

Kielenoppimisen kysymyksiä

toimittaneet

Kari Sajavaara

Arja Piirainen-Marsh



Soveltavan kielentutkimuksen teoriaa ja käytäntöä

KIELITAIDON ARVIOINTI

Ari Huhta & Sauli Takala

Artikkelimme esittelee kielitaidon arviointia lähinnä vieraiden kielten näkökulmasta, mutta samat periaatteet sopivat yleensä äidinkielenkin arviointiin. Arvioinnilla tarkoitamme kaikenlaista toimintaa, jossa oppijoiden kielitaidosta hankitaan näytteitä jotakin tarkoitusta varten mm. testien, jatkuvan observoinnin tai itsearviointin avulla. Käymme läpi kielitaidon arvioinnin keskeisiä kysymyksiä, kuten kielitaitoa, arvioinnin tarkoituksia ja sen laatua. Aloitamme katsauksella tieteenalan historiaan ja sen nykytilaan.

Kielitaitoa on arvioitu kautta aikojen siellä, missä kieliä on opetettu tai tehty päätöksiä ihmisten tulevaisuudesta kielitaidon perusteella (ks. Raamattu, Tuomarien kirja 12: 4–6 esimerkkinä 'kielitestistä'). Arviointi on ollut luonnollinen, jokapäiväinen osa tällä vuosisadalla yleistynyttä vieraan kielen opetusta. Monissa maissa on lisäksi kehitetty suuria, kansallisia tutkintoja osoitukseksi tietyn oppimäärän suorittamisesta, esimerkiksi Suomessa ylioppilastutkinto (vuodesta 1864 lähtien). Näissä testataan yleisesti äidinkielen hallintaa, monesti myös vieraita kieliä. Isoja kielitestejä on laadittu myös suurten kielialueiden oppilaitosten tarpeisiin, helpottamaan vieraskielisten opiskelijoiden valintaa (Spolsky 1995).

Tyypillistä sekä äidinkielen että vieraankielen arvioinnille ennen 1960-lukua oli se, ettei testeihin liittynyt paljонkaan tutkimusta: uskottiin, ettei arviointi periaatteessa ollut kovin ongelmallista. Monessa tapauksessa tutkimukseen on herätty vasta aivan viime vuosina.

Vuosisadan alusta aina 1960- ja 1970-luvuille asti kielitaidon arviointi oli tieteenalana vahvasti sidoksissa psykologiaan ja psykometriikkaan, jossa tutkitaan mittareiden laatua tilastotieteen keinoin. Vaikutus oli selvintä Yhdysvalloissa 1950- ja 1960-luvuilla, jolloin herättiin tutkimaan arvioinnin luotettavuutta (reliabeliutta) ja uskottiin, että luotettavaan mittaamiseen päästäisiin vain monivalinnan kaltaisten menetelmien avulla.

Kielitaidon arviointiin on luonnollisesti vaikuttanut kehitys yleisemmässä arviointitutkimuksessa. Amerikkalainen Ralph Tyler esitti jo 1920-luvulla, että kasvatuksessa vallitsee läheinen suhde tavoitteiden, oppimiskokemusten ja arvioinnin välillä. Näkemykset ilmenevät mm. Bloomin kehittämässä taksonomioissa, joissa esitetään hierarkkinen malli kognitiivisen toiminnan tasoista – tietäminen, ymmärtäminen, soveltaminen, analyysi, synteesi, arviointi. Näitä alettiin yleisesti käyttää testien sisällön määrittelyyn myös vieraissa kielissä (esim. Valette 1971). Monet uudet suuntauokset, kuten kriteeriviitteisyys, latenttien piirteiden testiteoria, käsitevalidiuden korostaminen, kiinnostus tulosten yleistettävyyteen ja mitausvirheen tarkempi analyysi, ovat myös peräisin yleisemmästä mittaamisteoriasta. Viimeisten kolmenkymmenen vuoden aikana monet muutkin tieteenalat ovat alkaneet vaikuttaa kielitaidon arviointiin, merkittävimmin sosiologia ja antropologia, joista ovat peräisin monet kommunikatiivisen arvioinnin ajatukset.

Nykyään kielitaidon arviointi on oma tieteenalansa, jonka merkitys ja yhteydet lähialoihin ovat kasvamassa, useastakin syystä. Arviointi on monesti vallan käyttämistä – testien avulla päätetään usein arvosanoista ja pääsystä tiettyihin ammatteihin tai opintoihin. Vallan käyttöön tulisi liittyä myös vastuuta arvioinnin laadusta ja

seurauksista. Tutkintojen laatijoita vaaditaankin yhä useammin osoittamaan testitulosten luotettavuus. Arvioinnin etiikka, testien sosiaaliset vaikutukset ja osallistujien oikeudet ovat näin ollen nousseet huomion kohteeksi (Norton 1997; Hamp-Lyons 1997), ja puhutaan avoimesti huonoista testeistä ja testien väärinkäytöstä (ks. Shepard 1992). Toisaalta kielen oppimisen ja opettamisen tutkimus on löytänyt kielitaidon arvioinnista hyödyllisen yhteistyökumppanin. Testien tutkimus on usein samalla kielitaidon tutkimusta, mistä on hyötyä muulle soveltavalle kielitieteelle sekä sisällöllisesti että parantuneena tutkimusvälineiden tuntemuksena. Vastaavasti kielitaidon arvioijat hyötyvät muusta soveltavasta kielitieteestä, sillä parempi käsitys kielestä ja sen oppimisesta auttaa laatimaan parempia arvioinnin välineitä (ks. tarkemmin Bachman 1988).

1 MITÄ KIELITAIDON ARVIOINNISSA ARVIOIDAAN?

Seuraavassa esittelemme kielitaidon arviointiin eniten länsimaissa vaikuttaneet käsitykset kielitaidosta. Etenemme perinteisestä taitojatottelusta (lukeminen, puhuminen jne.) uudempiin kommunikatiivisiin käsityksiin kielitaidosta. Laajempia katsauksia teorioista ovat esittäneet mm. Takala ja Havola (1983), Gardner (1985), Botha (1992), Huhta (1993b) ja McNamara (1995, 1996).

Teoreettista tietoa tarvitaan sitä enemmän mitä monipuolisemmin oppijan kielitaitoa halutaan arvioida. Arviointi pienen kielikursin yhteydessä ei ehkä tunnu tilanteelta, jossa tarvitaan teoreettista tietämystä, koska arvioinnin sisältö näyttäisi määräytyvän suoraan opetuksen sisällöistä. Kuitenkin tässäkin käsitykset kielitaidosta vaikuttavat arviointiin: Miten kommunikatiivista arvioinnin tulisi olla ja mitä menetelmiä pitäisi käyttää? Arvioidaanko (tai onko opetettu) oleellisia asioita? Voiko arviointi integroida opetettuja

asioita? Tällaisiin kysymyksiin on mahdollista etsiä vastauksia erilaisista teorioista ja malleista.

Lähtökohta kielitaidon arvioinnissa on, että arvioijan tulisi tuntea kielitaidon olemus mahdollisimman hyvin, jotta arvioinnin tulokset voitaisiin tulkita järkevällä ja hyödyllisellä tavalla. Arvioinnin perustarkoitus on auttaa ymmärtämään arvioinnin kohteena olevia henkilöitä ja näiden ominaisuuksia (Canale 1988). Arvioinnin tulosten käyttö valintoihin, arvosanojen antamiseen tai johonkin muuhun on vain tämän ymmärryksen soveltamista. Olisi toivottavaa, että käytössämme olisi tarkkaa ja luotettavaa tietoa kielen oppimisesta ja osaamisesta, koska kielitestin laatimisessa on aina kyse yrityksestä operationaalistaa (muuttaa konkreettiseen muotoon) jokin näkemys kielitaidosta. Näkemys voi olla tietoinen tai tiedostamaton, mutta väistämättä arvioijalla on jokin käsitys siitä, mitä kielitaidosta pitäisi arvioida. Mitä paremmin teoria kuvaa kielitaitoa ja kielen käyttöä, sitä helpompi sitä on käyttää apuna arvioinnissa. Kieliteoriat liikkuvat kuitenkin vielä niin yleisellä tasolla, ettei niistä ole testienlaadinnan 'resepteiksi', eikä siis moniin arvioinnin yksityiskohtiin teorioista löydy vastauksia (Huhta 1993b). Teorioista on eniten hyötyä suurissa linjanvedoissa, kun täytyy päättää arvioinnin pääsisällöistä ja keinoista.

Kielitaidon tutkimus ja määrittely on ollut keskeinen osa kielitaidon arviointia jo jonkin aikaa. Vielä 1970-luvun katsaukset kielitaidon arviointiin keskittyivät erilaisiin testaustapoihin (esim. Davies 1978), mutta uudemmat kiinnittävät enemmän huomiota kielitaitoon, mikä heijastaa alan painopisteiden muutosta (esim. Canale 1988; Skehan 1988/89; Bachman 1989; McNamara 1996). Muutoksen tärkein syy lienee kommunikatiivisen kielenopetuksen ja arvioinnin nousu, mikä on kyseenalaistanut aikaisemmat, melko kapealaiset näkemykset kielitaidosta. Halutaan selvittää kielitaidon suhdetta muihin ihmisen kognitiivisiin ja sosiaalisiin taitoihin ja sitä, mitä kielitestit oikein mittaavat.

Teoriat kieli- ja kommunikointitaidoista yrittävät kuvata ja selittää taitoja useasta erilaisesta näkökulmasta. Kognitiiviset teoriat tarkastelevat mm. kielitaidon rakennetta ja kognitiivisia (ajatteluun, muistamiseen, yleensä tiedon käsittelyyn liittyviä) prosesseja. Sosiaaliset teoriat puolestaan tutkivat kielen käyttöä ihmisten välisessä kommunikoinnissa ja mm. kielenkäytön tarkoituksia ja tilanteeseen sopivuutta. Kielitaidon arviointiin eniten vaikuttaneet käsitteet ovat sekoitus näitä näkemyksiä, ja ne voidaan jakaa neljään tai viiteen ryhmään, joita seuraavassa käsittelemme.

1.1 Perinteinen käsitys kielitaidon osataidoista

Vanhin näkemyksistä korostaa kielitaidon rakennetta ja on peräisin strukturalistisen kielitieteen ajoilta, jolloin kieli jaettiin osiin opettamista ja arviointia varten (esim. Lado 1961). Tältä ajalta on peräisin edelleenkin hyvin yleinen jako lukemiseen, kirjoittamiseen, puhumiseen ja kuunteluun sekä edelleen pienempiin tekijöihin kuten kielioppiin ja sanastoon (Bachman & Palmer 1984). 1970- ja 1980-luvuilla kielitaidon rakennetta tutkittiin empiirisesti, koska eräiden tutkimusten mukaan kielitaito muodostaa yhden yhtenäisen kokonaisuuden (ks. tarkemmin Skehan 1988/89; Huhta 1993b). Tutkimusten mukaan ainakin suulliset ja kirjalliset taidot eroavat usein toisistaan, eli ihminen voi olla selvästi parempi joissakin taidoissa kuin toisissa. Syynä eroihin on yleensä erilainen kielenoppimistausta, esimerkiksi formaali kielenopiskelu tai kielen omaksuminen sitä käyttämällä. Suuret erot taitojen hallinnassa ovat tyypillisempiä kielenoppimisen alussa oleville kuin pitemmälle ehtineille.

1.2 Kommunikatiivinen kielitaito

Toinen tärkeä teoreettinen vaikuttaja kielitaidon arvioinnissa ovat olleet ne lukuisat sosiolingvistiset näkemykset, joita 1970-luvulta lähtien on yritetty toteuttaa kommunikatiivisena kielenopetuksena

ja arviointina (ks. Huhta 1993b). Nämä toivat mukaan pyrkimyksen autenttisuuteen, todelliseen kielenkäyttöön (tai ainakin sen simulointiin) ja sosiaalisen kontekstin huomioon ottamiseen (ks. myös Bachman 1988). Seuraava luettelo antaa hyvän kuvan funktionaalisesta, kommunikatiivista kielitaitonäkemyksestä (Morrow 1979):

- kielenkäyttäjien välinen interaktio, vuorovaikutus on keskeistä
- diskurssi- ja sosiokulttuurinen konteksti on tärkeää
- ennakoimattomuus, yllätyksellisyys, luovuus ovat tyypillistä kielenkäytölle
- kielenkäytöllä on jokin tarkoitus
- kielenkäyttöä rajoittavat ja haittaavat monet tekijät (muisti, melu, väsymys, Jane)
- kieli on autenttista
- onnistuminen voi riippua muistakin tekijöistä kuin kielestä
- merkitykset vaihtelevat kontekstista toiseen; merkityksistä on neuvoteltava
- kielenkäytön ja kommunikoinnin onnistumista arvioidaan jatkuvasti
- kielitaidon osa-alueita käytetään usein integroidusti

Myös Chomskyn (1965) esittämä kompetenssi-performanssijako on vaikuttanut arviointiin: testajat ovat perinteisesti olleet kiinnostuneita suorituksen (performanssin) taustalla olevasta 'kyvystä' (kompetenssista), koska se tuntuisi mahdollistavan paremmat yleistyksiset suorituksista. Osa kommunikatiivisista testajista halusi unohtaa koko kompetenssin käsitteen ja pyrki luonnollisten viestintätilanteiden mahdollisimman tarkkaan ulkoiseen simulointiin (ns. performanssitestit; ks. Shohamy 1995).

1.3 Canalen ja Swainin malli kommunikatiivisesta kompetenssista

1980-luvun alussa Canale ja Swain (1980; Canale 1983) laativat yhteenvedon kommunikatiivista kompetenssia käsittelevistä teorioista ja tutkimuksista ja päätyivät malliin, joka on soveltavan kielitieteen tunnetuimpia näkemyksiä kielitaidosta. Erityisen selvästi

mallissa näkyvät Hymesin (1972) ajatukset sosiolingvistisen tietämyksen tärkeydestä kielitaidossa. Mallissa on neljä aluetta eli 'kompetenssia':

1. *Kieliopillinen kompetenssi*: kielikoodin hallinta (esim. sanan ja sananmuodostussäännöt, lauseiden muodostus, sananmukainen merkitys, ääntäminen ja oikeinkirjoitus).
2. *Sosiolingvistinen kompetenssi*: kyky käyttää (ja ymmärtää) tilanteeseen sopivaa kieltä erilaisissa sosiaalisissa tilanteissa; pääpaino on merkityksen ja muodon sopivuudella.
3. *Diskurssikompetenssi*: kyky yhdistellä ja tulkita kielen muotoja ja merkityksiä niin, että niistä muodostuu yhtenäisiä, järkeviä kokonaisuuksia ('tekstejä'), esimerkiksi keskustelu, argumentatiivinen esse tai vaikkapa resepti. Tämä tapahtuu (a) koherenssisääntöjen avulla, joilla organisoidaan asiat loogiseen järjestykseen, ja (b) koheesiokeinojen avulla, joita ovat ne kielelliset ilmaisut, joilla asioiden väliset suhteet ilmaistaan (esim. pronominit, synonyymit, siirtymiä osoittavat sanat, sanojen toisto).
4. *Strateginen kompetenssi*: kyky käyttää kielellisiä ja ei-kielellisiä strategioita (a) korvaamaan puuttuvaa kykyä muissa kompetenssissa (esim. kyky käyttää kiertoilmausta, jos ei muista tiettyä sanaa) ja (b) lisäämään sanottavan retorista vaikutusta (tehokkeinot). (Canale 1983, 339.)

Kielitaidon arvioijat voivat käyttää tällaista mallia keinona hahmottaa arvioinnin suuria linjoja, kuten tärkeimpiä sisältöalueita. Mallin tarkastelu pakottaa arvioijan mm. esittämään itselleen kysymyksen: Ovatko sosiolingvistinen ja diskurssikompetenssi tärkeitä taitoja ja pitäisikö niitä yrittää arvioida? Koska malli kuvaa kielitaitoa melko yleisellä tasolla, siitä on kielitaidon arvioijien piirissä kehitetty arviointiin paremmin soveltuvaa versiota.

1.4 Bachmanin ja Palmerin malli kielitaidosta

Nykyään kielitaidon arvioinnissa käytetään eniten Bachmanin ja Palmerin (1984; 1996; Bachman 1990; 1991) mallia kielitaidosta, joka perustuu Canalen ja Swainin jaotteluun, mutta on yksityiskohdallisempi ja monipuolisempi. Sen avulla yritetään kuvata paitsi

kielen rakennetta myös kompetenssin toteutumista kielen käyttönä, performanssina. Malli nojaa myös perinteiseen mittaamisteoriaan, joka selvittää testituloksiin vaikuttavia tekijöitä (vrt. nykyiset näkemykset älykkyydestä esim. Sternberg 1990). Näistä syistä mallia voidaan helpommin käyttää arvioinnin suunnitteluun kuin kieliteorioita yleensä (Skehan 1988/89).

Bachman (1990) määrittelee kielitaidon kyvyksi käyttää kielellistä tietoa viestintätilanteiden vaatimusten mukaisesti tarkoituksena luoda ja tulkita merkityksiä. Malliin kuuluu kaksi komponenttia: kielitieto ja metakognitiiviset strategiat (strateginen kompetenssi). Kielitietoon kuuluvat ne tiedon alueet, jotka ovat tyypillisiä vain kielen käytölle ja jotka eroavat esimerkiksi matemaattisesta tai musikaalisesta 'tiedosta'. Metakognitiiviset strategiat ovat yleisempiä kognitiivisia prosesseja, joiden avulla kielitieto muutetaan kielen käytöksi.

Bachmanin malli kielitaidon alueista koostuu seuraavista tekijöistä:

- A. Kielitietokomponentti (language knowledge):
 - 1. Organisatorinen tieto
 - a) Kieliopillinen tieto
 - b) Tekstuaalinen tieto
 - 2. Pragmaattinen tieto
 - c) Propositionaalinen tieto
 - d) Funktionaalinen tieto
 - e) Sosiolingvistinen tieto
- B. Strateginen komponentti (strateginen kompetenssi)

Mallin ensimmäinen komponentti eli *kielitieto* on Bachmanin (1990, 1991) mukaan tietoa, joka on ominaista lähinnä vain kielelle ja sen käytölle. On tärkeitä muistaa, etteivät mallin osat ole itsenäisiä, toisistaan riippumattomia kokonaisuuksia, vaan kielen käytössä ne ovat jatkuvassa vuorovaikutuksessa paitsi keskenään myös kielen käytön kontekstin, tilanteen kanssa. Lisäksi kielen oppimisen myötä eri osa-alueet muuttavat jatkuvasti muotoaan.

Organisatorisen tiedon ensimmäinen osa, *kieliopillinen* tieto, kattaa morfologian, syntaksin, fonologian ja kirjaimiston, joiden avulla kielen käyttäjä muodostaa yksittäiset lauseet ja puhunnokset. *Tekstuaalinen* tieto koskee sitä, miten lauseet tai puhunnokset liitetään toisiinsa teksteiksi ja diskursseiksi. Tämä liittyy koherenssiin, koheesioon ja tekstien retoriseen organisointiin (vrt. diskurssikompetenssi Canalen mallissa).

Toinen laaja kielitiedon alue mallissa on *pragmaattinen tieto*, jossa on kyse kielen käyttöön ja tulkintaan liittyvien sääntöjen hallinnasta. *Propositionaalisen* (lähinnä sanojen merkityksiä koskevan) tiedon avulla annetaan lauseille ja puhunnoksille tilanteeseen sopiva merkitys. *Funktionaalista* tietoa on puolestaan kyky ymmärtää ja tuottaa ilmauksia, joilla on tietty tarkoitus, kuten tiedon ja tunteiden välittäminen tai vaikuttaminen esimerkiksi käskyjen ja lupauksen avulla (Bachman, 1990; ks. myös Euroopan neuvoston 1997 funktiolistat).

Pragmaattista tietoa on myös *sosiolingvistinen* tieto: taito käyttää tilanteeseen sopivaa kieltä. Tätä on kyky ymmärtää ja käyttää erilaisia kielen *variantteja* (maantieteellisiä tai sosiaalisia) ja kyky hallita erilaisia *rekistereitä*. Kolmas sosiolingvistisen tiedon aspekti on kyky tulkita *kulttuurisidonnaisia viittauksia ja sanontoja* eli taito ymmärtää vertauksia, metaforia ja myös niitä laajempia merkityksiä, joita kulttuuri antaa tietyille tapahtumille, paikoille, instituutioille tai henkilöille (esim. Arkadian mäki, Uno Turhapuro, Waterloo). Sosiolingvististä tietoa on myös kyky käyttää ilmauksia, jotka eivät pelkästään ole kieliopillisesti oikeita, mutta sen lisäksi *idiomaattisia*, 'luonnollisia' kyseiselle kielelle. Ajatus on lähellä Hymesin (1972) näkemystä, että oleellinen osa kielitaitoa on käsitys siitä, kuinka todennäköisesti tiettyjä ilmauksia käytetään, joko yleensä tai tietyssä tilanteessa (ks. tarkemmin Huhta 1993b).

Bachmanin (1990, 1991) mukaan kielen käyttötaitoon kuuluu kielitiedon lisäksi toinen olennainen komponentti, *strateginen kompetenssi*. Toisin kuin Canale ja Swain (1980) Bachman ei pidä tätä

kiellellisenä kykynä vaan yleisempänä kognitiivisena ominaisuutena, josta kuitenkin on paljon apua kielen käytössä. Sen avulla paikataan puutteita esimerkiksi kielellisissä kompetensseissa ja huolehditaan niiden tarkoituksenmukaisesta käytöstä.

Edellä esitettyjen mallien lisäksi on myös Euroopan neuvostossa tehty kauan työtä kielitaidon määrittelyä (mm. Threshold- eli kynnystasojulkaisut), mitä on käytetty laajalti hyväksi kielenopetuksessa ja nyttemmin myös arvioinnissa (Euroopan Neuvosto 1997).

Kun teorioita kielitaidosta tarkastelee kokonaisuutena, kielitaidosta hahmottuu selvästi hyvin moniulotteinen kuva, millä on selviä käytännön vaikutuksia arviointiin. Näyttää epätodennäköiseltä, että kielitaidosta voisi saada luotettavan ja kattavan kuvan mittaamalla vain yhtä kielitaidon puolta tai käyttämällä vain yhtä lyhyehköä testiä. (Tietysti voi olla tilanteita, joissa riittää karkea, osittainen ja epäluotettavampi kuva kielitaidosta, esimerkiksi valinta eritasoisille kielikursseille, jos väärät valinnat voidaan helposti korjata siirtämällä oppija toiseen, sopivampaan ryhmään.)

Käsitys kielitaidosta on huomattavasti laajentunut viime vuosikymmeninä. Yhteistä kaikille uusimmille malleille on sosiolingvistisen puolen korostuminen: taito käyttää kieltä tilanteeseen sopivalla tavalla katsotaan oleelliseksi osaksi kielitaitoa. Tällaisia taitoja mittaavien arviointivälineiden kehittäminen on vielä lapsenkengissä, mutta vaikeuksista huolimatta sosiolingvistisia ja yleensäkin pragmaattisia seikkoja ei tule mielestämme unohtaa kielitaitoa arvioitaessa.

Lopuksi on vielä todettava, ettei kielitaidon arvioinnissa yleensä mitata kaikkia kielenopetuksen tavoitteita. Kielitaito ei yleensä ole ainoa opetuksen tavoite virallisten opetussuunnitelmienkaan mukaan, vaan tavoitteena voi olla mm. myönteinen suhtautuminen opiskeltavaan kieleen, sen puhujiin ja heidän kulttuuriinsa, omaehtoinen, motivoitu ja päämäärätietoinen kielenopiskelu sekä kansainvälisyys. Useimmiten näiden tavoitteiden saavuttamista on hyvin

vaikea testata, vaikka tietyt arvioinnin muodoissa, kuten jatkuvassa arvioinnissa, voidaan yrittää ottaa niitä huomioon. Kuten muissakin oppiaineissa, arvosanat ja todistukset kielen hallinnasta kertovat siis vain osan opetuksen tavoitteiden saavuttamisesta ja kieltenopetuksen vaikutuksista (ks. Cohenin 1990 kriittinen arvio).

2 MITÄ TARKOITUKSIA VARTEN KIELITAITOA ARVIOIDAAN?

Tärkeimmät päätökset, joita kielitaidon arvioija joutuu tekemään, koskevat arvioinnin kohdetta ja arvioinnin tarkoitusta. Arvioijan tulisi tuntea arvioinnin kohde, kielitaito, mahdollisimman hyvin, muuten tulosten tulkinta on vaikeaa. Arvioinnin tarkoituksen ymmärtäminen on aivan yhtä oleellista: yhteen tarkoitukseen hankittu ja luotettavaksi todettu tieto oppijoiden kielitaidosta ei välttämättä sovi käytettäväksi muihin tarkoituksiin.

Arvioinnin tarkoituksia on mahdotonta ryhmitellä yhdellä yleisesti hyväksytyllä tavalla, koska arviointityypit eivät ole aina selvärajaisia. Taulukon 1 ryhmittely perustuu Bloom ym. (1971) jaotteluun, ja se kuvaa arviointia yleensä eikä pelkästään kielitaidon arviointia. Seuraavassa käymme taulukon läpi pääpiirteittäin käsittelemättä kuitenkaan jokaista yksityiskohtaa. Arviointi voidaan jakaa kahteen päätyyppiin (taitotaso vs. opetukseen liittyvä) lähinnä sen mukaan, millä perusteilla arvioinnin kohde, sisältö määräytyy. Kahtiajako ei suinkaan ole täysin vedenpitävä: esimerkiksi testeissä, joista saadaan tutkintotodistuksia, yhdistyvät usein molempien piirteet.

Ensimmäisessä arviointityypissä on kyse *taitotason* (proficiency) arvioinnista, ja siinä halutaan saada selville oppijan senhetkinen, absoluuttinen *osaamisen taso* ja usein samalla *taitojen riittävyys* johonkin tiettyyn tarkoitukseen. Tällöin ei olla kiinnostuneita siitä, millä tavalla osaaminen on hankittu. Testien sisältö ei siis määräydy

suoraan niitä edeltäneen opetuksen perusteella, vaan perustuu yleisempiin seikkoihin, kuten teorioihin kielitaidosta tai käsityksiin siitä, millaista kielitaitoa tarvitaan tiettyjen tehtävien suorittamiseen. Toinen taulukossa esiteltävä arviointityyppi on nimetty *opetuksen liittyväksi arvioinniksi*. Siinä arvioinnin kohteet määräytyvät pääasiassa meneillään olevan tai juuri päättyneen *opetuksen sisällöstä* käsin ja myös opetuksen *tavoitteista* käsin (nämä eivät aina ole sama asia). Toki opetuksen sisältö ja tavoitteet voivat alunperin pohjautua teorioille tai tarveanalyysille, mutta tässä arviointityypissä korostuu arvioinnin välitön yhteys opetukseen.

On myös hyödyllistä ajatella arviointityyppejä kimppuina erilaisia piirteitä: tyypit eroavat toisistaan joidenkin piirteiden suhteen, mutta voivat muistuttaa toisiaan muilta ominaisuuksiltaan. Taulukon rivit kuvaavat näitä erottavia ja yhdistäviä piirteitä, esimerkiksi arvioinnin päätarkoitusta, suorittajia, ajankohtaa ja välineitä. Seuraavassa käsittelemme arvioinnin päätyyppejä tarkemmin.

TAULUKKO 1. Kielitaidon arvioinnin tarkoitukset

ARVIOINNIN TYYPI

TAITOTASON JA TAITOJEN RIITTÄVYYDEN ARVIOINTI

	KIELELLISEN LAHJAKKUIDEN ARVIOINTI	VALITSEMISEN JA SIJOITAMINEN	DIAGNOSOINTI	PÄTEYTYSKELPOISUUS- TAI TULKINTO- TAITOKUVAAMINEN	TUTKIMUS	FORMATIIVISEN TIEDON HANKKIMINEN	SUMMAATIIVISEN TIEDON HANKKIMINEN (OPPISAAVUSTUSTEN ARVIOINTI)	OPPILAITOSTEN JA OPETUSOHJELMIEN ARVIOINTI
ARVIOINNIN FUNKTION/TÄRKEYS	Kieliohjelmien toteutumisen ja oppilaiden osaamisen arvioiminen.	Välitsemisen ja sijoittamisen taitojen arvioiminen.	Oppimisympäristöjen ja oppimistilanteiden arvioiminen.	Oppilaiden osaamisen arvioiminen.	Laaja valtakunnallinen tutkimus kielten opettajien ja oppilaiden taitojen arvioimiseksi.	Palautteen antaminen oppijalle ja opettajalle siitä, miten oppija edistyy.	Todistusten tai arvosanojen antaminen kurssin, lukukauden lukuvoimien tai yleensä opinnoitten lopussa.	Oppilaitoksen tai tietyn opetusohjelman tehokkuuden arviointi.
ARVIOINNIN SUORITTAJA	Yleensä opettajan ja oppilaiden yhteistyö (tai rahoittaja).	Välitteistä vastaus- tai kysymys- tai oppimisympäristöjen arvioimiseksi.	Yleensä opettaja (oma tai tukeva opetusantaja).	Yleensä itsekkinen testi, voi olla myös oppilaitosta tai opettajasta.	Kansallinen tai kansallinen tutkimus, johon osallistuu opettajia ja oppilaita.	Yleensä opettaja, mahdollisesti myös oppija.	Opettaja tai ryhmä opettajia ja oppilaita (tai koulun tasolla).	Yleensä kansalliset opetusviranomaiset ja / tai muut rahoittajat.
ARVIOINNIN AJATUS	Ennen opinnoitten alkua.	Opintojen / työn alussa tai ennen niiden alkamista.	Ennen opetuksen oppimisen alkamista.	Periaatteissa milloin vain, kun oppijalla on tarpeeksi taitoa suorittaa.	Kansalliset selvitykset, muutamien voimassa olevien tutkimusten ja kyselyjen avulla.	Opetuksen laajuus.	Yleensä kurssin, lukukauden, lukuvoimien tai yleensä opinnoitten lopussa.	Esim. kerran vuodessa tai useammin vuodessa tai koulun kuluessa uuden (laajan) opetusohjelman alottamisesta.
PAINOPISTE ARVIOINNIN KOKONAISUUS	Kognitiiviset taidot, myös persoonallisuus, tunteet ja sosiaaliset taidot.	Kaikki opinnoissa / työssä menestykseen tarvittavat taidot.	Kognitiiviset, affektiiviset ja psyykkiset taidot.	Kognitiiviset taidot.	Kognitiiviset taidot, myös selviytymiskyky, kognitiiviset taidot, kognitiiviset taidot.	Kognitiiviset taidot.	Yleensä kognitiiviset taidot, myös persoonallisuus, tunteet, esim. kouluyhteydet.	Opetusohjelman keskeiset tavoitteet.
ARVIOINNIN MENETELMÄT (INSTRUMENTIT)	Ennen taitotason standardoitu testi.	Yleensä erikseen tarkoitusta varten laadittu testi (oppilaitos) tai yleisemmät testit (tutkimuslaitokset).	Standardoidut testit.	Yleisemmät testit ja tulokset.	Ennen taitotason standardoitu testi.	Ennen taitotason standardoitu testi.	Yleensä erikseen tarkoitusta varten laadittu testi.	Eriksen tarkoitusta varten laadittu testi.

Institutin testit ovat näyttökokeita, jotka eivät perustu minkään tietyn kielikoulun opinto-ohjelmaan, vaan ne testaavat yleensä näiden kielialueiden oppilaitoksissa vaadittavaa kielitaitoa.

Tätä tyyppiä edustavat myös sellaiset kokeet, joiden läpäisy osoittaa kelpoisuuden johonkin ammattiin tai virkaan. Monissa maissa eri ammatinharjoittajat joutuvat osoittamaan pätevyytensä kelpoisuuskokeiden avulla. Myös kielellisen pätevyyden voi joutua osoittamaan testin avulla, esimerkkeinä Suomessa valtion viranomaisilta vaadittava ns. pieni ja suuri kielitutkinto. On myös ammattiteja, kuten kielenopettaja, kääntäjä tai tulkki, joissa kielitaito on osa ammattitaitoa, joka täytyy osoittaa ammattipätevyuden saamiseksi.

Ero näyttökokeiden ja summatiivisten, oppimäärään perustuvien loppukokeiden välillä ei aina ole kovin selvä. Näyttökoe mittaa, millaista kielitaito on ja mitä kielen avulla osaa tehdä, esimerkiksi työmatkalla ulkomailla. Summatiivinenkin koe voi mitata oppimäärää, joka liittyy samaan aihepiiriin (työmatka ulkomaille), mutta mikäli oppimäärä tai -materiaalit eivät kata koko aihealuetta tai eivät ole sopivan tasoisia, hyväkään oppimäärän hallinta ei vielä kerro riittäkö oppijan kielitaito kyseiseen tarkoitukseen. Mutta on toki mahdollista, että summatiivinen loppukoe on käytännössä samanlainen kuin taitotasokoe. Mikäli oppimäärä, materiaali, kielitaidon taso, opetus ja opetuksen tavoitteet perustuvat realistiseen näkemykseen siitä, millaista kielitaitoa kohdetilanteissa tarvitaan, niin summatiivinen loppukoe mittaa aivan samoja asioita kuin vastaava näyttökoe.

Monissa tutkinnoissa yhdistyvät taitotasokokeen (näyttökokeen) ja summatiivisen saavutuskokeen ominaisuudet. Esimerkiksi Suomen ylioppilastutkinto vaikuttaa ensi näkemältä summatiiviselta kokeelta, koska se pidetään koulun loppuvaiheessa ja koska sen sisällön tulisi pitkälti perustua koulun oppimäärään. Kuitenkin kielten oppimäärät ovat varsinkin 'pitkissä' A-kielissä niin laajat, että niitä mittaava saavutuskoe alkaa käytännössä muistuttaa näyttökoea. Nykyisin koulut voivat lisäksi päättää itse opetuksestaan ja

Jatkotoiminta ei tietenkään riipu vain opettajasta, vaan yhä voimakkaammin on alettu korostaa oppijoiden omaa, suorastaan ratkaisevaa panosta oppimisen edistymisessä. Näin ollen diagnoositiedon käyttäjä on yhä useammin oppija/koulutettava/asiakas eikä pelkästään opettaja/kouluttaja. Diagnoositiedon kohteita ei tällöin voida sanella ulkopuolelta, ja siksi tiedon tarve voi olla monenlaista. Oppijat ovat usein kiinnostuneita saamaan ulkopuolisen arvion kielitaitonsa yleistasosta, mutta autonomisen oppimisen kannalta on järkevää pyrkiä hankkimaan myös yksityiskohtaisempaa palautetta, jonka avulla kielenkäyttäjä voi tehokkaammin ohjata oppimistaan.

Tutkimus

Kielitaitoa arvioidaan myös tutkimusmielessä. Laajat kansalliset ja kansainväliset tutkimukset pyrkivät yleensä selvittämään maan tai maiden kielitaidon tasoa ja siihen vaikuttavia tekijöitä (esim. opetusresurssit, opettajien koulutus, materiaalit). Usein pyritään hankkimaan myös tarkempaa tietoa kielitaidosta tai seurataan sen kehittymistä muutaman vuoden välein tehtävillä seurantatutkimuksilla. Tällaista tietoa tarvitaan mm. perustellun koulutuspolitiikan sekä -suunnittelun perustaksi. Myös yksittäiset tutkijat ja opinnäytteiden tekijät voivat käyttää arviointimenetelmiä tutkiessaan kieltä, kielitaitoa ja kielenoppimista. Onnistunut kielitaidon mittaaminen on näissä kaikissa avainasemassa, sillä tutkimusten tulokset ovat juuri niin luotettavia kuin ne välineet, joilla ne on hankittu. Tietty kielitaidon arvioinnin tuntemus on siis välttämätön osa monen soveltavan kielitieteilijän ammattitaitoa.

Suomessa kansallista arviointitutkimusta on tehnyt pääasiassa Jyväskylän yliopistossa sijaitseva Koulutuksen tutkimuslaitos (entinen Kasvatustieteiden tutkimuslaitos, KTL). Se on järjestänyt peruskoulussa muutaman vuoden välein englannin ja ruotsin kielen osaamista selvittäviä tutkimuksia, joissa on mm. todettu englannin kuulun ymmärtämisen huomattava kehittyminen 1970-luvulta lähtien (Havola & Saari 1993; Saari 1996; Takala 1998).

Kansainväliset kielitaidon tutkimukset ovat erittäin harvinaisia. IEA-järjestön (International Association for the Evaluation of Educational Achievement) tutkimus 1970-luvun alussa on harvinainen poikkeus, johon Suomikin osallistui (Lewis & Massad 1975; Takala & Saari 1979). Toinen tutkimus aloitettiin 1994, mutta se jouduttiin keskeyttämään rahoituksen puutteessa. Vaikka kielitaidon tasoa ei päästykään arvioimaan, koottiin kuitenkin tietoa kielenopetuksen ja arvioinnin järjestelyistä eri maiden koululaitoksissa (Dickson & Cumming 1996; Huhta ym., tulossa). IEA on tutkinut myös äidinkielen kirjoitustaitoa (Purves ym. 1988; Purves 1992) ja lukutaitoa (Linnakylä 1995). Näiden tutkimusten raportit ovat hyödyllistä luettavaa myös vieraan kielen arvioijille, koska useimmat ongelmat ja tehtävien laadintaperiaatteet ovat yhteisiä äidinkielessä ja vieraissa kielissä (esim. Lehmann 1992, 1993). Nykyisin OECD-järjestö on ottamassa IEA:n roolia kansainvälisten tutkimusten tekijänä esimerkiksi äidinkielessä, mutta toistaiseksi se ei ole ollut kiinnostunut vieraista kielistä.

2.2 Opetukseen liittyvä arviointi

Formatiivinen arviointi

Opetusta suoraan tukeva formatiivinen arviointi on luultavasti yleisin arvioinnin muodoista. Siinä hankitaan tietoa siitä, miten hyvin oppija hallitsee opetetut asiat ja missä on vielä puutteita. Useimmiten arvioinnin suorittaja ja arviointitiedon ensisijainen käyttäjä on opettaja, joka sen perusteella pystyy suuntaamaan opetustaan havaittuihin puutteisiin tai eriyttämään opetusta oppilaiden taitojen mukaan. Oppilaan rooli on yleensä yrittää hyödyntää opettajan antamaa palautetta; kuitenkin, jos käytettävissä on esim. itseopiskelutehtäviä, voi oppija hankkia itsenäisestikin oppimista tukevaa tietoa osaamisestaan.

Formatiivista tietoa voidaan hankkia hyvin monenlaisin menetelmin, joista luultavasti yleisin on oppilaiden toiminnan jatkuva

epämuodollinen observointi. Arvioija on yleensä opettaja, mutta myös oppijan oma itsearviointi soveltuu erinomaisesti tähän tarkoitukseen (Takala 1995, 1997). Arvioinnin välineitä voivat olla myös erilaiset tunnilla ja kotona suoritettavat harjoitustehtävät, jopa testitkin. Formatiiviseen arviointiin soveltuu hyvin myös *portfolioarviointi*, jossa pitkäkhön ajan kuluessa kerätään näytteitä oppijan osaamisen eri puolista ja vaiheista (Linnakylä ym. 1994; Pollari ym. 1996).

Formatiivista tietoa voi saada myös summatiivisista kokeista. Onkin varsin houkuttelevaa yrittää lyödä kaksi kärpystä yhdellä iskulla ja hankkia samasta testistä tietoa sekä opetuksen tueksi että arvosanojen antamiseksi. Tämä voi onnistua, mutta asiaan liittyviä ongelmia, joista on hyvä olla selvillä. Formatiivisen tiedon hyödyntäminen voi olla vaillinaista, mikäli opetus joko loppuu kokonaan summatiivisen kokeen jälkeen (yksittäisen kielikurssin loppukoe) tai siirtyy osittain muihin sisältöihin ja taitoihin. Useampaan kuin yhteen tarkoitukseen sopivan kokeen tulee yleensä olla laajempi ja pitempi kuin yhtä asiaa varten laadittu testi. Oppilaissa voi myös herättää epävarmuutta se, mikä kokeen 'todellinen' tarkoitus on. Perinteisesti arviointi samaistetaan niin vahvasti arvosanojen antamiseen, että opettaja ei ehkä saakaan kaikkea haluamaansa formatiivista tietoa (vaikeuksista käyttää arvioinnin tuloksia yhtä aikaa useampaan eri tarkoitukseen ks. Bloom ym. 1971: 61, Brindley 1998).

Ongelmista huolimatta arvioinnin tehokkuutta lisäisi, jos summatiivisen arvioinnin antamia tietoja käytettäisiin formatiivisesti myös laajemmin kuin yksittäisen luokan tasolla. Esimerkkinä voisi mainita ylioppilastutkinnon. Toistaiseksi formatiivinen palaute tutkinnosta on ollut hyvin vähäistä, mutta mitään periaatteellista estettä tämän puolen kehittämiseksi ei ole. Päinvastoin tutkinto tarjoisi mahdollisuuden jatkuvaan ja hyödylliseen palautteeseen sekä opettajille että oppilaille siitä, missä asioissa osaaminen on vankkaa ja missä on vielä kehittämisen varaa.

Summatiivinen arviointi

Summatiivisen arvioinnin tarkoitus on antaa arvosanoja sen perusteella, miten kurssilla tai tietynä lukukautena opetetut asiat hallitaan. Nimitystä 'saavutustestaus' (achievement testing) käytetään usein melkein samassa merkityksessä. Summatiivinen arviointi tapahtuu hyvin yleisesti testien avulla. Usein niiden tukena käytetään muitakin menetelmiä, kuten jatkuvaa arviointia, mutta lienee tavallista, että testin tulokset painavat enemmän.

Monissa maissa summatiivinen arviointi on täysin opettajien ja koulujen vastuulla, ilman tarkkoja kansallisia ohjeita tai valtakunnallisia testejä. Näin on pääpiirteittäin Suomessa ja muissa Pohjoismaissa, jos ei oteta huomioon ylioppilastutkinnon kaltaisia poikkeuksia. Toisaalla kouluja koskevissa laeissa määrätään yksityiskohtaisesti arvosanojen antamisen ja testauksen ajankohdasta, keinoista tai sisällöistä. Joissakin maissa summatiiviseen arviointiin käytetään runsaasti ulkopuolisia valtakunnallisia testejä (esim. viime aikoina Englannissa; ks. tarkemmin Huhta ym., tulossa).

Valtakunnallisten testien käyttö summatiiviseen loppuarviointiin vaihtelee kovasti eri maissa. Suomessa ylioppilaskokeesta annetaan oma todistuksensa, mutta lukiolaiset saavat myös päästötodistuksen, josta opettajat ovat yksin vastuussa. Ruotsissa kansallisia kokeita käytetään opettajan antamien arvosanojen tukena (erillistä todistusta ei niistä anneta), ja opettaja voi (mutta hänen ei ole pakko) muuttaa antamia arvosanoja ulkopuolisen testin perusteella. Englannissa taas summatiiviset loppuarvosanat perustuvat täysin ulkopuolisiin testeihin (Huhta ym., tulossa). Erot maiden välillä johtuvat usein perinteistä ja koulutusjärjestelmän yleisestä luonteesta, esimerkiksi siitä onko päätöksenteko keskitetty (tällöin käytetään usein keskitettyä valtakunnallista testausta; vrt. Englanti), tai yleisestä luottamuksesta opettajien ammattitaitoon, myös arvioinnin alalla (mitä suurempi luottamus, sitä vähemmän tarvetta keskitettyyn testaukseen). Myös aika ajoin esille nousevat vaatimukset oppilaitosten ja koulutusjärjestelmän 'tulosvastuusta' voivat johtaa

keskitetyn testauksen yleistymiseen. Joissakin maissa käytänteet voivat myös riippua siitä, mikä poliittinen puolue on kulloinkin vallassa (Brindley 1997).

Oppilaitosten ja opetusohjelmien arviointi

Koulutuksesta päättävät viranomaiset ja poliitikot, oppilaiden vanhemmat ja oppilaat itse ovat kautta aikojen olleet kiinnostuneita opetuksen laadusta (vrt. vaatimukset 'tulosvastuusta' edellä). Vaatimus saada luotettavaa tietoa koululaitoksen ja yksittäisten koulujen tuloksellisuudesta on epäilemättä oikeutettu, vaikka tällaisen tiedon tärkeys näyttääkin vaihtelevan eri maissa. Tämä on johtanut opetuksen tuloksellisuuden arviointiin eri tavoin, usein opetusviranomaisten ja tutkimuslaitosten laatimien testien avulla.

Vertaileminen on kuitenkin hyvin vaikeaa, ja vain harvoin onnistutaan keräämään sellaista vertailutietoa, joka todella kertoo koulun opetuksen tai oppimäärän laadusta ja tehokkuudesta. Koulujen asettaminen paremmuusjärjestykseen ylioppilastutkinnon kaltaisten summatiiivisten kokeiden perusteella on yleensä täysin perusteetonta, vaikka aiheesta lehdet saisivatkin myyviä otsikoita. Tutkiminto on laadittu toisiin tarkoituksiin, joten sen tulokset eivät välttämättä sovellu tällaisiin johtopäätöksiin. Koulujen vertaileminen on oikeutettua ja kiinnostavaa silloin, kun koulujen tuloksia vertaillaan esimerkiksi järkevään odotusarvoon, joka ottaa huomioon oppilasaineuksen koostumuksen ja koulun resurssit. Vertailun pohjaksi tarvitaan siis paljon muutakin tietoa kuin pelkät testitulokset. Vain näin voidaan saada hyödyllistä tietoa 'oikeasta' tehokkuudesta ja voidaan yrittää selvittää siihen vaikuttavia syitä, sekä kouluissa itsessään että kouluhallinnossa. Yksittäisten opettajien arvioimista oppilaiden testisuoritusten perusteella pidetään hyvin epäluotettavana (esim. Shepard 1992 tuomitsee sen kategorisesti), koska testituloksiin vaikuttavat monet muutkin tekijät kuin opettajan opetus- taidot.

Suomalaisena esimerkkinä kielikoulutuksen arvioinnista voidaan mainita Soveltavan kielentutkimuksen laitoksen projekti, jossa on arvioitu maamme kielikoulutuksen tehokkuutta kokonaisvaltaisesti (Takala & Sajavaara 1998).

Taulukossa 1 esitetty jaottelu arvioinnin päämääristä antaa kuvan kielitaidon arvioinnin monimuotoisuudesta. Kyse ei siis suinkaan ole yhtenäisestä ilmiöstä. Valitsemamme luokittelu ei toki ole ainoa mahdollinen, mutta melko yksityiskohtaisena se sopii tähän yhteyteen. Arviointia voidaan hahmottaa myös yleisemmin tarkastelemalla eri arviointimalleja (esim. tavoite- ja päätöksentekomallit; ks. tarkemmin Laukkasen 1994 yhteenveto). Toinen täydentävä näkökulma on tarkastella arviointia toisaalta yksilön ja toisaalta erilaisten ryhmien tasolla (opettaja, perhe, kansakunta) ja samalla kiinnittää huomiota arvioinnin myönteisiin ja kielteisiin seurauksiin (Little 1990).

3 MITÄ VAIHEITA KUULUU KIELITAIDON ARVIOINTIIN?

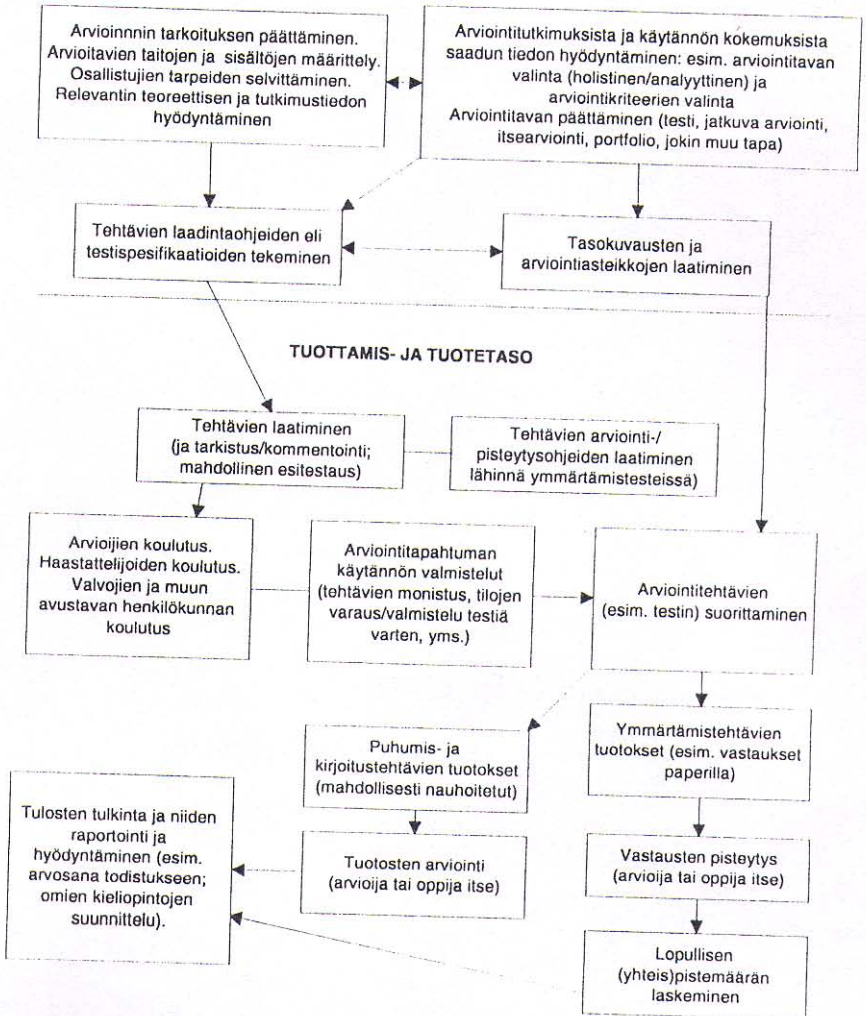
Kielitaidon arvioinnissa on useita vaiheita, joilla kaikilla on oma vaikutuksensa arvioinnin lopputulokseen ja laatuun. Aina kaikki osatekijät eivät ole yhtä tärkeitä: suuressa ja vaikutuksiltaan merkittävässä kielitestissä on paljon sellaista, mitä ei pienimuotoisemmassa luokkahuonetestauksessa ole mahdollista eikä tarpeenkaan toteuttaa. Julkisessa keskustelussa korostuu usein testien 'subjektiivisuus' tai 'objektiivisuus', missä on yleensä kyse vain yhdestä testausprosessin vaiheesta, tulosten arvioinnista. On tärkeää muistaa, että arvioinnin kokonaislaatu riippuu tämän lisäksi myös siitä, millaisia valintoja on tehty arvioinnin suunnittelussa.

Kaavio 1 kuvaa arviointiprosessin keskeisimmät vaiheet, ja se perustuu kirjoittajien aikaisempaan työhön (Takala 1988; Huhta 1993a, 1997). Tässä ei ole mahdollista esittää yksityiskohtaista

kuvausta arvioinnin suunnittelusta ja testien laadinnasta. Saatavilla on havainnollisia testinlaadinnan oppaita, mutta ei kuitenkaan vielä suomeksi. Takalan 1970-luvun suomenkieliset julkaisut (1973, 1975) ovat suppeita esityksiä, jotka myös heijastavat aikakauden painotuksia. Huhta (1993a) ja Huhta ja Suontausta (1993) käsittelevät vain suullisen kielitaidon arviointia. Takala (1997) ja Linnakylä ym. (1994) kuvaavat arvioinnin uusia suuntia yleisemmällä tasolla. Englanniksi on kuitenkin julkaistu kattavia oppaita, joista parhaita ovat Hughes (1989), Weir (1990, 1993), Cohen (1994), Alderson ym. (1995) ja Bachman ja Palmer (1996). Mitään testien laadinnan 'reseptejä' eivät nämäkään ole, eikä sellaisia voida laatiakaan, sillä arvioinnin konteksti määrää pitkälti sen, mitä kulloinkin on järkevää tehdä.

Kaavion 1 ylin taso (käsitteellinen taso) esittää tehtävien laadintaa edeltävät vaiheet testausprosessissa. Niissä määritellään testin tarkoitus, sisällöt ja arvioitavat ominaisuudet, minkä pohjalta valmistetaan tehtävien laatimisohteet (ns. testispesifikaatiot). Suurissa testeissä nämä ovat aikaa vieviä vaiheita, ja niiden tuloksena on joukko kirjallisia dokumentteja tehtävien laatijoiden avuksi. Pienimuotoisemmassa testauksessa arvioinnin kohteen määrittelyn ei tietenkään tarvitse olla näin tarkkaa ja eksplisiittistä. On kuitenkin hyvä muistaa, että jokainen testin laatija käy tämän vaiheen läpi enemmän tai vähemmän tietoisesti, olipa kyse miten epämuodollisesta arvioinnista tahansa. Arviointi ilman käsitystä siitä, mitä ollaan arvioimassa on hukkaan heitettyä aikaa, pahimmillaan jopa oppimista haittaavaa toimintaa. Vaikka kaikissa tilanteissa ei ole järkevää laatia yksityiskohtaisia ohjeita arvioinnista, on kuitenkin hyödyllistä miettiä arvioinnin tavoitteita ja sisältöjä samoin kuin sitä, millaista käsitystä kielitaidosta ne heijastavat. Vielä hyödyllisempää on keskustella omista käsityksistä toisten, esimerkiksi opettajakollegoiden kanssa, vaikei opetettava kieli (ehkei oppiaineeseen) olisi yhteinen.

KÄSITTEELLINEN TASO



KAAVIO 1. Kielitaidon arvioinnin vaiheet

Toinen taso arviointiprosessissa koskee konkreettista toimintaa ja tuotteita ja sitä voidaan kutsua 'tuottamistasoksi'. Siihen kuuluvat:

- koulutustilaisuudet testien laatijoille, arvioijille ja muille testin järjestämiseen osallistuville henkilöille (esim. haastattelijat, testien valvojat),
- testit tehtävien laatiminen kutakin testauskertaa varten,
- arviointiohjeiden laatiminen tehtäviin, ja
- testisuoritusten arviointi tai pisteytys.

Nämä vaiheet kuvaavat toimintaa, joka tapahtuu arvioinnin suunnittelun tai testin suorittamisen jälkeen. Myös tällä tasolla 'suuret' ja 'pienet' testit eroavat toisistaan. Isoissa testijärjestelmissä tarvitaan paljon arvioijia ja muuta henkilökuntaa, jonka koulutukseen täytyy uhrata tarpeeksi aikaa. Monesti testitehtävät esitetaan eli pilotoidaan niiden toimivuuden tarkistamiseksi. Yhden opettajan 'projektissa' tämä kaikki ei tietenkään ole tarpeen: hän on itse kokeen laatija, valvoja ja arvioija. Oman toiminnan monitorointi etenkin suoritusten arviointivaiheessa on tietenkin järkevää, jotta kaikkien oppijoiden arviointi tapahtuisi samoin perustein. Mikäli opettaja käyttää samoja tehtäviä useamman kerran, on hänenkin toki mahdollista ja järkevää korjata niissä ilmenneitä puutteita ennen seuraavaa käyttöä eli toteuttaa käytännössä se, mitä isommissa testeissä kutsuttaisiin pilotoinniksi.

3.1 Eri taitoja arvioidaan eri tavalla

Vaikka jako lukemiseen, kirjoittamiseen, kuunteluun ja puhumiseen ei kuvaa kielitaitoa läheskään tyydyttävästi (ks. esim. Bachmanin 1990 malli edellä), käytetään sitä edelleen yleisesti testinlaadinnan perustana. Uudemmat näkemykset mm. sosiopragmaattisten taitojen tärkeydestä yritetään yleensä ottaa huomioon laatimalla tehtäviä, joista suoriutuminen vaatii tällaisten seikkojen hallintaa tai lisäämällä arviointiasteikkoihin sellaisia piirteitä kuin 'kielenkäytön sopivuus'. Usein testien perusrakennetta ei silti muuteta, vaan edel-

leen tietyillä tehtävillä pyritään mittaamaan esimerkiksi 'lukemista', toisilla taas 'puhumista'. Tarve arvioida kielen osa-alueita erillään – tai sitten toisiinsa integroiden – riippuu yleensä arvioinnin tavoitteesta: mitä yksityiskohtaisempaa tietoa kielitaidosta halutaan, sitä helpompi on perustella eri osia mittaavia tehtävistä (integroiduista tehtävistä on vaikea päätellä, missä mahdolliset ongelmat piilevät ja millaista palautetta pitäisi antaa). Mikäli tavoite on taas antaa yksi, koko kielitaitoa yleisesti kuvaava arvosana, ei ole mitään periaatteellista estettä hyvinkin integroitujen tehtävien käytölle.

Testien laadinnassa oleellisin jako ei itse asiassa kulje niin sanottujen neljän päätaidon mukaan, vaan tuottamisen (puhuminen ja kirjoittaminen) ja ymmärtämisen välillä. Monissa suhteissa puhumisen ja kirjoittamisen arviointi muistuttaa enemmän toisiaan kuin vaikkapa lukemisen ja kirjoittamisen arviointi. Tarkastelemme seuraavassa eroa hiukan tarkemmin, samoin kuin sen vaikutuksia tehtävien laadintaan (ks. tarkemmin Huhta 1997).

Perussyyy tuottamisen ja ymmärtämisen arvioinnin eroihin on siinä, miten 'suoraan' niitä voidaan havainnoida testitehtävien avulla. Ymmärtämistä ei voi tarkastella niin suoraan kuin kielen tuottamista (vaikka tuottamisenkin arviointi on aina jonkin verran epäsuoraa). Todellisessa elämässä ymmärtäminen ilmenee monin eri tavoin: asiasta puhutaan tai kirjoitetaan; siihen liittyy fyysistä toimintaa (esim. saako laitteen toimintakuntoon ohjeiden perusteella), tai ymmärtämistä voidaan päätellä kyvystä osallistua keskusteluun toisen henkilön kanssa. Useimmiten on kuitenkin mahdotonta yrittää simuloida tällaisia tilanteita testeissä, koska se veisi aivan liian paljon aikaa eikä tuottaisi tarpeeksi tarkkaa tietoa, ja huolimatta aivotutkimuksen edistysaskeleista emme myöskään pääse kurkistamaan oppijan pään sisälle ja arvioimaan ymmärtämistä suoraan. Tästä syystä joudutaan turvautumaan esimerkiksi kysymyksiin, joihin annetuista vastauksista ymmärtämisen laatu arvioidaan epäsuorasti.

Puhumista ja kirjoittamista voidaan havainnoida suoremmin siitä yksinkertaisesta syystä, että ne ovat luonnostaan toimintaa, josta jää konkreettisia jälkiä, joiden laatua arvioija voi sitten tarkastella. Selvintä tämä on kirjoittamisessa, jonka tuotokset jäävät elämään omaa elämäänsä kirjeinä, raportteina ja muina kirjoituksina. Myös puhe on tarkasteltavissa sen hetken, jona se puhutaan, tai vaikka kuinka pitkään, mikäli se äänitetään. Tämä on tärkeä periaatteellinen ero, joka johtaa suuriin eroihin myös käytännön arvioinnissa.

Ymmärtämisen testaaminen on mahdollista vain epäsuorasti vastausten perusteella. Ymmärtämisen arvosanoja voidaan antaa suoraan vertaamalla testattavan vastauksia arvosanojen kuvauksiin. Ymmärtämistä ei osata määritellä niin selkeästi ja yksityiskohtaisesti, että olisi mahdollista laatia ymmärtämisen arviointiasteikko (näistä tarkemmin myöhemmin tässä artikkelissa). Vaativuudeltaan eritasoisia ymmärtämistehtäviä voidaan toki tehdä, mutta tällöin on pakko tyytyä varsin karkeaan luokitteluun (esim. helppo, keskivaikea, vaikea), mikä ei riitä käytännön arviointiin, jos arvosanoja käytetään enemmän kuin kaksi–kolme (esim. Suomen koulujärjestelmän asteikossa 4:stä 10:een on seitsemän eri arvosanaa).

Ymmärtämisen arviointi on tästä johtuen *pisteyttämistä*: tehtävistä annetaan pisteitä, monesti erillisen pisteytysohjeen avulla (ks. kaavio 1). Tehtävien pisteet lasketaan yhteen, ja vasta yhteispisteiden perusteella päätetään lopullinen arvosana. Tässä on kyse samalla testitulosten tulkinnasta ja ehkä kaikkein subjektiivisimmasta vaiheesta ymmärtämisen arvioinnissa.

Puhumista ja kirjoittamista voidaan sen sijaan *arvioida* suoraan: arvioija tarkastelee suoritusta ja päättää arvosanan vertaamalla suoritusta suoraan arvosanojen kuvauksiin eli arviointiasteikkoon, jossa jokaista arvosanaa vastaa lyhyt sanallinen kuvaus kielitaidon laadusta. Se, että suoritukset ovat konkreettisesti olemassa kirjoitelmina ja nauhoitteina, helpottaa myös arvioijien koulutusta ja parantaa arvioinnin laatua. Kun arviointia tarkastelee kokonaisuutena, on

vaikea sanoa, onko joidenkin taitojen arviointi luotettavampaa kuin toisten: ymmärtämisen arviointia on perinteisesti pidetty luotettavampana, mutta ilmeiset vaikeudet määritellä ymmärtämistä tyydyttävästi vievät pohjaa tällaisilta yleistyksiltä.

Poikkeuksiakin edellä esitettyyn jakoon löytyy. Tiivistelmän kaltaisia osittain integroituja tehtäviä voidaan arvioida arviointias- teikon avulla aivan kuin kirjoitelmiakin, vaikka niillä pyrittäisiin mittaamaan luetun ymmärtämistä. Toisinaan puhumisteisteissäkin yritetään arvioida ymmärtämistä, jolloin ymmärtämisestä laaditaan arviointiasteikko, mutta tämän onnistumisesta tiedetään kuitenkin vielä vähän. Toisaalta puhumista ja kirjoittamista on joskus arvioitu pisteyttämällä eli laskemalla esimerkiksi virheiden määriä. Nykyisen näkemyksen mukaan tällainen 'arviointi' ei kuitenkaan kerro kuin osan kirjoitelman tai puheen laadusta.

Ymmärtämisen ja tuottamisen erojen takia myös testitehtävien laadinta eroaa. Ymmärtämisen testaaajan täytyy tuntea testattavat hyvin voidakseen valita aiheeltaan ja vaikeustasoltaan sopivia tekstejä tai nauhoitteita, erityisesti tietylle ammattialalle suunnatussa testissä. Myös tehtävien vaikeustaso täytyy sovitaa oikeaksi. Valinnoillaan laatija asettaa rajat sille, miten hyvin testiin osallistujat pystyvät osoittamaan kielitaitonsa, ja samalla myös sille, millaista tietoa hän voi saada näiden taidosta. Näin testiin osallistuja on ymmärtämistestissä hyvin pitkälle kokeen laatijan tekemien valintojen armoilla: hän voi osoittaa taitonsa vain siinä määrin kuin tehtävät antavat siihen mahdollisuuden. Koska testin suunnitteluvaihe on poikkeuksellisen tärkeä ymmärtämisteisteissä, vie testien laadinta suhteellisesti paljon aikaa.

Puhumisen ja kirjoittamisen testien laatijan ei aina tarvitse kiinnittää aivan niin paljon huomiota tehtävän vaikeustasoon. Usein sama tehtävä sopii sellaisenaan eri tasoille oppijoille, he vain suoriutuvat siitä eri tavalla. Puhumistehtävissä testaaaja (haastatteli- ja) voi lisäksi ottaa testattavan kielitaidon huomioon ja sopeuttaa testin kulun tälle sopivaksi, ellei testin vaiheita ole liian pitkälle

saneltu etukäteen ohjeissa. Monet tuottamistaitoja mittaavat tehtävät voidaan siis laatia paljon väljemmin – ja nopeammin – koska testattava itse luo suuren osan niiden ‘sisällöstä’.

Edellä sanottu voidaan tiivistää seuraavasti: Tehtävien valinnalla ymmärtämistestin laatija asettaa rajat sille, miten oppija voi testin avulla osoittaa osaamisensa. Tämä asettaa haasteita tehtävien laadinnalle. Puhumisessa ja kirjoittamisessa kielenoppija ei sitä vastoin ole samalla tavoin tehtävänannon kahleissa, koska hän on paljolti itse vastuussa testin sisällöstä, jolloin myös tehtävien laadinta helpottuu. Toisaalta tuottamistestien arviointi on vaikeampaa ja vie enemmän aikaa kuin ymmärtämistestien pisteytys.

4 MILLÄ MENETELMILLÄ KIELITAITOA ARVIOIDAAN?

Tavallisin kuva, joka luultavasti tulee mieleen kielitaidon arvioinnista on tehtävävihkonen, joka sisältää tekstiä ja kysymyksiä. Tällaiset ‘kynä ja paperi’ -testit ovatkin hyvin yleinen arviointikeino, ja ennen useimmat virallisemmat kielitestit olivat kirjallisia. Jonkin verran suullisia testejä on käytetty kielitaidon arviointiin aikaisemminkin (mm. Yhdysvaltain ulkoministeriön testit 1950-luvulla), mutta vasta kommunikatiivisen testauksen mukana tulivat useimmat suulliset testityypit yleisempään käyttöön.

Kielitaidon arviointimenetelmiä valittaessa kannattaa pitää mielessä arvioinnin tarkoitus ja laajuus. Suuriin tutkintoihin, joiden tulokset ovat tärkeitä niiden tuhansille osallistujille, kohdistuu sellaisia vaatimuksia, että vain osa tehtävätyypeistä tulee niissä kyseeseen. Tehtävien tulee tällöin olla mahdollisimman selkeitä (esimerkiksi ennestään tuttuja) ja niiden arvioinnin mahdollisimman luotettavaa ja helppoa. Pienimuotoisemmassa arvioinnissa on helpompaa kokeilla monenlaisia tehtäviä, sellaisiakin, joiden suorittaminen vaatii harjoittelua ja arviointi aikaa ja vaivaa. Varsinkin

jatkuva arviointi mahdollistaa monipuolisen tehtävien käytön, sillä se voi sisältää useimmat niistä testeistä ja harjoitustehtävistä, joita opetuksen kuluessa käytetään. Myös portfoliot antavat mahdollisuuden tavallista monimuotoisempaan arviointiin (Linnakylä ym., 1994; Pollari ym., 1996).

Seuraavassa esitellään lyhyesti tavallisimpia kirjallisia ja suullisia testitehtävätyyppejä sekä viime aikoina yleistyneitä tietokoneavusteisia tehtäviä (tehtävien laatimisesta tarkemmin ks. Weir 1990, 1993, sekä alan tutkimuksesta Bachman 1990, Douglas 1995, Skehan 1998).

Tekstin ja puheen ymmärtämisen sekä *sanaston ja kieliopin* testauksessa tavallisimmat tehtävätyypit ovat monivalinta, avoimet kysymykset ja erilaiset aukkotäydennystehtävät. On vaikea suositella yhtä tehtävätyyppiä. Järkevämpää on monipuolisuus ja kunkin tehtävän etujen ja puutteiden tunteminen. Periaatteessa useimmat tehtävätyypit mahdollistavat esimerkiksi pääasioiden – tai yksityiskohtien – ymmärtämisen testaamisen.

Kirjoittamisen arvioinnin tehtävätyypit voidaan jakaa ainekirjoituksen kaltaisiin yleisiin tehtäviin ja ohjatumpiin tehtäviin. 'Aine' tai 'essee' antaa mahdollisuuden arvioida useita kirjoitustaidon piirteitä, mutta kirjoituksen tavoite ja lukijakunnan hahmottamattomuus voivat haitata kirjoittamista. Kommunikatiivinen arviointi on tuonut mukanaan ohjatumpia ja kontekstualisoidumpia tehtäviä, joissa määritellään tarkemmin, miksi, kenelle ja mitä kirjoitelman tulisi viestiä. Pyritään myös kattamaan diskurssi- ja sosiolingvistisiä taitoja, joita nykykäsitykset kielitaidosta korostavat. Tyypillisiä ohjattuja tehtäviä ovat mm. kaavakkeet, kirjeet, ohjeet, viestit, lyhyet raportit tai tiivistelmät (Weir 1990).

Puhumisen arvioinnissa käytetään sekä kasvokkain toisen henkilön kanssa suoritettavia tehtäviä että kielistudiossa tapahtuvia puhumistehtäviä. Yleisin ja perinteisin testausmuoto, haastattelu, on eduistaan huolimatta myös yksipuolinen: se on usein testaajan johtamaa 'kuulustelua', joka ei aina kerro kaikkea oppijan taidoista

(van Lier 1989; Weir 1990). Tästä syystä myös muunlaiset tehtävät ovat yleistyneet, varsinkin roolileikit ja ongelmanratkaisutehtävät (Weir 1990, 1993; Huhta & Suontausta 1993). Kuten kirjoittamisessa, monipuolisilla puhumistehtävillä pyritään testaamaan myös oppijoiden funktionaalisia ja sosiolingvistisiä taitoja. Koska puhumisen testaus on kallista ja aikaa vievää, on suullista kielitaitoa ryhdytty testaamaan myös kielistudiossa, jossa oppijat suorittavat puhumistehtäviä ohjeiden mukaan ja heidän puheensa nauhoitetaan myöhemmin arvioitavaksi. Suomessakin on kehitetty näitä kokeita sekä tutkintoihin että tutkimuskäyttöön (ks. Opetushallitus 1995; Salmela & Tarnanen 1996; Luoma 1997; Saleva 1997).

Tietotekniikan kehittyminen on tehnyt mahdolliseksi arvioida kielitaitoa myös tietokoneen avulla. Enää tekniikka ei rajoita arviointimenetelmiä niin paljon kuin vielä 1980-luvulla. Tietokoneiden multimediaominaisuuksien kehittyminen on mahdollistanut äänen, kuvien ja jopa videokuvan integroimisen tehtäviin ja tehnyt näin tietokonetesteistä autenttisempia. Toistaiseksi tietokone ei pysty kunnolla tunnistamaan puhetta, joten puheen automaattinen arviointi (kuten kirjoitelmienkin) on vielä hyvin vajavaista eräistä yrityksistä huolimatta (Fraser ym., 1997). Eräät kansainväliset kielikokeetkin ovat siirtymässä tietokoneella suoritettaviksi.

Internet alkaa myös näkyä kielitaidon arvioinnissa, ja jo nyt se sisältää lukuisia epävirallisia testejä ja harjoitustehtäviä. Lisäksi tekeillä on laajempiakin yrityksiä arvioida oppijoiden kielitaitoa Internetin kautta, esimerkiksi Euroopan unionin DIALANG-järjestelmä, joka tuottaa diagnostista palautetta monen eri kielen osaamisesta. (Internet-testeistä tarkemmin mm. Soveltavan kielentutkimuksen keskuksen kotisivuilla, <http://www.jyu.fi/tdk/kkkk/>.)

Tehtävät ovat tärkeä osa kielitaidon arvioinnin kokonaisuutta myös siksi, että tehtävä itsessään vaikuttaa oppijan suoritukseen ja arvosanoihin (Bachman 1989, 1990; Douglas 1995). Siksi onkin perusteltua olla luottamatta liikaa vain yhteen tai kahteen tehtävätyyppiin, vaan käyttää tilanteen sallimassa määrin mahdollisimman

monenlaisia menetelmiä. Ei kannata myöskään unohtaa eri taitojen yhtäaikaista testaamista integratiivisesti, mikä voi lisätä arvioinnin autenttisuutta, tarvitaanhan monissa viestintätehtävissä ja -tilanteissa monien taitojen hallintaa.

Tehtävien tutkimukseen on vakavammin herätty vasta viime aikoina, osaksi kiinnostuksesta tehtäväpohjaiseen kielenopetukseen (Skehan 1998), osaksi osana muuta arviointitutkimusta (Bachman 1990). On ilmeistä, etteivät tehtävät ole yhtenäisiä kokonaisuuksia, vaan koostuvat eri piirteistä, jotka vaikuttavat oppijoiden suoritukseen. Tehtävien syvällisempi tunteminen auttaisi arvioimaan paremmin testitehtävien autenttisuutta ja tulosten yleistettävyyttä (Skehan 1998).

5 MITÄ ON LAATU KIELITAIDON ARVIOINNISSA?

Arvioinnin laadulla tarkoitetaan yleensä arvioinnin reliaabeliutta, validiutta ja käytännöllisyyttä. Reliaabelius (luotettavuus) oli aina 1970-luvulle arvioinnin tutkituin ja tärkeimpänä pidetty ominaisuus. Sitten on ryhdytty korostamaan validiutta (arvioinnin perusteltavuutta, tulosten järkevää käyttöä), joka nykyään nähdään laadukkaan arvioinnin ehdottomasti tärkeimpänä ominaisuutena. Reliaabelius ja validius ovat tavoittelemisen arvoisia ominaisuuksia, mutta usein käytäntö asettaa rajat sille, miten luotettavaan ja validiin mittaamiseen kulloinkin päästään. Arvioinnin tarkoitus ja seuraukset vaikuttavat yleensä suuresti tarvittavaan laatutasoon ja laadun osatekijöiden painotukseen.

5.1 Reliaabelius

Reliaabelius tarkoittaa arvioinnin tai mittauksen samanlaisuutta, toistettavuutta eri arviointikerroilla, ja sitä pidetään yleisesti laadukkaan arvioinnin välttämättömänä lähtökohtana. Mikäli oppijan taito

arvioidaan useamman kerran ja joka kerralla saadaan sama tulos, voidaan puhua reliabelista arvioinnista. Usein reliabeliudella tarkoitetaan paitsi johdonmukaisuutta myös arvioinnin luotettavuutta, tarkkuutta ja 'reiluutta' (fairness) (Henning 1987). Reliabeliuden paraneminen merkitsee samalla mittausvirheen vähentymistä. Virhe voi johtua lukuisista eri tekijöistä, kuten oppijan fyysisestä tai psyykkisestä tilasta arviointihetkellä, arvioijien epäsystemaattisuudesta, häiriöolttiista olosuhteista ja sellaisista testin ominaisuuksista kuten pituus, vaikeustaso, erottelukyky ja osien homogeenisuus (tarkemmin Henning 1987). Vaihtelu näissä saa aikaan sen, ettei kaksi eri arviointia anna aivan samanlaista tulosta, vaikka oppijan taidot eivät olisikaan välillä muuttuneet. Moniin näistä tekijöistä voidaan vaikuttaa, joten arvioinnin luotettavuutta on aina mahdollista parantaa.

Reliabeliutta voidaan tutkia järjestämällä sama testi kahdesti peräkkäin samoille oppijoille tai antamalla heidän suorittaa kaksi samanlaista testiä. Koska tämä on käytännössä hankalaa, yleensä turvaudutaan testin sisäisen yhtenäisyyden tarkasteluun tilastollisin keinoin (esim. Cronbachin alfa). Testin osioiden toivotaan mittaan van samaa ominaisuutta – osiot ovat tavallaan saman taidon peräkkäisiä mittauksia – ja mikäli ne näin tekevät, testiä pidetään reliabelina. Subjektiiivisen arvioinnin (esim. puhuminen) reliabeliutta tarkastellaan puolestaan laskemalla arvioijien antamien arvosanojen välinen korrelaatio tai keskivirhe. Yleinen suositus on, että suoritukset kaksoisarvioidaan, mikäli kyse on arvioinnista eikä pisteytyksestä, etenkin tärkeissä tutkinnoissa.

Ilman johdonmukaista arviointia on vaikea kuvitella arvioinnin voivan tuottaa kovinkaan mielekäästä tietoa oppijan taidoista tai muista ominaisuuksista. Mikäli arvioinnin tulos vaihtelee suuresti kerrasta toiseen tai arvioijasta toiseen, on vaikea tietää, mitä oikein ollaan arvioimassa, eli reliabeliutta voidaan pitää arvioinnin välttämättömänä, joskaan ei riittävänä vaatimuksena (ks. kuitenkin Mos-

sin 1994 kiistelty artikkeli esimerkkinä erilaisesta suhtautumisesta reliiabeliuteen).

5.2 *Validius*

Vaikka reliabilisuus on olennainen osa laadukasta arviointia, on tärkein arvioinnin ominaisuus kuitenkin validius. Validiuden tavanomaisin määritelmä oli pitkään se, että validi arviointi tai testi mittaa sitä, mitä on ollut tarkoituskin mitata. Nykyään validius nähdään kuitenkin pikemminkin testin tuloksista tehtävien johtopäätösten ominaisuutena: vain arvioinnin tuloksista vedetyt johtopäätökset voivat olla valideja eikä itse testi. Validiudessa on siis kysymys tulosten tulkintojen perusteltavuudesta, ja perusteluja voi olla monenlaisia aina arviointivälineiden sisällön uskottavuudesta empirisiin havaintoihin tulosten laadusta. Validius liittyy kiinteästi arvioinnin tarkoitukseen: arviointi ja tulosten tulkinta voi olla validia tiettyyn tarkoitukseen, mikä ei kuitenkaan takaa, että muunlaiset johtopäätökset tuloksista ovat yhtä perusteltuja. Testin validius ei ole itsestään selvää pelkästään sillä perusteella, että testi ja sen sisältö näyttävät uskottavilta. Kokemukset ympäri maailmaa ovat osoittaneet, että varsinkin seurauksiltaan tärkeiden testien tuloksia ja käyttöä on syytä tutkia huolella, ennen kuin voi väittää mitään varmaa siitä, mistä niiden tulokset oikein kertovat.

Monitahoisena ilmiönä validius voidaan jakaa useaan alalajiin – jopa pariin tusinaan erilaiseen (Cumming ja Berwick 1996) – joista tunnetuimmat ovat *sisällön validius* (onko testin sisältö relevantti ja sopiva), *kriteerivalidius* (mittaako testi samaa ominaisuutta kuin jokin ulkopuolinen kriteeri, esimerkiksi toinen testi), *ennustevalidius* (ennustaako testi myöhempää menestystä vaikkapa opinnoissa) ja *käsitemvalidius* (construct validity; onko testi sopuisuutta mitattavaa ilmiötä selittävien teorioiden kanssa). Nykyään validiuden lajit yhdistetään usein yhdeksi käsittevalidius-nimiseksi ilmiöksi, jonka tunnetuin malli on Messickin (1988) nelikenttä (ks.

kaavio 2). Messickin mukaan käsitevalidoinnissa on kaksi puolta: (1) todisteperusta (evidential basis) ja (2) seuraukset (consequential basis). Lisäksi käsitevalidointi voi perustua joko (3) testin tulkintaan (test interpretation) tai (4) testin käyttöön (use). Yhdessä nämä muodostavat kuviossa 2 esitetyn matriisin (Cummingin ja Berwickin 1996 muokkaama, ks. myös Takala 1996).

	Testin tulkinta	Testin käyttö
Todisteperusta	A. Käsitevalidius	B. Käsitevalidius + relevanssi/hyödyllisyys
Seuraukset	C. Käsitevalidius + arvot	D. Käsitevalidius + relevanssi/hyödyllisyys + arvot + yhteiskunnalliset seuraukset

KAAVIO 2. Validius Messickin (1988) mukaan.

Perinteinen tapa validoida testi, eli tutkia sen tulosten pätevyyttä, on tarkastella sen *tulosten tulkintaa* erilaisten todisteiden valossa (ruutu A kaaviossa 2). Näitä ovat mm. empiiriset tutkimukset testien ja mitattavan kielitaidon ominaisuuksista ja rakenteesta ja myös loogiset, teoreettiset perustelut. Lisätodisteita validiudesta voi saada tarkastelemalla testin ja sen tulosten *käyttöä* (ruutu B). Tällöin ollaan kiinnostuneita tutkimaan testattavien käyttämiä prosesseja ja strategioita samoin kuin sitä, miten heidän taustansa vaikuttaa tulokseen (koulutus, äidinkieli, ikä, sukupuoli). Kuten tuonnempana todetaan, nykyään yleistynyt kvalitatiivinen tutkimus on tuonut lisää tietoa nimenomaan tästä validiuden puolesta. Testin validoinnissa voidaan edetä vieläkin pidemmälle, jos tarkastellaan testauksen seurauksia. Tällöin pyritään selvittämään, mitkä arvot, ideologiat tai teoriat oikein ohjaavat itse testin validointiprosessia ja laadunvalvontaa (ruutu C). Tässä toistaiseksi vähän tutkitussa tarkastelutavassa ollaan kiinnostuneita mm. siitä, miten testin laatijoiden ja testattavien arvot ja käsitykset eroavat toisistaan – arvioivatko he testin

laatua samoilla perusteilla? Lisäksi voidaan tarkastella myös testin käytön sosiaalisia seurauksia (ruutu D), kuten testin vaikutusta opetukseen, eettisiä kysymyksiä ja vaihtoehtoja testaukselle (ks. Cumming & Berwick, 1996; Kunnan, 1997). Aikaisemmin arvioinnin validointi rajoittui testitulosten analyysiin (ruutu A), mutta vaihtelevassa määrin nykyään kiinnitetään myös huomiota muihinkin seikkoihin, kun arvioinnin laatua puntaroidaan.

5.3 Käytännöllisyys

Kolmas yleisesti esitetty vaatimus arvioinnin laadulle on käytännöllisyys. Käytännöllinen arviointi ei saisi viedä liikaa aikaa ja resursseja (rahaa, työtä, ihmisiä). Se, mitä tämä tarkoittaa käytännössä, vaihtelee tietysti tilanteesta ja tarkoituksesta toiseen. Sekä pienissä että suurissa testeissä vallitsee aina jatkuva jännite laatuvaatimusten (reliaabelius ja validius) ja käytännön rajoitusten välillä. Mikäli kyse on tärkeästä testistä, jonka tulisi täyttää tiukat laatuvaatimukset, käytettävissä pitäisi olla tarpeeksi resursseja varmistaa testauksen reliaabelius ja validius, sillä kokemus ulkomaisista ja kotimaisista arviointiprojekteista osoittaa, ettei arvioinnin laatua voi pitää itsestään selvänä. Seurauksiltaan vähäisemmässä arvioinnissa resurssitkin ovat yleensä paljon pienemmät, muttei silloinkaan kannata uhrata arvioinnin laatua käytännöllisyyden alttarille: ellei arvioinnin tuloksiin voi luottaa tai niitä ei voi tulkita, on arviointiin käytetty aika hukkaan heitettyä.

6 MILLAISIA KEHITYSLINJOJA KIELITAIDON ARVIOINNISSA ON NYKYÄÄN?

6.1 Metodologiset edistysaskeleet

Kielitaidon arvioinnin tutkimuksessa ollaan yleensä kiinnostuneita selvittämään arvioinnin laatua (validius, reliaaбелиus) eli kysymystä siitä, voidaanko arvioinnin tuloksiin ylipäänsä luottaa ja mistä ne oikein kertovat. Toisaalta voidaan myös haluta tutkia erilaisia kielitaitoon ja kielen oppimiseen liittyviä kysymyksiä (ks. taulukko 1). Kielitaidon arviointien ja testien tutkimuksessa käytettävät menetelmät ovat kehittyneet voimakkaasti 1980-luvulta lähtien. Varsinkin perinteisesti käytetyt kvantitatiiviset menetelmät ovat parantuneet, mutta myös kvalitatiivinen tutkimus on saanut vankan jalansijan testaustutkimuksessa. Jako kvantitatiiviseen ja kvalitatiiviseen tutkimukseen ei tosin aina ole kovin selvä, kuten toisaalla tässä julkaisussa todetaan (Piirainen-Marsh & Huhta).

Kvantitatiivinen arviointitutkimus

Uusi latenttien piirteiden testiteoria ja siitä johdetut testien analyysimallit on tunnettu jo jonkin aikaa (ks. Takala 1980a), mutta niiden käyttö perinteisen analyysin täydentäjänä on yleistynyt vasta viime aikoina (Skehan 1988/89; McNamara 1996).

Perinteisessä, ns. klassisessa testiteoriassa tehtävien laadun tunnusluvut (esim. reliaaбелиus, erottelukyky) lasketaan tulosten vaihtelusta (varianssista) ja tehtävien korrelaatioista (Henning 1987). Nämä riippuvat kuitenkin kulloisestakin kohdejoukosta, joten testin laatu riippuu paljolti siitä, keille se on pidetty (Hambleton ym. 1993). Eri testejä on vaikeata vertailla keskenään, kuten myös testattavia: sama pistemäärä eri testeistä ei välttämättä merkitse samanlaista taitoa. Tämä on suuri ongelma varsinkin laajoissa, seurauksiltaan tärkeissä tutkinnoissa ja myös tutkimusmielessä

käytetyissä testeissä, joissa tulosten vertailukelpoisuus on yleensä tärkeää.

Latenttien piirteiden teoriassa oletetaan, että testisuorituksen taustalta on löydettävissä 'piilossa' oleva (latentti) taito, piirre tai ominaisuus, jonka määrää testitehtävät mittaavat. Teoriasta johdetut IRT (Item Response Theory) -mallit kuvaavat, millä todennäköisyydellä oppija suoriutuu tehtävästä, kun tiedetään hänen taitonsa ja tehtävän vaikeustaso (Konttinen 1980). Tehtävien vaikeustasosta ja osallistujien taidoista saadaan näin absoluuttisempaa tietoa.

IRT-mallit ovat mahdollistaneet useita merkittäviä uudistuksia. Testien järjestäjät voivat laatia vertailukelpoisia testejä analysoiduista (kalibroiduista) tehtävistä (ns. tehtäväpankit). Tämä auttaa erityisesti muuten hankalasti laadittavien ymmärtämistestien kehittämistä. IRT-pohjaiset tehtävät mahdollistavat myös adaptiivisen testauksen tietokoneen avulla, jossa oppija saa suoritettavakseen vain hänen taitotasolleen sopivia tehtäviä, mikä säästää aikaa ja tuottaa tarkempia tuloksia kuin perinteinen testaus. Uudet mallit ovat monipuolistaneet myös puhumisen ja kirjoittamisen arviointia. Voidaan mm. selvittää arvioijien käyttäytymistä ja arviointiasteikojen ja tehtävien toimivuutta. IRT sopii myös kielitaidon analyysiin, mistä myös monet oppimisen tutkijat ovat kiinnostuneita (ks. McNamaran 1996 selkeä johdatus IRT:n sovelluksiin; muiden kvalitatiivisten menetelmien kehittymisestä ks. tarkemmin Skehan 1988/89; Bachman 1989).

Kvalitatiivinen arviointitutkimus

Kielitaidon arviointitutkimus on perinteisesti ollut hyvin kvantitatiivista. On kuitenkin selvää, että kvantitatiivisella tutkimuksella ei pystytä kovin hyvin selvittämään monia tärkeitä kysymyksiä, kuten testattavan käyttämiä strategioita ja prosesseja (ks. myös luku validiudesta edellä). Kommunikatiivinen arviointi korostaa testaustilanteen autenttisuutta, mutta pelkkä ulkoinen samankaltaisuus jonkin viestintätilanteen kanssa ei vielä välttämättä riitä. On tärkeää selvit-

tää, miten testattava prosessoi kieltä ja millä tavalla hän ylipäänsä lähestyy tehtäviä, jotta voidaan sanoa jotakin syvempää testin ja ulkopuolisen todellisuuden vastaavuudesta (Bachman 1990). Samoin on tärkeää tietää enemmän arvioijien toiminnasta ja perusteista, joilla he antavat arvosanoja. Näihin kysymyksiin on pyritty vastaamaan esimerkiksi haastattelemalla testattavia ja arvioijia tai pyytämällä heitä kertomaan suorituksestaan testin tai arvioinnin kuluessa. Cohenin (1997) mukaan tärkeä syy tiettyjen kvalitatiivisten menetelmien yleistymiseen vieraan kielen arvioinnin tutkimuksissa on ollut niiden menestyksellinen käyttö äidinkielen tutkimuksissa, varsinkin lukemisessa.

Banerjee ja Luoma (1997) jakavat kvalitatiiviset menetelmät viiteen luokkaan: (1) sanalliset raportoinnit (verbal reports), joissa tehtävän kuluessa tai sen jälkeen oppija kertoo suorituksestaan; (2) observoinnit, joissa tutkija havainnoi testin tms. suorittamista; (3) haastattelut ja kyselylomakkeet, joilla on tutkittu varsinkin testattavien asennoitumista tehtäviin ja testeihin; (4) testisuoritusten kielellinen analyysi, esimerkiksi eri tyyppisten tehtävien diskurssianalyysi; (5) analyysikehykset, joita on kehitetty teorian pohjalta (esim. Bachmanin ja Palmerin) ja joita käytetään testitehtävien ominaisuuksien, varsinkin sisällön, tutkimiseen. Monilla näistä tavoista yritetään valottaa arvioinnin validiuden niitä puolia, joihin vasta viime aikoina on alettu kiinnittää huomiota (ks. Messickin nelikenttä aiemmin tässä artikkelissa).

Viime vuosina on myös kehitetty tietokoneohjelmia kvalitatiivisen tutkimuksen helpottamiseksi käytännössä (ks. Weitzman & Miles 1995). Useissa tutkimuksissa yhdistyy nykyään sekä kvantitatiivinen että kvalitatiivinen näkökulma, ja tutkimuskohteen triangulointi eli tiedon hankinta erilaisin menetelmin on yleistynyt.

6.2 Kriteeriviitteinen arviointi

Kriteeriviitteisen mittaamisen yleistymisen on tärkeimpiä suuntauksia kielitaidon arvioinnissa viimeisten 20 vuoden aikana (ks. Takala 1980b). Yksinkertaisin määritelmä kriteerimittaukselle on se, että arvioinnin tulokset yritetään selittää jonkin ulkopuolisen kriteerin avulla, joka voi olla vaikkapa sanallinen kuvaus kielitaidosta. Mittauksen tuloksille pyritään siis *antamaan jokin merkitys*. Kriteerimittauksen ymmärtämistä auttaa vertailu ns. *normiviitteiseen mittaamiseen*.

Normiviitteisessä mittaamisessa pääasia on testattavien asettaminen paremmuusjärjestykseen toistensa suhteen. Testin tulos tai arvosanat määräytyvät testattavien keskinäisen *vertailun* perusteella. Tätä käytetään etenkin valintoja tehtäessä, esimerkiksi oppilaitoksen pääsykokeissa, mutta normimittaus on ollut yleistä myös perinteisessä kouluarvioinnissa. Nykyisin monien maiden kouluarvioinnissa suuntaus on kuitenkin kohti kriteeriviitteisyyttä (Huhta ym., tulossa).

Luokan tasolla tapahtuva normimittaus tuo esille tämän arviointitavan pahimman puutteen: kun vertailuryhmä on pieni, on vaara, että oppijan arvosana ei määräydykään vain hänen omien taitojensa perusteella, vaan myös muiden oppijoiden suoritukset vaikuttavat arvosanaan. Esimerkiksi keskitasoisen oppijan taidoista saa erilaisen kuvan riippuen siitä, onko hänet sijoitettu luokkaan, jossa useimmat muut oppilaat ovat häntä parempia, tai luokkaan, jossa muut ovat häntä heikompia. Liikaa vertailuun perustuva arviointi tuottaa tällaisessa tilanteessa oppilaalle todennäköisesti eri arvosanan, vaikka hän osoittaisi aivan samanlaista osaamista kummassakin luokassa.

Normiviitteinen arvosanojen anto on vankemmalla pohjalla, kun viitejoukko on suuri. Näin on esimerkiksi amerikkalaisessa TOEFL-kielikokeessa, jonka viitejoukko voi olla miljoonia, tai ylioppilastutkinnossa, johon osallistuu yli 50 % koko ikäluokasta eli

kymmeniä tuhansia oppilaita. TOEFL:n tulokset ovat lisäksi kerrasta toiseen vertailukelpoisia, koska ne perustuvat IRT:n avulla kalibroituihin tehtäviin. Ylioppilastutkinnonkin tulokset ovat todennäköisesti vertailtavia 1–2 vuoden aikavälillä, koska niin lyhyessä ajassa ei kielitaidon yleinen taso ehdi muuttua koko maan mittakaavassa. Pitemmän ajan vertailut ovat sen sijaan epäluotettavia. Molemmissa tutkinnoissa on kuitenkin hankalaa määritellä, mitä kukin arvosana oikein kertoo oppijan taidoista, koska arvosanat määräytyvät pääosin normiviitteisesti.

Kriteeriviitteisessä arvioinnissa yritetään puolestaan määritellä mahdollisimman tarkoin, mitä mitattava kielitaito on ja mikä on testitehtävien ja testattavien taitojen suhde (Takala 1980b). Kriteeriarviointi on siis keskeinen syy siihen, että kielitaidon arviointioppaat ja katsaukset, kuten tämä artikkeli, kiinnittävät paljon huomiota kielitaitoon (ks. myös Skehan 1988/89; Bachman 1990; Weir 1993; McNamara 1996).

On mahdollista jakaa kriteeriarviointi muutamaan luokkaan sen mukaan, miten yksityiskohtaisesta määrittelystä on kysymys (Skehan 1988/89). Karkeimmalla tasolla käytetään IRT:n kaltaisia teknisiä menetelmiä testien vertailukelpoisuuden lisäämiseksi (vrt. TOEFL edellä). Tarkempaa kriteeriviitteisyyttä edustavat yritykset määritellä, milloin jokin suoritus on hyväksyttävä, milloin taas ei. Tällöin on tärkeä määritellä testin 'hyväksymisraja' (cut-off point) mahdollisimman tarkasti. Kolmannessa tyypissä määritellään jokin kielitaidon alue (domain) mahdollisimman tarkasti ja kehitetään menetelmiä, joilla alueelta laaditaan testitehtäviä (Takala 1989b; Skehan 1988/89). Alueita voivat olla vaikkapa sanasto tai suullinen kielitaito, ja jo näistä esimerkeistä on helppo arvata, ettei tarkka määrittely ole aina kovin helppoa. Esimerkki tämän kriteeriarvioinnin tyypin käytännön sovelluksista ovat tehtävänlaadintaohjeet (ks. kaavio 1), joita varsinkin suuret tutkinnot yleisesti käyttävät (Lynch & Davidson 1994; Alderson ym. 1995).

Toinen yksityiskohtainen tapa kuvata kielitaitoa ovat kuvaukset kielitaidon tasoista ja, mikäli mahdollista, kielitaidon kehittymisestä. Näitä käsitellään seuraavassa tarkemmin.

6.3 Kielitaidon tasokuvaukset

Kommunikatiivisen kielitaidon arvioinnin kautena on kehitynyt runsaasti sanallisia kielitaidon tasokuvauksia, mikä on keskeinen osa yleisempää suuntausta kohti kriteeriviitteistä arviointia. Tasokuvaukset ovat ehkä eniten käytetty keino pyrkiä kriteeriviitteiseen mittaamiseen.

Taitotasoasteikko koostuu yleensä numeroista ja niitä vastavista sanallisista kuvauksista. Usein asteikon toinen pää kuvaa aloittelevan oppijan taitoja ja toinen kielen hyvää hallintaa, mikäli asteikon halutaan kattavan kaikki taitotasot. Ääripäiden välissä on asteikosta riippuen eri määrä tasoja, kukin varustettuna omalla kuvauksellaan (ks. tarkemmin Huhta 1993b). Kuvaukset koskevat yleensä puhumista ja kirjoittamista. Ymmärtämisen määrittely ei sen sijaan ole kovin yleistä, koska sitä ei osata kuvata yhtä tarkasti eikä myöskään havainnoida yhtä suoraan.

Taitotasoasteikoilla on takanaan melko pitkä historia varsinkin työsuoritusten arvioinnissa (North 1993; ks. myös Lado 1961). Ensimmäiset kielitaidon arviointiin tarkoitetut sanalliset asteikot ovat ilmeisesti peräisin amerikkalaisesta Foreign Service Instituten (FSI) testistä, ja niistä ovat saaneet vaikutteita monet nykyisistäkin tasokuvauksista.

Sanallisia tasokuvauksia kielitaidosta on alettu käyttää arviointiin ja yleisemminkin kielenopetuksessa monesta syystä (Alderson 1991; North 1993). Ne parantavat arvioinnin luotettavuutta ja mahdollistavat erilaisten testien ja arviointien vertailun ja raportoinnin samaa ymmärrettävää menetelmää käyttäen. Ne voivat yhdistää opetuksen eri osat toisiinsa (sisältöjen ja materiaalien suunnittelu, alku- ja lopputestaus, formatiivinen arviointi). Ne voivat myös

yhdistää taitotasoa ja oppimisenäkökulmat: asteikon portaat toimivat opintojen loppu- ja välitavoitteina, mutta ovat samalla yleisempiä taitotason määritelmiä. Eri tarkoituksiin on paikallaan laatia hiukan erilaisia tasokuvauksia, koska käyttäjryhmillä (oppijat, arvioijat tai testien ja opetuksen suunnittelijat) on erilaiset tarpeet ja eri valmiudet ymmärtää kuvauksia (Alderson 1991).

Kielitaitoa kuvaavien asteikkojen laatiminen on yksi tärkeimpiä ja aikaa vievimpiä vaiheita puhumisen ja kirjoittamisen arvioinnissa, koska siinä määritellään arvioitavat ominaisuudet ja koska se on osa varsinaista mittavälinettä tehtävien ja arvioijan ohella (Huhta 1993a). Asteikkoihin liittyy lukuisia ongelmia, joten ne eivät lupaa-vuudestaan huolimatta ole mikään yksinkertainen ratkaisu kielitaidon arvioinnin ongelmiin. Ehkä vakavin huoli koskee sitä, kuinka hyvin asteikkojen sisältö vastaa kielenoppimista ja osaamista todellisuudessa. Tähän on mahdollista etsiä vastauksia vain käyttämällä hyväksi muiden soveltavan kielitieteen alojen tutkimusta mm. kielen oppimisesta ja omaksumisesta (North 1993; ks. myös Brindley 1998).

Kokonaisuutena kriteeriviitteinen arviointi ja kielitaitoasteikkojen kehittyminen ovat vieneet kielitaidon arviointia eteenpäin pakottamalla testien laatijat ja muut arvioinnin suunnittelijat keskittämään huomionsa arvioinnin ehkä oleellisimpiin kysymyksiin: Mitä arvioinnin tulokset oikein tarkoittavat? Mitä ne kertovat osallistujan taidoista? Mitä johtopäätöksiä tuloksista voidaan tehdä? Ne ovat myös lisänneet arvioinnin 'läpinäkyvyyttä' ja ymmärrettävyyttä ylipäänsä, entisen, testin tuloksiin keskittyneen tutkimuksen haittapuoli kun oli se, että siitä välittyi kuva hyvin vaikeaselkoisesta ja sisäänpäin lämpiävästä alasta.

7 YHTEENVETO

Kielitaidon arviointi on nykyään monitieteinen ala, jossa yhdistyvät psykometriikka ja kasvatustieteellinen arviointi erilaisiin kielitieteen, sosiologian ja psykologian näkemyksiin kielestä ja kielitaidosta. Koska arvioinnin tärkein päämäärä on auttaa ymmärtämään arvioinnin kohdetta, on tärkeää pyrkiä selvittämään kohteen, siis kielitaidon luonne mahdollisimman tarkasti. Teoriat kielitaidosta pystyvät tähän vain osittain, mutta joitakin keskeisiä seikkoja on mahdollista löytää niistä. Useimmat nykyiset teoriat korostavat kielen käyttämisen tärkeyttä kielitiedon ohella. Tällöin arvioinninkin (oli sitten kyse testeistä tai jatkuvasta epämuodollisesta observoinnista) täytyy ottaa huomioon luonnolliseen kielenkäyttöön liittyviä seikkoja, kuten sosiaalinen ja diskursssikonteksti, jossa kommunikointi tapahtuu. Muutokset kielitaidon arvioinnissa viime vuosikymmeninä ovat myös käytännössä alkaneet toteuttaa näitä näkemyksiä, varsinkin ns. kommunikatiivisessa ja kriteeriviitteisessä arvioinnissa.

Kielitaitoa voidaan arvioida monin eri tavoin. Useimmiten ei ole syytä luottaa liikaa yhteen testi- tai tehtävätyyppiin, sillä jokaisessa on omat puutteensa. Paras strategia on käyttää useita erilaisia tehtäviä ja pyrkiä samalla simuloimaan luonnollista kielen käyttöä.

Arvioinnilla voi olla monia käytännön tavoitteita ja sovelluksia kouluarvosanojen ja diagnostisen palautteen antamisesta aina kokonaisen koulutusjärjestelmän tuloksellisuuden arviointiin saakka. Siksi on mahdotonta asettaa kaikille arviointivälineille samoja laatuvaatimuksia. Perusvaatimukset ovat toki samat: arvioinnin tulisi olla perusteltua (validia), luotettavaa (reliabelia), oikeudenmukaista ja käytännöllistä. Se, miten tarkkaa ja luotettavaa arviointitietoa tarvitaan, riippuu kuitenkin suuresti arvioinnin tarkoituksesta. Yhteen tarkoitukseen sopiva testi ei välttämättä sovi johonkin toiseen ilman muutoksia ja tutkimusta.

Kaikessa arvioinnissa – suurista testeistä luokassa tapahtuvaan arviointiin – toistuvat tietyt vaiheet (ks. kaavio 1), vaikka kaikkia yksityiskohtia (esim. kirjalliset laadintaohjeet tai esitestausta) ei aina tarvitakaan. Ehkä keskeisin vaihe on arvioinnin tarkoituksen ja sisällön määrittely, mikä sisältyy kaikkeen arviointiin: arvioijalla on aina jokin käsitys siitä, mitä kielitaidossa on tärkeää ja mittaamisen arvoista, vaikka käsitys olisi tiedostamatonkin. Mitä perustellumpi näkemys kielestä on, sitä perustellumpia ja informatiivisempia ovat myös arvioinnin tulokset.

KIRJALLISUUS

- Alderson, J. C. 1991. Bands and scores. Teoksessa J. C. Alderson & B. North (toim.), *Language testing in the 1990s*. London: British Council & Macmillan, 71–86.
- Alderson, J.C., C. Clapham & D. Wall 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Bachman, L. 1988. Language testing: SLA research interfaces, *Annual Review of Applied Linguistics* 9, 193–209.
- Bachman, L. 1989. Assessment and evaluation, *Annual Review of Applied Linguistics* 10, 210–226.
- *Bachman, L. 1990. *Fundamental considerations in language testing*. Cambridge: Cambridge University Press.
- Bachman, L. 1991. What does language testing have to offer? *TESOL Quarterly* 25, 671–704.
- Bachman, L. & A. Palmer 1984. Some comments on the terminology of language testing. Teoksessa C. Rivera (toim.), *Communicative competence approaches to language proficiency assessment: Research and application*, Clevedon: Multilingual Matters, 34–43.
- *Bachman, L. & A. Palmer 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Banerjee, J. & S. Luoma 1997. Qualitative approaches to test validation. Teoksessa Clapham & Corson (toim.) 1997, 275–287.
- Bloom, B., J. Hastings & G. Madaus 1971. *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bollen, K. 1989. *Structural equations with latent variables*. New York: Wiley Interscience.
- Botha, R. 1992. *Twentieth century conceptions of language: Mastering the metaphysics market*. Oxford: Blackwell.
- Brindley, G. 1998. Outcomes-based assessment and reporting in language learning programmes: A review of the issues, *Language Testing* 15, 45–85.
- 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. 1988. The measurement of communicative competence, *Annual Review of Applied Linguistics* 8, 67–84.
- Canale, M. & M. Swain 1980. Theoretical bases of communicative approaches to second language teaching and testing, *Applied Linguistics* 1, 1–47.
- Carroll, J. B. 1962. The prediction of success in intensive foreign language training. Teoksessa R. Glaser (toim.), *Training research and education*, Pittsburgh: University of Pittsburgh Press, 87–136.
- Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.
- Clapham, C. 1996. *The development of IELTS: A study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.
- Clapham, C. & D. Corson (toim.) 1997. *Encyclopedia of language and education 7: Language testing and assessment*. Amsterdam: Kluwer.
- Cohen, A. 1994. *Assessing language ability in the classroom*. 2nd edition. Rowley, Mass.: Newbury House/Hcinle and Heinle.
- Cohen, A. 1997. Towards enhancing verbal reports as a source of insights on test-taking strategies. Teoksessa Huhta ym. (toim.) 1997, 339–365.
- Cohen, D. 1990. Reshaping the standards agenda: From an Australian's perspective of curriculum and assessment. Teoksessa P. Broadfoot, R. Murphy & H. Torrance (toim.), *Changing educational assessment: International trends and perspectives*, London: Routledge, 32–52.
- Cumming, A. & R. Berwick (toim.) 1996. *Validation in language testing*. Clevedon: Multilingual Matters.
- Davies, A. 1978. Language testing: Survey articles 1 & 2, *Language Teaching and Linguistics Abstracts* 11, 145–159 and 215–231.
- Dickson, P. & A. Cumming (toim.) 1996. *Profiles of language education in 25 countries*. Slough: National Foundation for Educational Research.
- Douglas, D. 1995. Developments in language testing, *Annual Review of Applied Linguistics* 15, 167–187.
- Euroopan neuvosto 1997. Modern languages: learning, teaching, assessment: a common European framework of reference. Strasbourg: Euroopan neuvosto.
- Frase, L.T., C. Chapelle, A. Ginther, P. Chawla, J. Burstein, J. Klavans & J. Bernstein 1997. Technology for language assessment and learning. Teoksessa Huhta ym. (toim.) 1997, 519–543.
- Gardner, H. 1985. *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- Hambleton, R., M. Dirid & M. de Brisay 1993. New measurement models and methods for constructing language tests, *Carleton Papers in Applied Language Studies* 10, 63–81.
- Hamp-Lyons, L. 1997. Ethics in language testing. Teoksessa Clapham & Corson (toim.) 1997, 323–333.
- Havola, L. & H. Saari 1993. Kaksikymmentä vuotta englannin kuullun ymmärtämistä – onko taidoissa edistytty. Teoksessa P. Linnakylä & H. Saari (toim.), *Oppiiko oppilas peruskoulussa?* Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Henning, G. 1987. *A guide to language testing*. Cambridge, Mass.: Newbury House.
- Hughes, A. 1989. *Testing for language teachers*. Cambridge: Cambridge University Press.
- Huhta, A. 1993a. Suullisen kielitaidon arviointi. Teoksessa Takala (toim.) 1993, 143–225.

- Huhta, A. 1993b. Teorioita kielitaidosta – Onko niistä hyötyä testaukselle? Teoksessa Takala (toim.) 1993, 77–142.
- Huhta, A. 1997. Yleisten kielitutkintojen arvosteluperusteet. Teoksessa A. Huhta, A. Määttä, S. Takala & M. Tarnanen (toim.), *Yleiset kielitutkinnot ja kieltenopetus*, Helsinki: Opetushallitus, 21–43.
- Huhta, A. & T. Suontausta 1993. Suullisen kielitaidon testausmenetelmiä. Teoksessa Takala (toim.) 1993, 227–266.
- Huhta, A., V. Kohonen, L. Kurki-Suonio & S. Luoma (toim.) 1997. *Current developments and alternatives in language assessment: Proceedings of LTRC96*. Jyväskylä: University of Jyväskylä.
- Huhta, A., E. Shohamy & A. Hasselgren, tulossa. Assessment practices and responsibilities in school language education,
- Hymes, D. 1972. On communicative competence. Teoksessa J.B. Pride ja J. Holmes (toim.), *Sociolinguistics*, Harmondsworth: Penguin, 269–293.
- Konttinen, R. 1980. *Testiteoria. Johdatus kasvatukseen- ja käyttäytymistieteellisen mittauksen teoriaan*. Helsinki: Gaudeamus.
- Kunnan, A. 1995. *Test taker characteristics and performance: A structural modeling approach*. Cambridge: Cambridge University Press.
- Kunnan, A. 1997. Connecting fairness with validation in language assessment. Teoksessa Huhta ym. (toim.) 1997, 85–105.
- Lado, R. 1961. *Language testing: The construction and use of foreign language tests*. London: Longman.
- Laukkanen, R. 1994. Koulutuksen tuloksellisuuden arvioinnin keskeiset kysymykset ja trendit keskushallinnon näkökulmasta. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Lehmann, R. 1992. The achieved quality of essay writing. Teoksessa Purves (toim.) 1992, 239–274.
- Lehmann, R. 1993. Rating the quality of student writing: findings from the IEA study of achievement in written composition. Teoksessa A. Huhta, S. Takala & K. Sajavaara (toim.), *Language testing: New openings*, Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto, 186–204.
- Lewis, G. & C. Massad 1971. *The teaching of English as a foreign language in ten countries*. Tukholma: Almqvist & Wiksell.
- *Linnakylä, P., P. Pollari & S. Takala (toim.) 1994. *Portfolio arvioinnin ja oppimisen tukena*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Linnakylä, P. 1995. *Lukutaidolla maailmankartalle. Kansainvälinen lukutaitotutkimus Suomessa*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Little, A. 1990. The role of assessment, re-examined in international context. Teoksessa P. Broadfoot, R. Murphy & H. Torrance (toim.), *Changing educational assessment: International perspectives and trends*, London: Routledge, 9–22.
- Luoma, S. 1997. Comparability of a tape-mediated and a face-to-face test of speaking: a triangulation study. Lisensiaatintutkimus. Jyväskylän yliopisto.
- Lynch, B. & F. Davidson 1994. Criterion-referenced language test development: linking curricula, teachers and tests, *TESOL Quarterly* 28, 727–743.
- McNamara, T. 1995. Modelling performance: Opening Pandora's Box, *Applied Linguistics* 16, 159–179.
- *McNamara, T. 1996. *Measuring second language performance*. London: Longman.

- Messick, 1988. The once and future issues of validity: Assessing the meaning and consequences of measurement. Teoksessa H. Wainer & H. Brown (toim.), *Test validity*, Hilledale, NJ: Erlbaum, 33–45.
- Morrow, K. 1979. Communicative language testing: Revolution or evolution? Teoksessa C. Brumfit & K. Johnson (toim.), *The communicative approach to language teaching*, Oxford: Oxford University Press.
- Moss, P. 1994. Can there be validity without reliability? *Educational Researcher* 23, 2, 5–12.
- North, B. 1993. The development of descriptors on scales of language proficiency. Washington, D.C.: The National Foreign Language Center.
- Norton, B. 1997. Accountability in language assessment. Teoksessa Clapham & Corson (toim.) 1997, 313–322.
- Opetushallitus 1995. *Yleisten kielitutkintojen perusteet*. Helsinki: Opetushallitus.
- Parry, T. & C. Stansfield (toim.) 1990. *Language aptitude reconsidered*. Englewood Cliffs, NJ: Prentice Hall.
- Pollari, P., M. Kankaanranta & P. Linnakylä (toim.) 1996. *Portfolion monet mahdollisuudet*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Purves, A., E. Degenhart & T. Gorman (toim.) 1988. *The IEA study of written composition 1: The international tasks and scoring scales*. Oxford: Pergamon Press.
- Purves, A. (toim.) 1992. *The IEA study of written composition 2: Education and performance in fourteen countries*. Oxford: Pergamon Press.
- Saari, H. 1996. Englannin kielitaidot. Puhcen ymmärtäminen peruskoulun 9. luokalla v. 1995. Teoksessa R. Jaku-Sihvonon, A. Lindström & S. Lipsanen (toim.), *Toteuttaako peruskoulu tasa-arvoa?* Helsinki: Opetushallitus, 451–463.
- Saleva, M. 1997. *Now they're talking: Testing oral proficiency in a language laboratory*. Jyväskylä: Jyväskylän yliopisto.
- Salmela, T. & M. Tarnanen 1997. Yleiset kielitutkinnot – kielitaitotesti moneen tarpeeseen. Teoksessa A. Huhta, A. Määttä, S. Takala & M. Tarnanen (toim.), *Yleiset kielitutkinnot ja kielenopetus*, Helsinki: Opetushallitus, 9–20.
- Shepard, L. 1992. Uses and abuses of testing. Teoksessa M. Alkin (toim.), *Encyclopedia of educational research*, New York: Macmillan, 1477–1485.
- Shohamy, E. 1995. Performance assessment in language testing, *Annual Review of Applied Linguistics* 15, 188–211.
- Skahan, P. 1988/89. Language testing: Survey article 1 & 2, *Language Teaching* 21/22, 211–221 and 1–13.
- Skahan, P. 1998. *A cognitive approach to language learning*. Oxford University Press.
- *Spolsky, B. 1995. *Measured words*. Oxford: Oxford University Press.
- Sternberg, R. 1990. Metaphors of mind: Conceptions of the nature of intelligence. Cambridge: Cambridge University Press.
- Takala, S. 1973. *Kielitaidon mittaaminen*. POPS-Opas 3b. Helsinki: Kouluhallitus.
- Takala, S. 1975. Oppimistulosten arvioiminen. Teoksessa I. Nuotio (toim.), *Kielenopettajan käsikirja*, Helsinki: WSOY, 172–181.
- Takala, S. 1980a. Kielitaidon mittaamisesta. Teoksessa K. Sajavaara (toim.), *Soveltava kielitiede*, Helsinki: Gaudeamus, 181–201.

- Takala, S. 1980b. *Kriteerimittaamisen 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. Degenhart (toim.), *The IEA study of written composition 1: The international writing tasks and scorin scales*, Oxford: Pergamon, 3–14.
- Takala, S. 1995. Arviointikulttuurin kehittäminen. Teoksessa K. Salmio & K. Lindroos-Himberg (toim.), *Askelia yleissivistävän koulutuksen arviointiin*, Helsinki: Opetushallitus, 6–17.
- Takala, S. 1996. Suoritusarviointi puntarissa: Mahdollisuuksia ja ongelmia valtakunnallisissa kokeissa. Teoksessa Pollari ym. (toim.) 1996, 207–222.
- Takala, S. 1997. Vieraan kielen kehittymisen arviointiperusteita. Teoksessa R. Jakkusihvonon (toim.), *Onnistuuko oppiminen – oppimistuloksien ja opetuksen laadun arviointiperusteita peruskoulussa ja lukiossa*, Helsinki: Opetushallitus, 87–114.
- Takala, S. 1998. Englannin kielen taidon taso Suomessa – hyvää kehitystä. Teoksessa Takala & Sajavaara (toim.) 1998, 73–89.
- *Takala, S. (toim.) 1993. *Suullinen kielitaito ja sen arviointi*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Takala, S. & L. Havola 1983. Kielitieteen, psykologian ja filosofian väliset yhteydet kielen tutkimuksen näkökulmasta, *Kasvatus* 4, 319–326.
- Takala, S. & Saari 1979. *Englannin kielen opetus ja koulusaavutukset Suomessa 1970-luvun alussa: IEA:n kansainväliseen yhteistoimintaan perustuva vertaileva ja kuvaileva tutkimus*. Jyväskylä: Kasvatustieteiden tutkimuslaitos, Jyväskylän yliopisto.
- Takala, S. & K. Sajavaara (toim.) 1998. *Kielikoulutus Suomessa*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto.
- Valette, R. 1971. Evaluation of learning in a second language. Teoksessa B. Bloom, J. Hastings & G. Madaus (toim.), *Handbook on formative and summative evaluation of student learning*, New York: McGraw-Hill, 717–853.
- van Lier, L. 1989. Reeling, writhing, drawing, stretching, and fainting in coils: Oral proficiency interview as conversation, *TESOL Quarterly* 23, 489–508.
- Weir, C. 1990. *Communicative language testing*. New York: Prentice Hall.
- *Weir, C. 1993. *Understanding and developing language tests*. New York: Prentice Hall.
- Weitzman, E. & B. Milcs, B. 1995. *Computer programs for qualitative data analysis: A software sourcebook*. Thousand Oaks, Ca.: Sage.