselyllä kartoitetaan kielenopiskelun resursseja ja kehystekijöitä, opettajakyselyllä opettajien kelpoisuutta, ammatillista identiteettiä, arvoja ja käsityksiä kielenopetuksesta sekä opetuskäytänteitä. Oppilaskyselyssä pureudutaan vakiintuneiden taustatietojen ja asenneosuuden lisäksi oppilaan tapoihin edistää kielitaitoaan koulussa ja sen ulkopuolella. Kyselytieto rinnastetaan kielikohtaisiin oppimistuloksiin, joiden yhteyksiä tutkimalla pyritään selittämään eritasoisia oppimistuloksia ja tekemään johtopäätöksiä ja antamaan suosituksia seuraavien opetussuunnitelman perusteiden laatijoiden käyttöön.

### Arviointitutkimushankkeita

Helsingin yliopistossa toteutettu HY-TALK -hanke (2006 –2010) tutki suullisen kielitaidon kehittymistä eurooppalaisen viitekehyksen taitotasoasteikon eri tasoilla viidessä kielessä. Kielikohtaisten tutkimustehtävien ohella hankittiin tietoa opetussuunnitelman perusteisiin sisältyvän taitotasoasteikon erottelukyvystä. Tutkimustulosten perusteella puhumisen arviointitehtävien laadinta on vaativaa etenkin yhteismitallisuuden ja ulkoisen autenttisuuden tasapainottamisen osalta. Taitotasojen käyttöön kouliintuneet arvioitsijat antavat hyvin yhteneviä arvioita keskenään ja ovat enimmäkseen myös johdonmukaisia omissa arvioissaan. Taitotasojen erottelukyky on myös hyvällä tolalla, joskin tutkimus pitäisi toistaa erityyppisillä arvioijaryhmillä. (Hildén & Huhta 2011.)

Jyväskylän yliopiston Soveltavan kielentutkimuksen keskus ja Kielten laitos ovat vuodesta 2007 lähtien tutkineet useissa hankkeissa vieraan ja toisen kielen kirjoittamisen kehittymistä, nyttemmin myös lukemisen kehittymistä. Näistä tärkeimmät ovat kolme Suomen Akatemian rahoittamaa tutkimusta:

- CEFLING (2007-09) Linguistic Basis of the Common European Framework for L2 English and L2 Finnish ([www.jyu.fi/cefling](http://www.jyu.fi/cefling))

- TOPLING (2010-13) Toisen kielen oppimisen polut ([www.jyu.fi/topling](http://www.jyu.fi/topling))
- DIALUKI (2010-13) Toisen tai vieraan kielen lukemisen ja kirjoittamisen diagnosointi ([www.jyu.fi/dialuki](http://www.jyu.fi/dialuki))

Kaikissa näissä hankkeissa tarkastellaan, miten EVK:n ja opetussuunnitelman taitotasolla kuvattu toiminnallinen kielitaito liittyy kielen lingvististen piirteiden esiintymiseen ja kehittymiseen. Toisin sanoen hankkeissa selvitetään, millaiset kielelliset (esim. kieliopilliset ja sanastolliset) piirteet tai piirrekimput ovat tyypillisiä eri taitotasolle arvioiduissa oppijoiden kirjoitussuorituksissa. CEFLING-hanke tutki tätä suomalaisilla englannin oppijoilla ja maahanmuuttajataustaisilla suomen oppijoilla peruskoulun 7. - 9. luokilla. TOPLINGissa tutkitaan näiden kielten lisäksi myös suomalaisia ruotsin oppijoita. TOPLINGin tutkimusasetelma on pitkittäinen eli samojen oppilaiden kirjoitustaidon kehittymistä seurataan usean vuoden ajan. Tutkimuksessa on mukana oppilaita ala- ja yläkoulusta, lukiosta sekä yliopistosta. DIALUKI eroaa näistä hankkeista siten, että kirjoitussuoritusten lingvistinen analyysi on pienemmässä roolissa. Sen sijaan pyritään selvittämään, millaiset kognitiiviset, lingvistiset, motivaatio- ja taustatekijät ennustavat taitoa lukea ja kirjoittaa toisella kielellä. Eriyisen kiinnostuneita ollaan mahdollisten ongelmien ennustamisesta ja ymmärtämisestä. Tutkimus kohdistuu suomalaisiin englannin oppijoihin (alakoululaisista lukiolaisiin) ja venäjää kotikielensä puhuviin suomen oppijoihin (ala- ja yläkoululaisia).

Kaikissa näissä hankkeissa oppijoiden kirjoitussuoritukset on arvioitu sekä EVK-asteikolla että opetussuunnitelmien liitteenä olevalla taitotasosteikolla. Arvioijat koulutettiin käyttämään kumpaakin asteikkoa mm. eri taitotasojen edustavien esimerkkisuoritusten avulla, jotka oli kerätty kirjoitustehtävien pilotointivaiheessa tai aikaisemmassa tutkimuksessa (CEFLING-hanke toteutettiin ennen kahta uudempaa hanketta). Kaikki suoritukset arvioi vähintään kaksi arvioijaa, joskus jopa neljäkin. CEFLING-arviointien tilastollinen analyysi moderniin testiteoriaan perustuvien menetelmien avulla (esim. Linacre 2009) toi esille useita mielenkiintoisia seikkoja. Ensinnäkin monet arvioijat eroavat toisistaan arviointilinjansa suhteen: jotkut ovat yleensä ankarampia, toiset taas sallivampia, mitä on vaikea muuttaa koulutuksenkaan avulla (ks. esim. Knoch 2010). Tärkeää on kuitenkin, että arvioijat ovat johdonmukaisesti sellaisia kuin ovat, eivätkä anna välillä yllättävän ankaria tai lempeitä

arvioita yleislinjaansa verrattuna. Tutkimuksen yhteydessä järjestetty koulutus oli tässä mielessä onnistunutta, koska vain pieni osa arvioinneista osoittautui laadultaan ongelmallisiksi. Yksi johtopäätös tuloksista on se, että suoritusten kaksoisarviointi on tärkeää; vain yhden arvioijan käyttö voi helposti vinouttaa tuloksia, koska arvioijien peruslinjaa on vaikea muuttaa. Ehkä tärkein ja mielenkiintoisin havainto oli, että sekä EVK- että OPS-asteikko osoittautuivat toimivan arviointiasteikkoina, mistä ei aikaisemmin ole julkaistu tutkimusta (Huhta, Alanen, Tarnanen, Martin & Hirvelä, arvioitavana). Koulutetut arvioijat pystyivät keskimäärin erottelemaan jopa OPS-asteikon 10 tasoa toisistaan. Toki analyysi osoitti, että jotkut tasot näyttivät hiukan hankalammin erottuvilta, joka voi merkitä sitä, että niiden kuvauksia voisi olla syytä vielä tarkentaa.

Mielenkiintoisia suomalaisia ja kansainvälisiä arviointihankkeita on toki muitakin, joita ei tässä ole mahdollista yksityiskohtaisesti kuvailla. Näistä mainittakoon Yleiset kielitutkinnot, aikuisille tarkoitettu kielitaidon sertifiointitutkinto, jossa on jo 1990-luvulta lähtien tehty kriteeriperusteista testausta (lisätietoja Opetushallituksen verkkosivuilla [www.oph.fi](http://www.oph.fi)). KORU-hankeessa kehitettiin 2000-luvulla puolestaan korkeakoulujen ruotsin kielen arvioinnin yhdenmukaisuutta kriteeriperusteisen arvioinnin kautta (Elsinen & Juurakko-Paavola 2006). Eurooppalaisista yhteistyöhankkeista kannattaa mainita Webcef ja Cefcult. Edellisessä laadittiin suullisen kielitaidon arviointia tukeva verkkopohjainen ympäristö, jossa arviointia voi harjoitella Viitekehukseen perustuvien asteikkojen avulla (ks. [www.webcef.eu](http://www.webcef.eu)). Cefcult-hanke keskittyi puolestaan kulttuurien välisen suullisen viestinnän arvioinnin kehittämiseen (ks. [www.cefcult.eu](http://www.cefcult.eu)). Lopuksi mainittakoon vielä kaksi eurooppalaisena yhteistyönä syntynyttä EVK-perusteista arviointia tukevaa resurssia. Ceftrain-sivuston avulla voi tutustua EVK-tasoihin ja harjoitella niiden avulla tehtävää arviointia. Euroopan neuvosto on puolestaan tuottanut erinomaisen monikielisen puhumisen maamerkitön, joiden avulla havainnollistetaan Viitekehysten taitotasoa englannin, espanjan, italian, ranskan ja saksan kielessä (ks. <http://www.ciep.fr/publi/evalcert/dvd-productions-orales-cecrl/index.php>).

Kielitaidon arvioinnin ajankohtaisia haasteita ja suuntia tulevaan

Käsitlemme seuraavassa muutamaa kielitaidon arvioinnin ajankohtaista kysymystä, jotka käytännön ja teoriaperusteisen tutkimuksen valossa vaativat erityistä harkintaa. Nämä ovat arvioinnin tarkoitus, käsitevalidius eli kielitaidon ja muiden taitojen sekoittuminen arvioinnissa ja taitotasokuvauksiin perustuva arviointi. Lopuksi hahmottelemme kielitaidon arvioinnin todennäköisiä ja toivottuja kehityskulkuja parin seuraavan vuosikymmenen ajalle.

## Haasteita

Keskeisimmät arviointia yleensä ja siten myös kielitaidon arviointia koskevat kysymykset liittyvät arvioinnin päämääriin ja tulosten käyttöön päätöksenteon pohjana: Miksi arvioidaan? Mihin arvioinnin tuloksia käytetään? Vaikeudet erottaa arvioinnin eri tarkoituksia aiheuttavat paljon ongelmia ja väärinymmärrystä jopa opetuksen asiantuntijapiireissä.

Ihanteellista olisi, jos testin aiottu tarkoitus ja sen todellinen käyttö vastaisivat toisiaan, koska testin tarkoitus ohjaa vahvasti sen sisältöä, arviointikriteerejä ja muita keskeisiä ominaisuuksia. Testin tulosten käyttö jälkikäteen muihin tarkoituksiin olla joko mahdotonta tai tuottaa virheellisiä johtopäätöksiä. Pari esimerkkiä valaisee asiaa.

Jos kurssin lopussa järjestetyn puhekokoon tarkoituksena on mahdollistaa arvosanojen antaminen, suoritusten arviointitavaksi on voitu valita holistinen arviointi, joka perustuu kokonaisvaikutelmaan, jonka opettaja saa oppilaan suorituksesta. Kokeen jälkeen voi kuitenkin herätä tarve antaa oppilaille formatiivista palautetta suorituksista. Tämä on tietenkin vaikeaa, koska opettajan ei kokeen alkuperäistä tarkoitusta varten tarvinnut arvioida suorituksia yksityiskohtaisesti eritellen, eikä tehdä systemaattisia muistiinpanoja kunkin oppijan vahvuuksista ja heikkouksista. Arvosanan antaminenhan ei tätä edellyttänyt.

Jos arviointi olisi alun perin suunniteltu tuottamaan formatiivista tietoa opettajalle ja oppilaille, arviointitavaksi ei ehkä olisi päätetty valita kurssin lopussa pidettävää koetta, vaan jokin muu tapa, jolla tällaista tietoa olisi voitu saada niin hyvissä ajoin, että sitä olisi voitu hyödyntää jo kurssin aikana.

Toinen esimerkki arvioinnin tarkoituksen ja tulosten käytön välisestä ristiriidasta liittyy kansallisiin testeihin ja tutkintoihin. Ylioppilastutkinnon tuloksista laaditaan

lehdissä usein koulujen ranking-listoja, joiden uskotellaan kertovan koulujen, erityisesti niissä annettavan opetuksen laadusta. Ylioppilastutkinto on kuitenkin tarkoitettu yksilöiden osaamisen sertifiointiin; todistus kertoo siitä, miten hyvin oppilas on saavuttanut lukion oppimäärän verrattuna muihin samaan tutkintoon osallistuneisiin. Jos tutkinnon tarkoitus olisi selvittää koulujen laatueroja, testien yhteydessä olisi pitänyt kerätä hyvin paljon tietoa sekä oppilaiden että koulun taustoista ja nämä olisi myös sitten pitänyt raportoida (aiheesta lisää esim. Kuusela 2003). Koska tämä ei ole koskaan ollut ylioppilastutkinnon tarkoitus, tällaista tietoa ei tutkinnon yhteydessä kerätä. On lisäksi hyvin mahdollista, että tutkinnon käyttö uusiin tarkoituksiin vaatisi muutoksia myös sen sisältöön.

Rajanveto kielitaidon ja muun osaamisen välillä on yksi kielitaidon arvioinnin isoja ja vaikeasti ratkaistavia haasteita. Monesti pidetään ongelmallisena, jos kielitaitoa testattaessa mitataan samalla jotakin muuta kuten yleistietoa, älykkyyttä tai persoonallisuutta. Haluttaisiin testata 'puhdasta' kielitaitoa. Toive on ymmärrettävä, mutta monella tapaa ongelmallinen. Esimerkiksi lukemisen teorit, joissa mallinnetaan, miten tekstiä ymmärretään, ovat jo pitkään pitäneet itsestään selvänä, että lukijan on osattava integroida lukemansa teksti tietoonsa tekstin aihepiiristä ja yleensä tietoonsa maailmasta, jotta voisi kunnolla ymmärtää sen (Grabe 2009). Ulospäin suuntautuva persoonallisuus on puolestaan eittämättä avuksi monissa suullisen viestinnän tilanteissa, ei vain vieraalla vaan myös äidinkielellä. Oma alueensa on ns. erityisalojen kielitaito ja sen arviointi. Useimmat ihmiset tuntevat joitakin asioita ja aihealueita paremmin kuin muut keskimäärin, koska ovat niistä kiinnostuneita tai koska se kuuluu heidän työhönsä. He myös osaavat näihin liittyvää kieltä paremmin kuin muut, jos esimerkiksi testissä sattuu olemaan jotakin näihin aiheisiin liittyvää. Ellei kyse ole nimenomaan erityisalojen kielitaidon arvioinnista (Douglas 2000), on tietysti järkevää pyrkiä välttämään aiheita, joiden tiedetään suosivan joitakin oppijoita toisten kustannuksella. Tätä on kuitenkin käytännössä mahdotonta täysin välttää, joten ainoa keino on pyrkiä sisällyttämään testeihin mahdollisimman useita, vaihtelevia aiheita.

Aikaisemmin kuvasimme, miten Eurooppalainen viitekehys on lisännyt kriteeriperusteisen kielitaidon arvioinnin suosiota viime vuosina. Käytännössä tämä tarkoittaa useimmiten sitä, että oppijoiden suorituksia arvioidaan jonkin asteikon avulla, jossa on sanallisesti määritelty erilaisia osaamisen tasoja. Kriteeriperusteinen arviointi on tietysti useimmissa tilanteissa hyvin järkevää, koska on hyödyllistä, jos

opettajalla, oppilailla ja muilla mahdollisilla arviointitulosten käyttäjillä on jokin käsitys siitä, mitä testin tulos tai annettu arvosana tarkoittaa taitona käyttää kyseistä kieltä. Toki perinteinenkin numeroarviointi voi perustua kriteereihin, ja etenkin opettajalla voi olla suhteellisen hyvä käsitys siitä, miten eri arvosanat liittyvät tietyn kurssin sisältöjen osaamiseen. Taitotasoihin perustuvassa arvioinnissa on useimmiten kyse yleisemmällä tasolla tapahtuvasta arvioinnista, vaikka mikään ei periaatteessa estä laatimasta jokaiselle kurssille omaa sanallista arviointiasteikkoa. Suomessa tunnetuin kielitaidon tasokuvaus, jota voidaan käyttää myös arviointiin, on opetussuunnitelmiin sisältyvä 10-portainen asteikko, jossa määritellään osaamisen tasot puhumiselle, kuuntelemiselle, lukemiselle ja kirjoittamiselle.

Puhumisen ja kirjoittamisen arviointi on perinteisesti perustunut arviointiasteikkojen käyttöön, joten se on jo perusluonteeltaan kriteeriperusteista. Näiden taitojen arviointi esimerkiksi monivalintatehtävien avulla tuntuisi varsin kummalliselta, vaikka on sitä toki joskus yritettykin. Puhumisen ja kirjoittamisen arviointi voi kuitenkin olla varsin subjektiivista, jos arvioijat eivät jaa samaa käsitystä kielitaidosta, jos arviointiasteikot ja kriteerit eivät ole selkeitä, jos käytettävissä ei ole hyviä taitotasoja havainnollistavia maamerkkejä tai jos arvioijia ei ole koulutettu. Tarpeeksi luotettavaan arviointiin on kuitenkin mahdollista päästä, ja lisäksi on kehitetty tilastollisia analyysimenetelmiä, joiden avulla on mahdollista mallintaa ja ottaa huomioon arvioijien välisiä eroja ankaruudessa (esim. McNamara 1996).

Erityinen haaste on käyttää arviointiin EVK-asteikkoja, koska ne eivät luonteeltaan ole ihanteellisia arviointiin; useimmat niistä ovat kielitaitoa varsin yleisellä tasolla kuvaavia asteikkoja (vrt. Alderson 1991). Suomalaisen opetussuunnitelmien asteikot (OPS-asteikot) suunniteltiin niin, että niihin lisättiin arviointia helpottavia kuvauksia mm. oppijoiden tekemistä virheistä ja muista puutteista heidän suorituksissaan, jotka tyystin puuttuvat EVK-asteikoista (vrt. Hildén & Takala 2007). OPS-asteikkojen toimivuudesta arvioinnissa ei kuitenkaan ole vielä paljon näyttöä; erityisen haasteelliselta tuntuu asteikon hienojakoisuus - pystyvätkö arvioijat erottelemaan kaikki asteikon tasot toisistaan? Eräät suomalaiset tutkimushankkeet, joita kuvaamme tuonnempana, näyttävät osoittavan, että sekä EVK- että OPS-asteikot voivat toimia tuottamistaitojen arvioinnissa, ainakin kun arvioinnin suunnitteluun kiinnitetään tarpeeksi huomiota.



Ymmärtämisen arviointi on lähtökohdiltaan vaikeampaa, kuin puhumisen ja kirjoittamisen arviointi, koska ymmärtämistä ei voida tarkastella samalla tavalla 'suoraan' kuin puhe- ja kirjoitussuorituksia. Poikkeuksen tästä muodostavat sellaiset tilanteet, joissa oppija voi toiminnallaan tai käyttäytymisellään osoittaa ymmärtäneensä esimerkiksi annetun toimintaohjeen. Useimmiten joudumme kuitenkin turvautumaan epäsuoriin testaustapoihin: annamme oppijoiden luettavaksi tekstin, josta tehtyihin kysymyksiin annetuista vastauksista teemme johtopäätöksiä ymmärtämisestä. Monivalinnan kaltaiset objektiivisesti arvioitavat tehtävätyypit saattavat tosin antaa harhakuva ymmärtämisen testauksen selkeydestä. Kun testausprosessia tarkastelee kokonaisuutena, on kuitenkin selvää, että esim. tekstien valintaan ja kysymysten laadintaan liittyy hyvin paljon sattumanvaraisuutta ja subjektiivisuutta, joka jää ehkä piiloon, mutta jolla voi olla huomattava vaikutus tuloksiin.

Ymmärtämisen mallintaminen taitotasoina on vielä lapsenkengissään, joten ymmärtämisen tasojen laatiminen ei ole kovin tarkkaa puuhaa. Vieraan kielen ymmärtämisen kehittymistä on ylipäänsä tutkittu varsin vähän kielenoppimisen tutkimuksessa, joka tyypillisesti keskittyy puhumisen ja kirjoittamisen tarkasteluun (Alderson & Huhta 2011). Lisähaasteen tuo pyrkimys raportoida jonkin tietyn ymmärtämistestin tulokset taitotasoina; miten voidaan perustella se, että vaikkapa 20 pistettä testistä tarkoittaa A2-tasoista osaamista, mutta 21 pistettä merkitsee jo B1-tason ymmärtämistä? Tällaisiin kysymyksiin vastaamista varten on arviointikirjallisuudessa kehitetty lukuisia taitotasorajojen asettamismenetelmiä (engl. standard setting; esim. Kaftandjjeva 2004; Cizek 2011), joita on ainakin isommissa testeissä ja tutkinnoissa mahdollista käyttää varsin menestyksellisesti.

## Tulevia kehityskulkuja

Arvioinnin lähtökohta on arvioitavan käsitteen määrittely. Kielitaidon käsite on kokenut suuria muutoksia viime vuosikymmenien aikana, kuten artikkelin alkupuolella kuvasimme. Latinan innoittamasta kieliopillisesta muutosysteemistä on edetty kielen luonnollisen käytön moninaisuuteen. Samalla kielen kirjalliset, staattiset ilmentymät ovat antaneet tilaa puhutun kielen ja sen käyttötilanteiden dynamiikalle. Kielenopetuksen saralla vuoden 1970 Peruskoulun opetussuunnitelma oli tärkeä

taitekohta siirryttäessä muoto-opillisesta painotuksesta elävän elämän kielenkäytön opetuksen suuntaan. Kieliopillisten elementtien ohien tavoitteiksi kirjattiin pragmaattisia ja sosiolingvistisiä tietoja ja taitoja. Viimeisimmät opetussuunnitelmaperusteet asettavat kieliaineille monia muitakin kuin kielellisiä tai edes viestinnällisiä tavoitteita. Perusopetuksen päättyessä tähdätään osallistuvan kansalaisen valmiuksiin ja todetaan kielet taito- ja kulttuuriaineiksi (Opetushallitus 2004, 136). Yleissivistävässä lukiokoulutuksessa kielet ovat myös tietoine.

Useimmissa Euroopan maissa virallisesti omaksuttu kielitaitokäsitys pohjautuu Euroopan neuvoston julkaisemaan Kielten oppimisen, opettamisen ja arvioinnin viitekehykseen, johon on koostettu Euroopan yhdentymiskehityksen kannustaman kokeilu- ja tutkimusyhteistyön sato noin 40 vuoden ajalta. Viitekehyyksessä kielitaito koostuu varsinaisista kielellisistä ja ns. yleisistä kompetensseista. Kieliopilliset, pragmaattiset ja sosiolingvistiset tiedot ja taidot lukeutuvat kielellisiin kompetensseihin. Yleiset kompetenssit kattavat kielenkäyttöä ja opiskelua ohjaavia valmiuksia, joista eräät ovat osin synnynnäisiä (kuten elämähallintataidot ja taipumukset suhtautua uusiin asioihin), toiset taas pitkälti opetuksen ja harjoituksen kautta muokattavia (kuten kulttuurienväliset tiedot ja taidot tai opiskelutaidot).

Opetussuunnitelman perusteissa sekä perusopetuksen että lukiokoulutuksen tavoitteisiin kuuluvat kieliaineissa kielitaito, kulttuuritaidot ja opiskelutaidot. Näiden lisäksi kielten opetusta ja opiskelua koskevat koulukasvatuksen yleiset tavoitteet, joihin lukeutuvat esimerkiksi valmius elinikäiseen oppimiseen, aktiivinen ja tavoitteellinen toimijuus kulttuurien välillä, suvaitsevaisuus ja yhteistyökyky. Voimassa olevissa opetussuunnitelman perusteissa kunkin oppimäärän kielitaitotavoitteet on kuvattu eurooppalaisen viitekehyyksen asteikon suomalaisen sovelluksen avulla. Siihen sisältyy viitekehyyksen kuudesta tasosta viisi, ja nämä on edelleen jaettu alatasoihin siitä syystä, että alkuperäiset viitekehyyksitasot vaativat niin isoja tuntimääriä opiskelua, että suomalaisen perusopetuksen päättyessä useimmissa oppimäärissä päästäisiin hädintuskin toiselle portaalle. Alatasojen asettaminen paikallisten tarpeiden mukaan on myös viitekehyyksen laatijoiden suosituksen mukaista. Suomalainen sovellus on verrannettu empiirisesti viitekehyyksitasoihin (Hildén & Takala 2006).

Kielitiedon ja -taidon keskinäinen suhde askarruttaa opettajia, mutta sitäkin enemmän kielitaidon arvioijia, koska yhteiskunnallisesti tärkeitä kielikasvatuksen tavoitteita ei kyetä nyky menetelmin kovinkaan kattavasti ja luotettavasti arvioimaan.

Tosiasialla kielikasvatuksen tavoitteiden toteutumista kuvaava arvosana perustuu todennäköisesti varsin usein kirjalliseen kielitaitoon ilman vuorovaikutusta tai kosketusta käytännön kielenkäyttötilanteisiin. Tämä selittyy osin perinteen painolastilla, osin käytännöllisistä tekijöistä.

Vieläkin murrosvaiheessa olevan kielitaitokäsityksen lisäksi myös arviointitavat ja käytänteet muuttuvat tulevaisuudessa. Ehkä näkyvin muutos on tietotekniikan vahvempi rooli kielitaidon(kin) arvioinnissa. Uudet, hyvinkin pitkälle tietotekniikkaa hyödyntävät kansainväliset testit, joita kuvattiin aiemmin osoittavat, mihin suuntaan suomalaisetkin tutkinnot väistämättä siirtyvät jossain vaiheessa. Myös opetusviranomaisten aloitteesta tehtävät arviointitutkimukset toteutetaan joskus tietokoneiden avulla, kuten osa Euroopan komission 2012 valmistuneesta vieraiden kielten oppisaavutustutkimuksesta jo tapahtui (osallistuvat maat ja koulut saivat päättää suorittivatko ne kielitestit perinteisesti paperilla vai tietokoneella).

Toinen tietekniikkaan liittyvä muutos on houkutteleva mahdollisuus, jonka toteutuminen ei kuitenkaan ole varmaa. Tietotekniikka mahdollistaa nimittäin opetukseen integroitavien formatiivisten harjoitus- ja arviointitehtävien järjestämisen ja välittömän ja yksityiskohtaisen palautteen antamisen oppilaille paljon systemaattisemmin kuin mihin kukaan yksittäinen opettaja pystyy opettaessaan kokonaisia oppilasryhmiä. Periaatteessa tarvittava teknologia on jo olemassa; haaste on pikemminkin pedagoginen ja kustannuspoliittinen. Miten parhaiten integroida tietokoneella tapahtuva formatiivinen arviointi opetukseen, millaiset arviointitehtävät ovatärkevimpiä ja erityisesti millainen ja miten ajoitettu palaute auttaa oppilaita tehokkaimmin?

Formatiivisen arvioinnin ja palautteen ei välttämättä tarvitse rajoittua siihen, mitä opettaja voi luokassaan tehdä, joko tietotekniikan avulla tai ilman sitä. Myös erilaiset kansalliset kielitutkinnot ja testit voisivat tuottaa huomattavasti nykyistä enemmän opetusta ja oppimista tukevaa tietoa. Huolimatta siitä, mitä aikaisemmin kirjoitimme vaaroista, joita voi syntyä, kun yhteen tarkoitukseen kehitettyä testiä aletaankin käyttää johonkin toiseen tarkoitukseen, tätä asiaa kannattaisi vakavasti miettiä sen ilmeisten hyötyjen takia. Esimerkiksi ylioppilastutkinnon kielikokeista voisi yleisarvosanan lisäksi tuottaa yksityiskohtaisempaa taito- ja tehtäväkohtaista tietoa tutkinnon suorittajien vahvuuksista ja heikkouksista lukioiden kieltenopettajien käyttöön. Ajatukseen elinikäisestä oppimisesta sopisi hyvin, että myös oppijat saisivat

Huhta, A. & Hildén, R. 2012. Kielitaidon arvioinnin metodologisia vaihtoehtoja.

tutkinnosta muutakin palautetta kuin arvosanan. Sama periaate koskee myös kansallisia arviointitutkimuksia; niistäkin voitaisiin joskus tulevaisuudessa antaa hyvinkin yksityiskohtaista palautetta tutkittujen oppilaiden osaamisesta. Tutkintojen ja testien tietokoneistuminen helpottaa tämänkaltaisten muutosten toteuttamista käytännössä.

## Lähteet

- Alderson, J.C. 1991. Bands and Scores. Teoksessa C. Alderson & B. North (toim.) *Language testing in the 1990s: The communicative legacy*. Lontoo: Macmillan, 71-86.
- Alderson, J. C. 2005. *Diagnosing Foreign Language Proficiency. The Interface between Learning and Assessment*. New York: Continuum.
- Alderson, J. C., & Huhta, A. 2005. The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing*, 22(3), 301-320.
- Alderson, J.C. & Huhta, A. 2011. Can research into the diagnostic testing of reading in a second or foreign language contribute to SLA research? Teoksessa L. Roberts, G. Pallotti & C. Bettoni (toim.) *EUROSLA Yearbook Volume 11 (2011)*. Amsterdam: John Benjamins, 30-52.
- Alderson, J. C., Figueras, N., Kuijper, H. Nold, G., Takala, S. & Tardieu, C. 2006. *Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR Construct Project*. *Language Assessment Quarterly*, 3 (1), 3-30.
- Bachman, L. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. & A. Palmer. 2010. *Language testing in practice*. Oxford: Oxford University Press.
- Bachman, L. & Palmer, A. 1996. *Language testing in practice. Designing and developing useful language tests*. Oxford: Oxford University Press.
- Bernstein, J., Van Moere, A. & Cheng, J. 2010. Validating automated speaking tests. *Language Testing* 27(3), 355-378.
- Bonnet, G. (toim.) 2004. *The Assessment of Pupils' Skills in English in Eight European Countries 2002. The European network of policy makers for the evaluation of education systems*.
- Brown, A., Iwashita, N., & McNamara, T. 2005. An examination of rater orientations and test taker performance on English for academic purposes speaking tasks. *TOEFL Monograph Series*. Princeton, NJ: Educational Testing.
- Canale, M. 1983. On some dimensions of language proficiency. Teoksessa J. Oller (toim.) *Issues in language testing research*, Rowley, Mass.: Newbury House, 333–342.
- Canale, M. & Swain, M. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.

Huhta, A. & Hildén, R. 2012. Kielitaidon arvioinnin metodologisia vaihtoehtoja.

Carroll, J. 1961. Fundamental considerations in testing for English language proficiency of foreign students. Washington, DC Testing Center for Applied Linguistics.

Cizek, G. 2011. Setting performance standards: Foundations, methods, and innovations (2.painos). New York: Routledge.

Chapelle, C., Enright, M., & Jamieson, J. (toim). 2008. Building a Validity Argument for the Test of English as a Foreign Language. New York: Routledge.

Douglas, D. 2000. Assessing languages for specific purposes. Cambridge: Cambridge University Press.

Elsinen, R. & Juurakko-Paavola, T. 2006. Korkeakouluopiskelijoiden ruotsin kielen taidon arviointi. Hämeen Ammattikorkeakoulu: Hämeenlinna.

Euroopan komissio. 2012. First European survey on language competences. Final report. Bryssel: European Commission / Education and Training.

(<http://ec.europa.eu/languages/eslc/index.html>)

Euroopan neuvosto. 2001. The Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Cambridge: Cambridge University Press.

Euroopan neuvosto. 2003. Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys. WSOY.

Grabe, W. 2009. Reading in a second language. Moving from theory to practise. Cambridge: Cambridge University Press.

Fulcher, G. 2003. Testing second language speaking. Harlow: Pearson.

Halliday, M. 1973. Relevant Models of Language. Teoksessa M. Halliday (toim.) Explorations in the Functions of Language. New York: Elsevier North-Holland.

Havola-Pitkänen, L. 1974. Englannin kielen koulusaavutukset peruskoulun ala-asteen ja yläasteen päättyessä. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitoksen julkaisuja 231.

Havola, L. & Saari, H. 1993. Kaksikymmentä vuotta englannin kuullun ymmärtämistä - onko taidoissa edistytty? Teoksessa P. Linnakylä & H. Saari (toim.) Oppiiko oppilas peruskoulussa. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 157-173.

Hildén, R. & Takala, S. 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. Reihe: Dichtung – Wahrheit – Sprache Bd. 7. Münster: LIT Verlag, 291-300.

Hildén, R. 2009. [In pursuit of validity: an empirical pilot on validating the Finnish version of the European Language Portfolio developed for upper grades of compulsory basic education](#). In Melbourne papers in language testing. 14 (2), 32-71.

Hildén, R. & Huhta, A. 2011. Validity of an analytic CEFR-based oral proficiency scale: The quality of the five subscales. Presentation at The Eight Annual Conference of EALTA, Siena, Italy 5th - 8th of May, 2011.

<http://www.ealta.eu.org/conference/2011/programme.html>

Huhta, A. 2012. Kansalliset ja kansainväliset kielitaidon arviointitutkimukset – millaista tietoa niissä tuotetaan? Kieli, koulutus ja yhteiskunta - Kielikoulutuspoliittisen verkoston verkkolehti. Lokakuu 2012. ([www.kieliverkosto.fi](http://www.kieliverkosto.fi))

Huhta, A., Alanen, R., Tarnanen, M., Martin, M. & Hirvelä, T. (jätetty arvioitavaksi) Same tasks, same scales, different languages - exploring the rating process in a SLA study. (Language Testing).

Huhta, A. & Hildén, R. 2012. Kielitaidon arvioinnin metodologisia vaihtoehtoja.

Huttunen, I., Paakkunainen, R. & Pohjala, K. (toim.) 1995. Suullisen kielitaidon opetus ja arviointi lukiossa. 1995. Lukion suullisen kielitaidon kokeilun loppuraportti. Oulun yliopiston kasvatustieteiden tiedekunnan tutkimuksia 95/ 1995.

Hymes, D. 1972. *Towards Communicative Competence*. Philadelphia: University of Pennsylvania Press.

Kaftandjieva, F. & Takala, S. 2002. Englannin A-tason ylioppilaskokeen tulosten linkittäminen Euroopan neuvoston taitotasosteikolle. Euroopan neuvoston seminaari, Helsinki 30.6 – 2.7, 2002. Opetushallitus & Council of Europe. Julkaisematon seminaariesitelmä. Lisätietoja: [sjtakala@hotmail.com](mailto:sjtakala@hotmail.com)

Kaftandjieva, F. 2004. Standard setting. Reference Supplement to the Preliminary Pilot version of the Manual for Relating Language examinations to the Common European Framework of Reference for Languages: learning, teaching, assessment. Strassbourg: Council of Europe, 1-43.

Kohonen, V. 2000. Student reflection in portfolio assessment: making language learning more visible. *Babylonia* 1, 13-16.

Kohonen, V. (toim.) 2005. Eurooppalainen kielisalkku Suomessa. Tutkimus- ja kehittämistyön tuloksia. Helsinki: WSOY.

Knoch, U. 2010. Investigating the effectiveness of individualized feedback to rating behavior – a longitudinal study. *Language Testing* 28(2) 179–200.

Kuusela, J. 2003. Lukioiden tuloksiin vaikuttavista tekijöistä. *Moniste* 13/2003. Helsinki: Opetushallitus.

Lado, R. 1961. *Language testing; The construction and use of foreign language tests: A teacher's book*. New York: PMcGraw-Hill.

Leimu, K. (toim.) 2004. Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa. Koulutuksen tutkimuslaitos, Jyväskylän yliopisto.

Linacre, M. 2009. *A User's Guide to FACETS v 3.66.0*. Chicago: Winsteps.

McNamara, T. 1996. *Measuring second language performance*. Harlow, UK: Addison Wesley Longman Ltd.

Luoma, S. 2004. *Assessing speaking*. Cambridge. Cambridge University Press.

McNamara, T. & Roever, C. 2006. *Language Testing: The Social Dimension*. Hoboken, NJ: Wiley-Blackwell.

Opetushallitus. 2003. *Lukion opetussuunnitelman perusteet 2003*. Helsinki:

Opetushallitus.

Opetushallitus. 2004. *Perusopetuksen opetussuunnitelman perusteet 2004*. Helsinki:

Opetushallitus.

Opetusministeriö. 2010.

Takala, Sauli. 2004. Englannin kielitaidon tasosta Suomessa. Teoksessa K. Leimu (toim.) *Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa*. Koulutuksen tutkimuslaitos, Jyväskylän yliopisto, 255–275.

Tannenbaum, R. & Wylie, C. 2008. *Linking English-language test scores onto the Common European Framework of Reference: An application of standard-setting methodology*. TOEFL iBT Research Report 06. Princeton: Educational Testing Service.

Tuokko, E. 2000. *Peruskoulun 9.vuosiluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi 1999*. Opetushallitus.

Tuokko, E. 2002. *Perusopetuksen päättövaiheen ruotsin kielen oppimistulosten kansallinen arviointi 2001*. Opetushallitus.

Huhta, A. & Hildén, R. 2012. Kielitaidon arvioinnin metodologisia vaihtoehtoja.

Tuokko, E. 2007. Mille tasolle perusopetuksen englannin opiskelussa päästään? : perusopetuksen päättövaiheen kansallisen arvioinnin 1999 Eurooppalaisen viitekehyksen taitotasoihin linkitetyt tulokset. Jyväskylä Studies in Humanities 69.

Tuokko, E. 2009. Miten ruotsia osataan peruskoulussa? Perusopetuksen päättövaiheen ruotsin kielen B-oppimäärän oppimistulosten kansallinen arviointi 2008. Opetushallitus.

Väisänen, T. 2003. Perusopetuksen 9.luokan A-kielenä opetettavan ranskan, saksan ja venäjän oppimistulosten kansallinen arviointi. Opetushallitus.