

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Ahola, Sari; Neittaanmäki, Reeta; Hirvelä, Tuija

Title: Laadunvarmennus yleisissä kielitutkinnoissa

Year: 2024

Version: Published version

Copyright: © Opetushallitus

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Ahola, S., Neittaanmäki, R., & Hirvelä, T. (2024). Laadunvarmennus yleisissä kielitutkinnoissa. In T. Leblay, & S. Ahola (Eds.), Yleiset kielitutkinnot 30 vuotta (pp. 91-101). Opetushallitus. Raportit ja selvitykset, 2024: 8. <https://www.oph.fi/fi/tilastot-ja-julkaisut/julkaisut/yleiset-kielitutkinnot-30-vuotta>

Sari Ahola
tutkimuskoordinaattori
Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus

Reeta Neittaanmäki
projektitutkija
Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus

Tuija Hirvelä
projektitutkija
Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus

Laadunvarmennus yleisissä kielitutkinnoissa

Johdanto

Yleisten kielitutkintojen (YKI) todistus on monelle tutkintoihin osallistuvalla tärkeä, koska sitä käytetään monissa eri tilanteissa, joissa halutaan osoittaa kielitaitovaatimusten edellyttämä kielitaito. Tällaiset osallistujan elämää koskettavat käyttötarkoitukset edellyttävät tutkintojärjestelmältä laadunvarmennusta, jotta todistuksessa mainitut arviot antavat oikeaa tietoa osallistujan senhetkisestä kielitaitotasosta. Tästä syystä laadunvarmennus on osa jatkuvia käytänteitä YKI:ssä, ja laatua tarkkaillaan tutkintojärjestelmässä niin laadullisten kuin määrällisten menetelmien avulla. Laatuun pyritään vaikuttamaan myös erilaisilla YKIn liittyvillä tutkimuksilla, koska niiden avulla saadaan syvempää tietoa tutkintojärjestelmän käytänteistä ja toimijoista (ks. lisää Ahola, Honko & Huhta tässä julkaisussa).

Tässä artikkelissa esittelemme ensin, miksi kielitestien laadunvarmennus on tärkeää, ja sen jälkeen kuvaamme, miten YKI-tutkintojen laatua seurataan tutkintojärjestelmässä. YKI-tutkintojen laatuun liittyvä osuus on jaettu kahteen kokonaisuuteen: tehtäviin ja testikokonaisuuksiin sekä arviointiprosessiin liittyvään laadunvarmennukseen. Artikkelissa keskitymme englannin, ruotsin ja suomen tutkintoihin liittyviin laadunvarmennuskäytänteisiin, koska pääosa YKIn osallistujista on näiden tutkintokielten osallistujia ja koska näissä tutkintokielistä osallistujamäärät mahdollistavat kehittyneempien, laajastikin testauksessa ja arvioinnissa käytettävien IRT (*Item Response Theory*) -pohjaisten tilastollisten menetelmien hyödyntämisen osana laadunvarmennusta. Kaikissa YKIn tutkinnoissa tehdään laatuun liittyviä varmistuksia, mutta prosesseissa on eroja. Esimerkiksi tutkinnoissa, joissa osallistujia on vähän, IRT-pohjaisia tilastollisia analyysimenetelmiä ei voida käyttää laadunvarmennuksessa.

Laadunvarmennus kielitaidon arvioinnissa

Validiteettia ja reliabiliteettia on pidetty keskeisinä arvioinnin laadukkuutta mittaavina piirteinä. Näistä kahdesta *validiteetti* on merkitykseltään varsin laaja, koska sillä voidaan viitata koko testijärjestelmän laadunvarmennukseen. Yksinkertaistettuna validius tarkoittaa sitä, että käytetty testi mittaa sitä taitoa, mitä sillä on tarkoitus mitatakin, ja että testin ja siihen liittyvien prosessien kautta syntyneet tulokset antavat luotettavaa tietoa taidosta. *Reliabiliteetti* liittyy arvioinnin johdonmukaisuuteen ja testauksen toistettavuuteen. Arviointi on johdonmukaista, kun testi ja arviointikäytännöt ennustavat ja mittaavat onnistuneesti ja odotuksenmukaisesti osallistujien osaamista: esimerkiksi käytettävät arviointikriteerit soveltuvat taidon ja tehtävien arviointiin ja arvioijat tulkitsevat näitä kriteereitä yhdenmukaisesti ja johdonmukaisesti. Lisäksi testi on reliaabeli, kun se mittaa osallistujan taitoa samalla tavoin testikerrasta tai arvioijasta riippumatta. (Bachman & Palmer 1996: Weir 1993.)

Kielitaidon arvioinnissa käytetään myös muita laadullisia kriteereitä, kuten käytettävyyttä, autenttisuutta, vuorovaikutuksellisuutta sekä arvioinnin eettisyyttä ja oikeudenmukaisuutta. *Käytettävyys* on yhteydessä mm. testin laajuuteen, saavutettavuuteen ja testiprosessiin liittyvään ajankäyttöön. *Autenttisuus* ja *vuorovaikutuksellisuus* ovat yhteydessä testisuorituksen edustavuuteen; miten hyvin esimerkiksi testitehtävät simuloivat arkielämän todellisia kielenkäyttötilanteita. Vuorovaikutuksellisuudella voidaan lisäksi viitata siihen, miten testi haastaa osallistujaa käyttämään yksilöllisiä ominaisuuksiaan erilaisissa kielenkäyttötilanteissa (maailmantieto ja aihetieto, kielitieto, strategiset kompetenssit, affektiiviset skeemat). (Huhta & Hildén 2013; Bachman & Palmer 1996; Weir 1993.) Lisäksi nykyään laadukkaaseen arviointiin kuuluvat *arvioinnin eettisyys ja oikeudenmukaisuus* (esim. Roever & McNamara 2006). Kun tutkintotodistuksella on merkitystä osallistujan elämälle, kuten YKIn kaltaisen *high stakes* -tutkinnon tapauksessa, on tutkintojärjestelmän sisäisillä käytänteillä huolehdittava siitä, että testit eivät suosi mitään osallistujaryhmää ja että testattavien saamat tulokset ovat oikeuden- ja totuudenmukaisia ja antavat luotettavan kuvan testattavan kielitaidosta.

Arvioinnissa yleisesti laadukkaina pidettyjen piirteiden lisäksi YKIn laadunvarmennus on yhteydessä ALTE:n (*The Association of Language Testers in Europe*) laatimien laatustandardeihin (*minimum standards*). Koska YKI on yksi ALTE:n kuuluvista testijärjestelmistä, sen laatua ja käytänteitä kehitetään ja seurataan ALTE:n laatustandardien avulla. YKIn suomen keskitason tutkinto on auditoitu kolme kertaa, ja osoituksena hyvästä laadusta sille on myönnetty ALTE:n Q-merkki. Tämä laatumerkki kertoo siitä, että tutkinto on läpäissyt kaikki ALTE:n 17 (nykyään 18) laatustandardin keskeistä vaatimusta. Laatustandardit liittyvät kaikkiin arviointiprosessin vaiheisiin ja niihin osallistuneisiin toimiin. Standardit keskittyvät mm. testin käyttötarkoitukseen ja sisältöihin, testin laatimiseen, arviointiin ja tilastollisiin analyyseihin liittyviin käytänteisiin, testin hallinnoinnin ja logistiikan toimenpiteisiin sekä yhteydenpitoon ja yhteistyöhön eri sidosryhmien kanssa. (Ks. tarkemmin ALTE 2024.)

Laadunvarmennuksen vaiheita yleisissä kielitutkinnoissa

Olemme jakaneet artikkelissamme laadunvarmennuksen esittelyn kahteen osaan: tehtävien ja testikokonaisuuksien sekä arviointiprosessin laadunvarmennukseen. Artikkelia lukevan on kuitenkin hyvä huomioida, että laadunvarmennuksen vaiheet todellisuudessa limittyvät toinen toisiinsa. Esimerkiksi testikohtaisia analyysejä (tehtävien ja testikokonaisuuden toimivuus) ei voida tehdä ilman arviointiin liittyviä analyysejä.

Kuviossa 1 (ks. sivu 5) on kuvattu YKIn tehtävien ja testikokonaisuuksien sekä arviointiprosessin laadunvarmennukseen liittyvät keskeiset vaiheet. Etenemme YKIn laadunvarmennuksen esittelyssä kuviossa esitettyjen vaiheiden mukaan (lue kuviota alhaalta ylöspäin). Ensin esittelemme tehtäviin ja testikokonaisuuksiin liittyvän laadunvarmennuksen (kuviossa vihreällä) ja sen jälkeen kuvaamme arviointiprosessiin ja arvioijien toimintaan liittyvää laadunvarmennusta (kuviossa sinisellä). Kuviossa eri vaiheiden välillä olevat yhtenäisillä viivoilla merkityt nuolet viittaavat joka testikerta tapahtuviin vaiheisiin. Sen sijaan katkoviivoilla piirretyt nuolet tarkoittavat sitä, että tämä vaihe tehdään testikierroksella vain tarpeen vaatiessa. Erityisen tärkeää on myös ymmärtää, että osallistujat, heidän taustansa ja tarpeensa osallistua tutkintoihin vaikuttavat keskeisesti laadunvarmennusprosessin eri vaiheiden taustalla (kuviossa kirkkaan vihreällä katkoviivalla). Nämä tiedot ovat tärkeimmät arviointia ja sen laatua ohjaavat tekijät. Sillä, keitä tutkintoon osallistuvat ovat ja mihin he käyttävät tutkintotodistusta, on merkitystä sille, millaiseksi tutkintojen sisällöt ja rakenne muodostuvat.

osaaminen mitattavissa taidoissa eikä hänen henkilökohtaiset ominaisuutensa, kuten ensikieli, sukupuoli tai ikä.

YKlissä saadaan tietoa eri tutkintokieliin ja tutkintotasoihin osallistuvista ilmoittautumisen yhteydessä ja keräämällä taustatietoja ennen tutkintopäivää vapaaehtoisesti täytettävällä Webropol-kyselyllä (ks. tarkemmin Hirvelä tässä teoksessa). Taustakyselyn avulla saadaan tietoa mm. tutkintotodistuksen käyttötarkoituksesta (esimerkiksi kielitaidon osoittaminen kansalaisuutta, työ- tai opiskelupaikkaa varten), mikä auttaa testin sisältöjen suunnittelussa. Lisäksi osallistujien demografisia taustamuuttujia (esim. ikä, sukupuoli) käytetään testitehtävien toimivuuden tarkasteluun, jotta voidaan varmistaa kaikkien osallistujien oikeudenmukainen kohtelu arviointiprosessin aikana.

Tehtäviin ja testikokonaisuuksiin liittyvä laadunvarmennus

Tehtävien laadinta

YKI-tutkintojen sisältöjä eivät ohjaa kurssi- eivätkä opintosuunnitelmat vaan Opetushallituksen määräystasoinen asiakirja, Yleisten kielitutkintojen perusteet (Opetushallitus 2011). Tutkintojen perusteissa on kuvattu tutkintojen kielitaitokäsitys, testattavat osataidot, arvioinnin perusteet ja tehtävien aihepiirit sekä kielenkäyttötilanteet. Lisäksi laadintaa varten on koottu ohjeet, joissa esitellään eri testattavat osataidot ja niihin liittyvät tehtävätyypit sekä yleisiä ohjeita tehtävänlaadinnasta. Laadintaohjeiden lisäksi laatijoiden koulutuksella ja ohjauksella pyritään varmistamaan se, että arvioijat noudattavat tehtäviä laatiessaan tutkintojärjestelmän kielitaitokäsitystä ja sovittuja käytänteitä. Laadunvarmennus tarkoittaakin laatijoiden kohdalla hyvää laadinnan ohjausta, selkeitä laadintaohjeita ja tehtävien laadullista kommentointia. (Ahola 2012.)

Testimateriaalia on nykyään helppo löytää internetistä, mutta kaikki materiaali ei välttämättä sovi testitarkoitukseen, ja myös tekijänoikeuskysymykset on huomioitava materiaalia valittaessa. Keskeistä testimateriaalin valinnassa on se, että materiaalin kielenkäyttötilanteet ja muut ominaisuudet simuloivat todellisia kielenkäyttötilanteita (tarve osallistua tutkintoon) ja että ne huomioivat testattavien ominaisuudet (esim. sukupuoli, ikä). Laatija joutuukin jo laadintavaiheessa tekemään laadullista valintaa sen suhteen, millainen materiaali on aiheeltaan sopiva osallistujajoukolle (aihepiiri, vaikeustaso) ja miten hän saa kyseisen materiaalin muokattua testitehtäväksi. Testimateriaalia valitessa on myös huomioitava se, että YKlissä tehtävät eivät ole kertakäyttöisiä, vaan niitä käytetään useammalla testikerralla. Tästä syystä tehtävien sisältöjen tulee kestää aikaa.

Laadintavaiheessa tehtävät käyvät YKlissä läpi kommentointivaiheen, jossa kunkin laatijan tehtäväversioita kommentoivat sekä YKlissä työskentelevät asiantuntijat että muut kyseisen kielen laatijatiimiin kuuluvat laatijat. Laadintavaiheessa yleensä kommentoidaan materiaalin soveltuvuutta testitarkoitukseen, tehtävän vaikeustason sopivuutta testitasolle sekä osioiden (osio = tekstiin liittyvä yksi kysymys) ja tehtävätyypin toimivuutta valitussa materiaalissa. Tällainen laadullinen tehtävien tarkastelu on tärkeä vaihe laadintaprosessissa, koska se muokkaa tehtäviä paremmaksi ja koska sen aikana tehtävien mahdolliset heikkoudet ja puutteet tulevat esille (Ahola 2012). Tarkempaa tietoa testitehtävien toimivuudesta saadaan kuitenkin vasta testikäytön jälkeen, jolloin tilastollisten analyysien avulla saadaan tietää tehtävien vaikeustaso ja johdonmukaisuus (erottelukyky).

Testin koostaminen

Tehtävien tilastollista tietoa hyödynnetään testikokonaisuuksien koostamisessa. YKlissä on käytössä kieli-, taso- ja osakoekohtaiset osiopankit, joilla tarkoitetaan hyvin toimivien tehtävien varastoa. Osiopankista löytyy tietoa olemassa olevien tehtävien käyttöhistoriasta ja toimivuudesta. Osiopankki

onkin perusta validien, luotettavien ja vertailukelpoisten testiversioiden laatimiselle. Lisäksi se on kustannustehokas tapa laatia testikokonaisuuksia, koska testikokonaisuuksiin ei tarvitse laatia useita uusia tehtäviä.

Testikokonaisuuden koostamisessa hyödynnetään myös tehtävien taitotasolle asettamista (ks. kohta Standard setting), koska tämän prosessin avulla saadaan tietää, minkä tasoista osaamista eri tehtäväsiot mittaavat. Tämä tieto on tärkeää, koska YKl:ssä yhdellä testikokonaisuudella mitataan kahta erillistä taitotasoa, esimerkiksi keskitasolla taitotasoja 3 ja 4. Tästä syystä testikokonaisuudessa on oltava vaikeustasoltaan tehtäviä molemmilta taitotasoilta.

Määrällisen laadunvarmennuksen lisäksi testikokonaisuuden koostamisessa huomioidaan YKIn perusteet (Opetushallitus 2011), joissa on kuvattu tutkinnoissa käytettävät aihepiirit ja kielenkäyttötilanteet. Koostamisvaiheessa testikokonaisuudessa tulee käyttää laajasti erilaisia aiheita ja kielenkäyttötilanteita, jotta kielitaidosta saadaan mahdollisimman monipuolinen otos. Lisäksi kielitaidon arvioinnissa on tärkeää käyttää erilaisia tehtävätyyppejä sekä eri pituisia ja monipuolisia teksti- ja diskurssilajeja edustavia tehtäviä. Ymmärtämisessä teksteihin kohdistuneiden kysymysten on lisäksi mitattava erilaisia luku- ja ymmärtämisstrategioita. Tuottamisen taidoissa tehtävien on taas hyvä testata muun muassa konventioiden, rekistereiden ja kirjoittajan tai puhujan pragmaattisia ja sosiolingvistisiä taitoja.

Koostamisen jälkeen testikokonaisuus käy läpi vielä kommentointikierroksen. Tässä vaiheessa sekä YKIn arvioinnin ja kielen asiantuntijat että ulkopuolinen asiantuntija käyvät tehtävät läpi. He tarkistavat, että tehtäviin ei jää virheitä, että tehtäväkokonaisuus testaa eri aihepiirejä ja kielenkäyttöfunktioita ja että tehtävät sopivat kyseiselle tutkintotasolle.

Testikohtaiset analyysit tehtävien ja testikokonaisuuden toimivuuden tarkistamiseksi testin jälkeen

Yleisten kielitutkintojen eri osataitojen testit analysoidaan sekä laadullisesti että määrällisesti jokaisen testin jälkeen. Määrällisillä menetelmillä analysoidaan yksittäisten testitehtävien sekä koko testin toimivuutta. Laadullista tarkastelua tehdään koko ajan määrällisten tulosten rinnalla. Analyysillä varmistetaan, että testit ovat vertailukelpoisia testikerrasta toiseen (sisältö, vaikeustaso, johdonmukaisuus ja tasoarvioiden rajat) ja että ne mittaavat luotettavasti ja tarkasti kyseessä olevaa taitoa (testitulokset ovat tarkkoja, johdonmukaisia ja mittausvirheettömiä). Nämä analyysit ovat tärkeitä myös siksi, että uudet tehtävät esitestataan varsinaisessa testissä. Jos tehtävät osoittautuvat toimimattomiksi ja vaikeustasoiltaan poikkeaviksi, ne poistetaan testituloksista. Tämä on kuitenkin harvinaista, koska tehtäviä on tarkasteltu laadullisesti useampaan otteeseen ennen testikäyttöä.

Testattava osataito vaikuttaa käytettävään tilastolliseen menetelmään, koska arvioinnin kautta saatu aineisto on erilainen. Tuottamistaidot arvioidaan suoraan arviointikriteereiden avulla taitotasolle, kun taas ymmärtämistaidoissa taitotason määrittely on yhteydessä osallistujan saamiin pistemääriin osakokeesta (ks. Standard setting). Ymmärtämistaitojen analyysit perustuvat pääasiassa IRT-pohjaiseen Raschin malliin (Rasch, 1960, 1980), mutta myös klassisen testiteorian menetelmiä käytetään tulosten tarkastelussa. Tuloksista nähdään, miten yksittäiset tehtävät toimivat ja miten luotettavasti testi toimii kokonaisuutena. Vaikka testikokonaisuudessa on aiemmin testeissä käytettyjä toimivia tehtäviä osiopankista, tämä on tärkeä analyysivaihe, koska tehtävien toimivuudessa havaittavat muutokset saattavat viestiä esimerkiksi osallistujajoukossa tapahtuneista muutoksista.

Tuottamistaitojen analysoinnissa käytetään Raschin mallin laajennusta, joka huomioi samanaikaisesti testitehtävän vaikeuden, osallistujan kyvyn ja arvioijan ankaruuden. Analyysien perusteella voidaan

tehdä ymmärtämistaitojen tapaan päätelmiä testitehtävien ja testikokonaisuuden toimivuudesta, mutta analyysillä (ks. arvioijien seuranta) varmistetaan myös se, että arvioijat ovat arvioineet johdonmukaisesti eivätkä ole olleet poikkeavan ankaria tai lempeitä. Lisäksi analyysien lopputuloksena saadaan jokaiselle suorittajalle määriteltyä heidän osaamistaan vastaava taitotaso.

YKlissä on käytössä myös menetelmiä, joiden avulla voidaan varmistaa, että osallistujien väliset erot suorituksissa liittyvät ensisijaisesti arvioitaviin taitoihin eivätkä testin ulkopuolisiin tekijöihin, kuten suorittajan taustaan (sukupuoli, ikä, ensikieli, kansalaisuus). Tämä on tärkeä osa laadukasta testauksesta ja arviointia sekä peruste sille, miksi testijärjestelmissä tarvitaan tietoa osallistujien taustoista.

Osiopankki päivitetään myös jokaisen testikierroksen jälkeen: uudet tehtävät lisätään osiopankkiin ja jo käytössä oleviin tehtäviin lisätään uutta tietoa. Tehtävien toimivuudessa tapahtuvien muutosten seuraaminen on yksi keskeinen osa laadunvarmennusta. Joskus tehtäviä myös poistetaan osiopankista, jos tehtävien sisältö on vanhentunut tai jos havaitaan selkeä muutos hyvin toimivan osion vaikeustasossa. Testitehtävät, jotka toimivat tilastojen perusteella huonosti tai joiden toimivuudessa havaitaan muutoksia, ohjataan takaisin kielen asiantuntijalle laadullista analyysia varten. Vasta sen jälkeen tehdään päätös siitä, voiko tehtävää enää käyttää jatkossa.

Standard setting

Ymmärtämistaitojen taitotasorajat perustuvat standard setting (taitotasojen asettaminen) -prosessiin, joka suoritetaan kerran tai kaksi kertaa vuodessa (ks. myös Ahola, Neittaanmäki & Hirvelä 2016). Standard setting -menetelmiä on paljon, ja eri menetelmät saattavat tuottaa erilaisia tuloksia (Cizek & Bunch, 2007). Testijärjestelmissä menetelmän valinta onkin tehtävä huolella, jotta käytetty menetelmä palvelee testauksen tarpeita. Taitotasojen asettaminen on testauksessa keskeinen prosessi, jonka avulla pyritään yhdenmukaistamaan ymmärtämistaitojen eri testiversiot. Tuottamistaidoissa tämä prosessi ei ole yhtä oleellinen, koska puhumisessa ja kirjoittamisessa suoritukset arvioidaan arviointikriteerien avulla suoraan taitotasolle. Sen sijaan tekstin ja puheen ymmärtämisen osakokeista osallistuja saa kokonaispistemäärän, josta ei voida suoraan päätellä, millä taitotasolla henkilö on. Taitotasojen asettamisen tavoitteena onkin määritellä, mitä taitotasoa eri pistemäärät tarkoittavat: mikä on se pistemäärä, joka erottelee peräkkäiset taitotasot.

Koska käsitykset taitotasosta ja niiden yhteydestä arvioitaviin tehtäviin sekä näkemykset riittävästä suoriutumuksesta tietyllä taitotasolla ovat subjektiivisia ja vaihtelevat asiantuntijoittain, tarvitaan päätöksen tekemiseen useamman asiantuntijan näkemys ja yhteisymmärrys. Taitotasolle asettamiseen liittyykin keskeisesti asiantuntijoiden välinen keskustelu, jossa haetaan yhteisymmärrystä siitä, millaista taitotasoa tehtävään vastaaminen edellyttää. YKlissä standard setting -tilaisuuksiin osallistuu 10–15 asiantuntijaa, jotka ovat kyseisen kielen arvioijia tai toimivat YKlissä tehtävän laatijoina. Testin pisterajojen taitotasolle määrittämisen lisäksi YKlissä määritellään yksittäisten testitehtävien taitotasoa.

Testin pistemäärän asettamista taitotasolle ei tarvitse tehdä jokaiseen testiversioon, koska eri testiversiot pyritään pitämään sisällöltään ja myös vaikeustasoltaan vertailukelpoisina osiopankin avulla. Taitotasojen asettamisella kerran kaksi vuodessa varmistetaan, että aiemmillä kerroilla määritellyt rajat ovat edelleen oikeita ja toimivia eri osatiedoissa.

Arviointiin liittyvä laadunvarmennus

Arvioijien peruskoulutus ja arviointikoulutus ennen testisuoritusten arviointia

Arvioinnissa laatu syntyy siitä, että arviointi on osuvaa ts. osallistujan saamat arviot kertovat hänen todellisesta osaamisestaan. Tästä syystä on tärkeää, että arvioijat tulkitsevat arviointikriteereitä ja -ohjeita mahdollisimman samankaltaisesti. Arvioinnin yhdenmukaisuuteen pyritään koulutusten avulla, sillä yhteiset keskustelut suoritusnäytteistä ja arviointikriteereistä auttavat arvioijia tuotos-ten asettamisessa taitotasolle. Koulutus on keskeinen osa arvioijana toimimista, vaikka koulutuksen merkityksestä tutkijoilla on erilaisia käsityksiä. Osa pitää arvioijakoulutuksen merkitystä vähäisenä ja lyhytkestoisena (Eckes 2008; Lumley & McNamara 1995; Shohamy, Gordon & Kraemer 1992), koska arviointikäyttäytyminen on osa arvioijan persoonaa (Zhu, Fung, Tse & Hsieh 2017). Toiset taas ovat todenneet, että arvioijien sisäinen johdonmukaisuus paranee ja että sillä on vaikutusta myös eri arvioijien väliseen johdonmukaisuuteen ja eroihin ankaruudessa tai lempeydessä (Davies 2016; Fahim & Bijami 2011; Lim 2011). Arvioijakoulutukset ennen arvioinnin aloittamista ovat tärkeitä arvioijille, koska heidän omien käsitystensä mukaan koulutukset auttavat heitä sisäistämään paremmin kielitaidon kehittymisen (taitotasokuvaimet) (Ahola 2016). Lisäksi arvioijayhteisöllä on suuri merkitys arvioijana toimimiselle: toisia arvioijia kuuntelemalla voi oppia myös itse arvioimaan, mutta muita aktiivisemmin koulutuksiin osallistuvat arvioijapersonat voivat myös vaikuttaa toisten arvioijien käsityksiin näytesuorituksissa olevien oppijoiden taidosta. Siksi on tärkeää, että koulutuksissa pyritään aktivoimaan kaikki arvioijat osallistumaan keskusteluihin.

Arvioijana toimimista YKl:ssä ohjaa laissa mainittu asetus (A 1163/2004), jossa on kuvattu arvioijalta vaadittava koulutus ja muut kelpoisuusvaatimukset. Lisäksi asetuksessa on maininta Opetushallituksen edellyttämän arvioijakoulutuksen suorittamisesta. Tässä Soveltavan kielentutkimuksen keskuksen organisoimassa peruskoulutuksessa arvioijat perehdytetään YKl:n tutkintojärjestelmään ja standardoidun arvioinnin erityispiirteisiin, käytettyihin testitehtäviin, arviointimenetelmiin, arviointitaitoihin ja eri taitotasoihin. Koulutuksessa myös harjoitellaan arviointia näytteiden ja oikeiden testisuoritusten avulla. (Ahola & Tossavainen 2014.) Kaksiosaisen peruskoulutuksen aikana tulevien arvioijien kanssa keskustellaan arviointiperusteista testisuoritusten avulla ja peilataan niitä arviointikriteereihin. Tällä tavoin perehdytetään arvioijat taitotasoihin ja arviointikriteereiden monipuoliseen käyttöön. Samalla heidät perehdytetään arviointikäytänteisiin (mm. aikataulut, arviointialustan käyttö), jotta arviointiprosessi sujuisi mahdollisimman jouhevasti tutkintotilaisuuden jälkeen. Peruskoulutuksen viimeisessä vaiheessa arvioijat osallistuvat varsinaisen arviointipäivän arviointiin ja heille annetaan palautetta arviointien osuvuudesta ja siitä, miten heidän arviointilinjansa näyttäytyy kokeneempien arvioijien joukossa.

Peruskoulutuksen lisäksi arvioinnin laatuun pyritään vaikuttamaan joka arviointikertaa ennen järjestettävillä arviointikoulutuksilla. Arvioijan on osallistuttava sen osakokeen arviointikoulutukseen, jota hän arvioi kyseisellä tutkintokierroksella. Arvioijien säännöllinen kouluttaminen on osa laadunvarmennusta, ja koulutuksella pyritään siihen, että arvioijat tulkitsevat arviointikriteereitä ja myös tehtäväkohtaisia kriteereitä mahdollisimman samalla tavalla. Vaikka arvioijilla on eroja lempeydessä tai ankaruudessa, pääosa arvioijista kuitenkin päätyy arvioinnissaan samaan arvioon. Arvioijien oman käsityksenkin mukaan koulutuksissa käydyt keskustelut auttavat heitä kehittymään arvioijina ja vaikuttavat heidän arviointikäyttäytymiseensä tuoden varmuutta arviointiin ja vahvistaen kuvaa eri taitotasosta (Ahola 2016), vaikka arviointiin liittyy aina jonkin verran epävarmuutta (Honko & Neittaanmäki 2023).

Arviointikäyttötymisen seuraaminen

Ihmisten suorittamat arvioinnit tuovat arviointiprosessiin subjektiivisuutta ja ei-toivottua vaihtelua. Arviointiin voivat vaikuttaa sekä osallistujiin että arvioijiin liittyvät tekijät, kuten osallistujan sukupuoli ja ensikieli (Halonen ym. 2023) sekä arvioijan kokemus arvioinnista (Davis, 2016; Lim, 2011) ja opettamisesta (Royal-Dawson & Baird, 2009). Arvioijat voivat myös tulkita ja soveltaa arviointikriteerejä eri tavalla (Barkaoui, 2010). Käytännössä tämä tarkoittaa sitä, että arviointikoulutuksista ja arvioijien pätevyysvaatimuksista huolimatta arvioijat voivat antaa erilaisia tai jopa ristiriitaisia arvioita osallistujan kielitaidosta, mikä on ongelma arviointitulosten oikeudenmukaisuudelle ja pätevyydelle. YKIn kaltaisessa *high stakes* -testissä on ehdottoman tärkeää valvoa arviointien laatua, jotta voidaan varmistaa, että kaikkien osallistujien kielitaito arvioidaan oikeudenmukaisesti ja luotettavasti ja että heidän saamansa testitulokset ovat vertailukelpoisia huolimatta siitä, milloin he ovat osallistuneet testiin ja kuka heidän suorituksensa on arvioinut.

Arviointikäyttötyminen ei ole muuttumaton ominaisuus, vaan se saattaa vaihdella testikerrasta toiseen tai siinä tapahtuu muutoksia kokemuksen ja arviointivuosien myötä (Neittaanmäki & Lamprianou, 2024b). Arvioijilla saattaa esiintyä ongelmia arvioinnissaan. Arviointikäyttötymisen ongelmilla viitataan tavallisesti poikkeavan lempeään tai ankaraan arviointilinjaan, arvioinnin epäjohtonmukaisuuteen tai taipumukseen käyttää vain osaa arviointiasteikosta. Näitä tekijöitä tarkkaillaan säännöllisesti myös YKlissä, jotta voidaan varmistaa, ettei yksittäisen arvioijan arviointi suosi tai syrji ketään osallistujaa. Lisäksi YKlissä seurataan arvioijan aktiivisuutta osallistua arviointeihin, koska poissaolojen on havaittu vaikuttavan arvioijan arviointilinjaan, erityisesti arviointien johdonmukaisuuteen (Neittaanmäki & Lamprianou, 2024a). Arviointikäyttötymisessä tapahtuvien muutosten ja arvioijien ammattitaidon kannalta onkin tärkeää, että arvioijat osallistuvat ja myös pääsevät arvioimaan eri osaitaitoja säännöllisesti.

YKlissä arviointikäyttötymisen seuraaminen alkaa jo arvioijien peruskoulutuksessa, jolloin arvioijat saavat palautetta arvioimistaan suorituksista. Tarkemmin arviointikäyttötymistä tarkkaillaan kuitenkin joka testikierrokseen liittyvien arviointien aikana. Lisäksi vielä jokaisen arvioijan testikohdasta arviointilinjaa verrataan hänen aikaisempaan arviointilinjaansa, jotta nähdään, että siinä ei ole tapahtunut merkittäviä muutoksia. Myös arviointiin käytetystä ajasta ja teknisistä ongelmista kerätään arviointikierroksen aikana tietoa. Näitä tietoja kerätään, jotta voimme selvittää muita mahdollisia syitä arvioinnista löytyneisiin ongelmiin.

Vaikka arviointikäyttötymistä seurataan joka arviointikierroksella, henkilökohtaista palautetta arvioinnista annetaan vain silloin, jos arvioijan arviointilinjassa havaitaan ongelmia. Palautteen antamisen jälkeen kyseisen arvioijan arviointikäyttötymistä seurataan tarkemmin 2–3 seuraavan arviointikierroksen aikana. Jos arvioijalla on edelleen ongelmia arvioinnissaan, hänelle tarjotaan lisäkoulutusta tai sovitaan joistakin muista toimenpiteistä arvioijan kanssa. Tutkimustulokset palautteen vaikutuksesta ja vaikutuksen kestosta ovat ristiriitaisia (ks. esim. Knoch, 2011; Hoskens & Wilson, 2001). Kuitenkin YKlissä saatujen kokemusten perusteella näyttää siltä, että arviointikäyttötyminen paranee heti palautteen jälkeen, erityisesti palautteeseen myönteisesti suhtautuvien arvioijien kohdalla, mutta vaikutus voi jäädä lyhytaikaiseksi. Jos useaan kertaan saadulla palautteella ja lisäkoulutuksella ei ole korjaavaa vaikutusta arvioijan arviointilinjaan, arvioijan on lopetettava arviointityö YKlissä. Tällaiset tapaukset ovat hyvin harvinaisia.

Taitotasoarvioiden viimeistely ja tarkistaminen

Kun suoritukset on arvioitu ja kun arviointikäyttäytyminen ja testien toimivuus on analysoitu, alustavat puhumisen, kirjoittamisen, tekstin- ja puheen ymmärtämisen taitotasoarviot ovat valmiina jokaiselle osallistujalle. Osallistujan saamia arvioita eli kielitaitoprofiilia tarkastellaan vielä tässä vaiheessa tarkemmin. Tarkastelussa huomioidaan ennen kaikkea sellaisten osallistujien kielitaitoprofiilit, joilla jonkin osataidon taitotasoarvio poikkeaa muiden osataitojen arvioista tai joilla jonkin osataidon taitotasoarvio on kahden taitotason rajalla (ns. rajatapaus). Tämä vaihe vie aikaa ja hidastaa jonkin verran lopullisten tulosten valmistumista, mutta osallistujien oikeudenmukaisen kohtelun vuoksi on tärkeää, että vielä ennen lopullisten tulosten valmistumista tulosten oikeellisuus tarkistetaan.

Tarkistusvaiheessa käydään läpi myös ns. vilppiepäilyt. Vilpillistä menettelyä havaitaan sekä tutkintopäivänä tutkintopaikalla että tulosten valmistumisvaiheessa Soveltavan kielentutkimuksen keskuksessa (Solki). Havaituista epäilyistä tehdään kirjallinen ilmoitus Opetushallitukselle. Tavallisesti Solkissa havaitut vilppiepäilyt tulevat esiin arviointiprosessin aikana tai kielitaitoprofiilien tarkistusvaiheessa. Arviointiprosessin aikana havaituissa epäilyissä kyse on yleensä siitä, että osallistujan tuotoksessa havaitaan jotain poikkeavaa. Kielitaitoprofiilien tarkistusvaiheessa havaittu tyypillinen vilppitapaus on sellainen, jossa osallistujan yksittäisen tai useamman osataidon kehittämisessä havaitaan lyhyellä aikavälillä poikkeuksellisen suuri muutos.

Arvioinnin oikaisuvaatimuspyyntö

Osallistujalla on oikeus pyytää maksullista tutkintosuorituksen arvioinnin tarkistamista 14 vuorokauden kuluessa tutkintotodistuksen saapumisesta (Muutoksenhaku, laki yleisistä kielitutkinnoista, 13§ 20.12.2022/1082). Oikaisuvaatimuspyynnön voi kohdistaa yhteen tai useampaan osataitoon. Osallistujan oikeusturvan takaamiseksi uudelleenarvioinnin tekee eri arvioija kuin alkuperäisessä arvioinnissa. Tutkinnon laadukkuudesta kertoo se, että vain pieni osa johtaa tasoarvion nousuun ja oikaisuvaatimuspyyntömaksun palautukseen. (Ks. lisää Opetushallitus 2024.)

Pohdintaa laadunvarmennuksesta ja tulevaisuuden näkymiä

Olemme yllä kuvanneet testien sisältöön ja arviointiin liittyvää laadunvarmennusta, mutta kuten alussa mainitsimme, laadukkaassa testijärjestelmässä laatua varmistetaan myös monessa muussa vaiheessa. Esimerkiksi erityisjärjestelyistä huolehtiminen, testimateriaalien valmistaminen ja niiden käsittely, suoritusten välittäminen arvioijille ja todistusten tuottaminen ovat testikierroksen aikana vaiheita, jotka vaikuttavat myös testijärjestelmän laatuun. YKlissä onkin pyritty kehittämään myös näihin vaiheisiin sellaisia käytänteitä, jotka noudattavat kielitutkintojärjestelmiltä edellytetyjä laadukriteereitä.

Tutkinnon jälkeen osallistujat yleensä odottavat malttamattomina tuloksia, ja heidän on välillä vaikea ymmärtää, miksi tulosten saaminen kestää niin kauan. Laadunvarmennusta tehdään kuitenkin osallistujien edun vuoksi. Jos tutkintotuloksien valmistumista halutaan nopeuttaa, tulee ennemmin kehittää tutkintojen suorittamista digitaalisesti kuin heikentää laadunvarmennusta, koska osallistujat käyttävät tuloksiaan yhteiskunnallisesti merkittäviin tarkoituksiin, ja kaikkien edun vuoksi testien on pystyttävä antamaan luotettavaa tietoa kielitaidosta.

Yleisten kielitutkintojen tutkinnot ovat kansainvälisestikin tarkasteltuna valideja, ja niiden arviointi on tarkkaa ja luotettavaa. Laatu ei kuitenkaan ole pysyvä ominaisuus, vaan laadukkaan tutkintojärjestelmän on pystyttävä kehittämään ja arvioimaan kriittisestikin omia käytänteitään. Tästä syystä

onkin tärkeää, että YKl:ssä seurataan testikierrokseen liittyvien käytänteiden rinnalla yhtä tarkemmin sitä, keitä tutkintoon osallistujat ovat ja mikä merkitys tutkinnoilla on ollut heidän elämälleen. Tällä hetkellä vapaaehtoinen, erillinen verkkopohjainen kysely ei anna tarpeeksi tietoa osallistujista, koska valitettavasti vain kolmannes osallistujista vastaa siihen. Tulevaisuudessa olisikin hyvä, jos taustakysely saataisiin osaksi Opetushallituksen ylläpitämää ilmoittautumisjärjestelmää. Näin voitaisiin paremmin varmistaa, että tutkintojen sisällöt vastaavat osallistujien tarpeita. Laadukkaaseen arviointiin kuuluu myös se, että testijärjestelmissä tiedetään, mitä merkitystä testin suorittamisella on ollut osallistujille ja kuinka oikein testi on ennustanut osallistujan kielitaidon riittävyyttä arkipäivän kielenkäyttötilanteissa. YKl:ssä tarvitaankin enemmän tutkimusta, joka keskittyy testien vaikutavuuteen – ja tämä tutkimus on jo aloitettu.

Kirjallisuus

- Ahola, S. 2016. Puhetta arvioinnista: Yleisten kielitutkintojen arvioijien käsityksiä arvioinnista. Teoksessa A. Huhta, & R. Hildén (toim.), *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. Suomen soveltavan kielitieteen yhdistys. AFinLA-e: soveltavan kielitieteen tutkimuksia 9, 89–109. <http://journal.fi/afinla/article/view/60848>
- Ahola, S. 2012. Yleisten kielitutkintojen laatijoiden käsityksiä kielestä ja tehtävien laadinnasta. *Lisensiaattitutkielma*, Jyväskylän yliopisto. <http://urn.fi/URN:NBN:fi:juu-201301241110>
- Ahola, S., R. Neittaanmäki & T. Hirvelä 2016. Puheen ymmärtämisen tehtävien taitotasolle asettaminen Yleisissä kielitutkinnoissa. Teoksessa A. Huhta, & R. Hildén (toim.), *Kielitaidon arviointitutkimus 2000-luvun Suomessa*. Suomen soveltavan kielitieteen yhdistys. AFinLA-e: soveltavan kielitieteen tutkimuksia, 9. <http://journal.fi/afinla/article/view/60847>. 68–88.
- Ahola, S. & H. Tossavainen 2014. Yleisten kielitutkintojen arvioijakoulutus. Teoksessa T. Leblay, T. Lammervo & M. Tarnanen (toim.), *Yleiset kielitutkinnot 20 vuotta. Raportit ja selvitykset 2014*: 16. Helsinki: Opetushallitus, 125–134.
- Bachman, L.F. & A.S. Palmer 1996. *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.
- Barkaoui, K. 2010. Variability in ESL Essay rating processes: The role of the rating scale and rater experience. *Language Assessment Quarterly*, 7(1), 54–74. <https://doi.org/10.1080/15434300903464418>
- Cizek, G. J. & M. B. Bunch 2007. *Standard Setting. A Guide to Establishing and Evaluating Performance Standards on Tests*. SAGE Publications.
- Davis, L. 2016. The influence of training and experience on rater performance in scoring spoken language. *Language Testing*, 33(1), 117–135. <https://doi.org/10.1177/0265532215582282>
- Eckes, T. 2008. Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing* 25 (2), 155–185. <https://doi.org/10.1177/0265532207086780>
- Fahim, M. & H. Bijani 2011. The effects of rater training on raters' severity and bias in second language writing assessment. *Iranian Journal of Language Testing* 1, 1–16.
- Halonen, M., Ahola, S., Hirvelä, T., Neittaanmäki, R., Ohranen, S. & R. Ullakonoja 2024. Kielitaidon arviointin kansalaisuuden portinvartijana. Teoksessa TA Renvik & M. Säävälä (toim.) *Kotoutumisen kokonaiskatsaus 2023: Näkökulmana väestösuhteet. TEM oppaat ja muut julkaisut 2024:1*. Helsinki: Työ- ja elinkeinoministeriö, 75–84.
- Hirvelä, T. 2023. Yleisten kielitutkintojen osallistujien monet kasvot. *Kieli, koulutus ja yhteiskunta*, 14(2). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-maaliskuu-2023/yleisten-kielitutkintojen-osallistujien-monet-kasvot>.
- Honko, M. & R. Neittaanmäki 2023. Arvioinnin epävarmuus ja sen vaikutukset arvioinnin laatuun suomen kielen puhumisen taidon testiarvioinnissa. *Puhe ja kieli*, 43(4), 171–193. <https://doi.org/10.23997/pk.144389>
- Hoskens, M., & M. Wilson 2001. Real-time feedback on rater drift in constructed-response items: An example from the Golden State Examination. *Journal of Educational Measurement*, 38(2), 121–145.
- Huhta, A. & Hildén, R. 2013. *Kielitaidon arvioinnin metodologisia vaihtoehtoja*. Teoksessa A. Räisänen (toim.), *Oppimisen arvioinnin kontekstit ja käytännöt*. Helsinki: Opetushallitus, 159–186.
- In'nami, Y. & R. Koizumi, 2016. Task and rater effects in L2 speaking and writing: A synthesis of generalizability studies. *Language Testing*, 33(3), 341–366. <https://doi.org/10.1177/0265532215587390>
- Knoch, U. 2011. Investigating the effectiveness of individualized feedback to rating behavior – a longitudinal study. *Language Testing*, 28(2), 179–200. <https://doi.org/10.1177/0265532210384252>

- Lim, G. S. 2011. The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4), 543–560. <https://doi.org/10.1177/0265532211406422>
- Lumley, T. & McNamara, T. F. 1995. Rater characteristics and rater bias: implications for training. *Language Testing* 12 (1), 54–71. <https://doi.org/10.1177/026553229501200104>.
- McNamara, T., & Roever, C. 2006. Psychometric Approaches to Fairness: Bias and DIF. *Language Learning*, 56(Suppl 2), 81–128. <https://doi.org/10.1111/j.1467-9922.2006.00381.x>
- Neittaanmäki, R., & I. Lamprianou 2024a. All types of experience are equal, but some are more equal: The effect of different types of experience on rater severity and rater consistency. *Language Testing*, 41(3). <https://doi.org/10.1177/02655322241239362>
- Neittaanmäki, R., & I. Lamprianou 2024b. Communal factors in rater severity and consistency over time in high-stakes oral assessment. *Language Testing*, 41(3).
- Opetushallitus 2011. Yleisten kielitutkintojen perusteet 2011. Määräys 24/011/2011. Määräykset ja ohjeet 2011:33 Tampere: Juvenes Print – Tampereen yliopistopaino Oy.
- Rasch, G. 1960/1980. Probabilistic Models for Some Intelligence and Attainment Tests, Chicago: The University of Chicago Press. <https://doi.org/10.1177/02655322241239363>
- Royal-Dawson, L., & Baird, J.-A. 2009. Is Teaching Experience Necessary for Reliable Scoring of Extended English Questions? *Educational Measurement: Issues and Practice*, 28(2), 2–8. <https://doi.org/10.1111/j.1745-3992.2009.00142.x>
- Roever, C. & McNamara, T. 2006: Language testing: the social dimension. *International Journal of Applied Linguistics* Vol. 16, No. 2, 242258.
- Shohamy, E., Gordon, C. M. & Kraemer, R. 1992. The effect of rater's background and training on the reliability of direct writing test. *The Modern Language Journal* 76, 1, 27–33. <https://doi.org/10.2307/329895>
- Zhu, Y., Fung, S. L., Tse, S. K. & Hsieh, C. Y. 2017. Exploring the Relationship Between Raters' Personality Traits and Rating Severity in Writing Assessment of Chinese as a Second Language: A Pilot Study. In D. Zhang & Lin CH (Eds), *Chinese as a Second Language Assessment*. Springer, 203–223. http://dx.doi.org/10.1007/978-981-10-4089-4_10.
- Weir, C. 1993. *Understanding and Developing Language Tests*. New York: Prentice Hall.

Verkkolähteet

- A 1163/2004. Valtioneuvoston asetus yleisistä kielitutkinnoista. Annettu Helsingissä 16.12.2004. [online] Finlex [luettu 18.4.2024] Saatavissa: <https://www.finlex.fi/fi/laki/ajantasa/2004/20041163>
- ALTE 2020. ALTE Principles of Good Practice [online]. ALTE [luettu 6.4.2024]. Saatavissa: <https://pt.alte.org/resources/Documents/ALTE%20Principles%20of%20Good%20Practice%20Online%20version%20Proof%204.pdf>
- ALTE 2024. Minimum standards [online]. ALTE [luettu 15.4.2024] Saatavissa: Minimum Standards 2024 EN.pdf ([alte.org](https://www.alte.org))
- Laki yleisistä kielitutkinnoista. Muutoksenhaku, 13§ 20.12.2022/1082. <https://www.finlex.fi/fi/laki/ajantasa/2004/20040964>
- Opetushallitus 2024. Tarkistusarviointi [online]. Opetushallituksen verkkosivut [luettu 2.5.2024] Saatavissa: <https://www.oph.fi/fi/koulutus-ja-tutkinnot/yki-testin-jalkeen>