

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

Author(s): Kekki, Jarkko; Ahola, Sari; Neittaanmäki, Reeta

Title: Englannin ja suomen arvioijien käsityksiä ylimmän tason puhumisen osakokeesta

Year: 2024

Version: Published version

Copyright: © 2024 Opetushallitus

Rights: In Copyright

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

Please cite the original version:

Kekki, J., Ahola, S., & Neittaanmäki, R. (2024). Englannin ja suomen arvioijien käsityksiä ylimmän tason puhumisen osakokeesta. In T. Leblay, & S. Ahola (Eds.), *Yleiset kielitutkinnot 30 vuotta* (pp. 112-119). Opetushallitus. Raportit ja selvitykset, 2024: 8. <https://www.oph.fi/fi/tilastot-ja-julkaisut/julkaisut/yleiset-kielitutkinnot-30-vuotta>

Jarkko Kekki
tutkimuskoordinaattori
Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus

Sari Ahola
tutkimuskoordinaattori
Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus

Reeta Neittaanmäki
projektitutkija
Jyväskylän yliopisto, Soveltavan kielentutkimuksen keskus

Englannin ja suomen arvioijien käsityksiä ylimmän tason puhumisen osakokeesta

Johdanto

Artikkelissa tarkastelemme englannin ja suomen ylimmän tason (YKI 5–6~EVK C1–C2) puhumisen osakoetta ja arvioijien käsityksiä osakokeeseen kuuluvista studiokokeesta ja haastattelusta sekä niissä käytettävistä tehtävätyypeistä. YKIn kaltaisissa tutkintojärjestelmissä käytetään tyypillisesti puhumisen taidon testaamisessa joko haastattelua tai studiokoetta, mutta YKlissä ylimmän tason puhumisen osakokeessa on aina ollut nämä molemmat osat: kielistudiossa suoritettava studio-osa ja kasvokkain tapahtuva haastattelu. Vaikka tutkintojärjestelmän jokaisessa tutkintokielessä on käytetty kyseisiä osia jo vuosia testattaessa ylimmän tason puhumisen taitoa, näihin kahteen osaan ja niiden arviointiin liittyvää tutkimusta ei ole tehty viime vuosina. Luoma (1997) vertaili lisensiaatintutkielmassaan studiokokeen ja kasvokkaisen haastattelun eroja. Vaikka otos oli pieni ja vaikka kyse oli keskitasolla tehdystä kokeilusta, Luoman tulosten mukaan osallistujat saivat lievästi parempia taitotasoarvioita haastattelusta kuin studio-osuudesta (1997,114). Toisaalta tutkimuksessa havaittiin, että osioiden taitotasoarvioiden välinen korrelaatio oli melko korkea (0,85), mikä käytännössä tarkoittaa sitä, että jos osallistuja sai haastattelusta paremman taitotasoarvion, sai hän yleensä studiostakin paremman taitotasoarvion. (Ks. myös Wigglesworth & O’Loughlin 1993; Stansfield ym. 1990.)

Artikkelimme aineisto koostuu englannin ja suomen arvioijille lähetetystä kyselystä, jonka avulla kuvaamme, mitä huonoja ja hyviä puolia arvioijien käsitysten mukaan puhumisen osien arviointiin liittyy, miten laajasti heidän mielestään näiden kahden erillisen osan arvioinnissa voidaan soveltaa puhumisen arviointikriteereitä ja miten hyvin studiokoe ja haastattelu mittaavat osallistujien puhumisen taitoa.

Ylimmän tason puhumisen osakoe YKlissä

Yleisten kielitutkintojen ylimmän tason puhumisen osakoe koostuu kaikissa kielissä kahdesta erillisestä osasta: studio-osasta (n. 15 min) ja haastattelusta (n. 15 min). Molemmilla osilla on oma funktionsa puhumisen taidon määrittelyssä, sillä osien tehtävät vaativat osallistujilta erilaisten kielenkäyttötilanteiden hallintaa. Puhumisen arvioinnissa on keskeistä, että osakoe sisältää erilaisia tehtäviä, jotka mittaavat monipuolisesti puhumisen taitoa. Yleisten kielitutkintojen perusteet ohjaavat osakokeiden sisältöjä: niissä on kuvattu taidon kehittyminen, osakokeen sisältö sekä tutkintojen aihepiirit ja kielenkäyttötilanteet. (Yleiset kielitutkinnot 2011.)

Studio-osan tehtävät koostuvat kolmesta erityyppisestä tehtävästä, ja se suoritetaan nimensä mukaisesti yleensä kielistudiossa tai tietokoneluokassa, jossa suoritukset tallennetaan myöhemmin

tapahtuvaa arviointia varten. Osa kestää taukoineen noin 15 minuuttia, josta itse puheaikaa on keskimäärin puolet. Osallistujan osuus studio-osassa on ajastettu siten, että taustalla kuuluva mastertalenne ohjaa tehtäviin valmistautumista ja puheaikaa. Vaikka tämänkaltaiset ajallisesti rajoitetut puhetilanteet eivät toki ole täysin normaaleja puhetilanteita jokapäiväisessä elämässä, puheajan rajaamisella voidaan kuitenkin varmistaa se, että osallistujat saavat saman ajan tuottaa samasta aiheesta puhetta.

Studio-osassa on kolme erillistä tehtävää, joista ensimmäinen tehtävä on kertominen. Kertomista voidaan kutsua myös ns. lämmittelytehtäväksi, koska sen yhtenä funktiona on antaa osallistujalle mahdollisuus puhua jostain henkilökohtaisemmasta aiheesta, jotta hän pääsee puhumisen rytmiin kielistudiossa ja jotta mahdollinen testijännitys pääsisi laukeamaan. Kertomisen aihe on hyvin vapaamuotoinen siten, että otsikon perusteella aiheen käsittelyä ei rajoiteta tarkemmin. Osallistujan tehtävänä on puhua valmistautumisajan (1½ minuuttia) jälkeen annetusta aiheesta 1½ minuuttia.

Toisena tehtävänä on kuusi lyhyttä tilannetta. Kaksi muuta tehtävää studio-osassa ovat puhtaasti monologityyppisiä tehtäviä, mutta lyhyillä tilannetehtävillä pyritään simuloimaan erilaisissa vuorovaikutustilanteissa tarvittavia reagoineja. Tilannetehtävät testaavatkin osallistujan kykyä reagoida nopeasti erilaisiin jokapäiväisen elämän kielenkäyttötilanteisiin. Osallistuja saa sanallisen kuvauksen tilanteesta, johon hänen tulee reagoida (20–30 s) lyhyen valmistautumisajan jälkeen (20 s). Tilannetehtävissä keskeistä on testata yleisten kielitutkintojen perusteissa kuvattuja kielenkäyttötilanteita, kuten mielipiteen tai asenteiden ilmaisemista tai tunteiden ilmaisemista ja kysymistä (Opetushallitus 2011). Puhetilanteiden konteksti ja sisällön kuvaus tulee olla rajattu tarkasti siten, että testattava pystyy itsenäisesti tuottamaan tehtävän edellyttämän reagoinnin sille annetussa ajassa (myös Ahola & Tossavainen 2013).

Studio-osan päättävässä mielipidetehävässä testataan sitä, miten osallistuja kykenee argumentoimaan annetusta aiheesta 2 minuuttia yleisellä ja käsitteellisellä tasolla. Testattava voi valita kahdesta aiheesta toisen käsiteltäväksi. Molempia otsikoita ohjaa muutama kysymys, joiden tehtävänä on helpottaa aiheen käsittelyä ja joilla pyritään ohjaamaan puhetta mahdollisimman käsitteelliselle tasolle.

Haastattelu kestää noin 15 minuuttia, ja se videoidaan arviointia varten. Osallistuja saa valita otsikon perusteella ennen haastattelun alkua neljästä aiheesta kaksi aihetta, joista hän haluaa keskustella haastattelijan kanssa. Hänellä on mahdollisuus valmistautua haastatteluun kaksi minuuttia. Haastattelu koostuu kolmesta eri vaiheesta: alkukeskustelu (2–3 min.), aihe 1 (n. 6 min) ja aihe 2 (6 min). Haastattelu suoritetaan koulutetun haastattelijan kanssa, joka saa tutkintojärjestelmästä listan valmiita kysymyksiä kustakin haastatteluaiheesta. Vaikka haastattelun eteneminen on ohjattu antamalla haastattelijalle lista kysymyksiä, haastattelut kuitenkin saattavat edetä varsin hyvin eri tavoin riippuen siitä, millainen kielitaito osallistujalla on ja mihin suuntaan haastattelu etenee. Lisäksi haastattelijalla on mahdollisuus keksiä aiheesta lisäkysymyksiä. Haastattelija keskittyy vain haastattelutilanteeseen eikä osallistu millään tavalla osallistujan suorituksen arviointiin. Haastattelijan täytyy kuitenkin tuntea kielitaitotasot, jotta hän voi haastaa osallistujaa osoittamaan kielitaitonsa parhaalla tavalla.

Kyselytutkimuksen sisällöt ja kyselyyn vastanneiden arvioijien taustamuuttajat

Aineisto kerättiin keväällä 2024 (maalis- ja huhtikuussa) Webropol-kyselyllä. Kysely lähetettiin yleisten kielitutkintojen suomen ja englannin arvioijille, joilla oli paljon kokemusta ylimmän tason puhumisen suorituksen arvioinnista YKl:ssä tai jotka olivat arvioineet ylimmän tason puhumisen suorituksia viimeisten vuosien aikana.

Kysely sisälsi sekä valintakysymyksiä että avoimia kysymyksiä. Kyselyn alussa arvioijilta tiedusteltiin heidän opetuskokemustaan ja kokemustaan eri tasoista oppijoista (*Oletko toiminut päätoimisesti opettajana viimeisen viiden vuoden aikana? /Miten paljon sinulla on kokemusta opetustyösi*

kautta ylimmän tason oppijoista?). Opettamiseen ja kokemukseen liittyvä kysymys on tärkeä, koska arvioijan kokemus ylimmän tason oppijoista auttaa arviointityössä. Tämän jälkeen arvioijia pyydettiin kertomaan studio-osan ja haastatteluosan hyviä ja huonoja puolia sekä mikä tai mitkä tehtävyyt ovat helpoimpia ja vaikeimpia arvioida (*Minkä tehtävän arviointi on helpointa/vaikeinta studio-osassa ja perustelut*). Arvioijia pyydettiin myös pohtimaan, vaikuttaako arviointiin haastattelijan tai suorittajan persoona (Likert-asteikollinen monivalintakysymys *Vaikuttaako haastatteluissa haastattelijan persoona arviointiin? /Vaikuttaako haastatteluissa osallistujan näkeminen/persoona arviointiin?* 1 = ei mitään vaikutusta, 2 = hieman, 3 = kohtalaisesti, 4 = paljon, 5 = hyvin paljon sekä avoin kysymys perusteluihin). Lisäksi arvioijilta kysyttiin Likert-tyyppisellä arviointiskaalalla (erittäin huonosti – erittäin hyvin), miten hyvin he pystyvät soveltamaan arviointikriteereitä eri tehtävätyypeissä. Heitä pyydettiin perustelevaan, jos he vastasivat, että arviointikriteerejä pystyy soveltamaan huonosti. Arvioijilla oli mahdollisuus myös kertoa, jos jotain kriteeriä ei pysty soveltamaan lainkaan ylimmän tason arvioinnissa. Kyselyn avulla haluttiin saada myös käsitys, ovatko arvioijien mielestä molemmat, studio-osa ja haastattelu, tarpeellisia osallistujan kielitaidon arvioimisen näkökulmasta (*Ovatko molemmat osat tarpeellisia, vai pystyisikö vain jommankumman perusteella arvioidaan osallistujan kielitaidon? Miksi?*).

Kyselyyn vastasi yhteensä 7 englannin arvioijaa ja 9 suomen arvioijaa. Kyselyyn vastanneet arvioijat erosivat toisistaan arvioijakokemukseltaan, joka vaihteli 1–30 vuoteen. Kuitenkin heistä reilulla kolmasosalla oli yli 20 vuoden kokemus YKI-arvioinnista, toisella kolmasosalla 10–20 vuoden kokemus YKI:ssä toimimisesta ja lopulla kolmasosalla alle 10 vuotta. 80 % (n = 14/16) vastaajista oli toiminut päätoimisesti opettajana viimeisen 5 vuoden aikana koko ajan tai lähes koko ajan, ja 70 %:lla vastaajista (n = 10/16) oli paljon kokemusta ylimmän tason oppijoista. Noin 70 %:lla (n = 11/16) vastaajista opiskelijat olivat olleet viimeisen 5 vuoden aikana pääasiassa ylimmän tason opiskelijoita.

Käsityksiä ylimmän tason puhumisen osakokeesta

Studio- ja haastatteluosan arviointiin liittyviä helppouksia ja vaikeuksia

Arvioijista studio-osan arviointi oli haastatteluosan arviointia vaikeampaa. Studio-osasta arvioijat (n = 13/16) mainitsivat etenkin lyhyet tilannetehtävät vaikeimmiksi tehtäviksi arvioida. Lyhyiden tilannetehtävien ongelmana mainittiin osallistujan näkökulmasta se, että ne vaativat osallistujilta mielikuvitusta ja kykyä tehdä nopeita päätöksiä tuotoksen sisällöstä. Siksi suoritukset eivät aina ole asiayhteyteen sopivia tai osallistujat eivät kykene tuottamaan annetussa ajassa arvioinnin näkökulmasta tarpeeksi tuotosta. Joku arvioijista piti myös yksin studiossa suoritettavia dialogityyppisiä tilannetehtäviä epäluontevina puhetilanteina verrattuna monologityyppisiin mielipidet tehtäviin. Tilannetehtävien arviointia vaikeutti myös se, että arvioijat eivät kirjallisten kommenttien mukaan voineet käyttää arvioidessaan kaikkia arviointikriteereissä kuvattuja puhumisen piirteitä, vaan arviointi kohdistui enemmän osallistujan kykyyn reagoida nopeasti tilanteessa. Lisäksi joku arvioija mainitsi, että heikotkin osallistujat saattoivat onnistua lyhyessä tilanteessa, jos he vain osasivat täyttää tehtävännannon vaatimukset. Koska tilannetehtävien tuotokset ovat lyhyitä, arvioijien oli myös joskus vaikea päättää, miten paljon virheitä vain 20–30 sekuntia kestävä suoritus voi sisältää. (Ks. myös Ahola ym. 2013.)

Muutama arvioijista (n = 3/16) mainitsi myös mielipidet tehtävän vaikeimmaksi studiokokeen tehtäväksi. Syyksi mainittiin se, että tehtävän kohdalla vaatimukset saattoivat nousta liian korkeaksi: *"saatan mielessäni vaatia suorituksesta liikaa: samaan aikaan loogisesti etenevää puhetta, mielipiteen selkeää kertomista ja tarkkaa perustelua, aiheen käsittelyä yleisemmällä ja abstraktimmalla tasolla"* (Arvioija 3). Mielipidet tehtävän vaikeudesta ja helppoudesta arvioijien käsityksissä oli ristiriitaisuuksia, sillä suurin osa arvioijista (n = 11/16) nimesi mielipidet tehtävän studio-osan arvioitavista

tehtävistä kaikkein helpoimmaksi arvioida. Helpoksi sen teki se, että osallistuja tuotti riittävästi sisältöä arviointia varten ja tehtävän aiheet ovat osallistujille tarpeeksi haastavia, jotta hän voi osoittaa paremmin kielitaitonsa ja kykynsä tuottaa koherenttia puhetuotosta.

Mielenkiintoista vastauksissa oli se, että vaikka tehtävä 1. kertominen on tilastollisesti helpoin studiotehtävä, sen mainitsi vain kaksi arvioijaa helpoimmaksi tehtäväksi arvioida. Tähän syynä on ehkä se, että vaikka tehtävä on osallistujille helppo, arvioijan saattaa olla vaikea arviointitilanteessa päättää, miten käsitteellistä sisältöä hän voi tehtävässä osallistujalta vaatia. Tehtävässä osallistujat tavallisesti pysyttelevät sisällöllisesti varsin henkilökohtaisella tasolla, kun taas esimerkiksi mielipide tehtävässä osallistujalta odotetaan kykyä puhua annetusta aiheesta käsitteellisellä tasolla.

Haastatteluosuus

Haastatteluosaa pidettiin yleisesti helppona arvioinnin näkökulmasta, koska arvioinnissa pystyi käyttämään arviointikriteereitä laajasti päätöksen tukena. Haastatteluosan arviointi vei paljon aikaa, mutta siitä huolimatta haastattelu oli arvioijista studio-osan tehtäviä luontevampi puhetilanne, jossa osallistuja voi halutessaan tuottaa paljon ja esitellä monipuolisesti kielitaitoaan sekä käyttää myös hyväkseen nonverbaalista viestintää. Haastattelua pidettiin kielitaidollisesti haastavana, koska osallistujan odotetaan puhuvan haastattelussa melko käsitteellisellä tasolla, mikä aiheuttaa osalle osallistujista vaikeuksia: *"Joskus osallistujalla ei ole tarvittavaa yleistietoutta vastata hyvin haastaviin tiettyä aihetta koskeviin kysymyksiin tai osallistuja ei ole tarpeeksi kouluttautunut pärjätäkseen haastattelusta"* (Arvioija 13). Yleensä haastattelu koettiin arvioitavan näkökulmasta kuitenkin reiluna tapana testata puhumisen taitoa, koska siinä osallistuja pystyy itse säätämään vuorovaikutusta ja vaikuttamaan siihen, mihin suuntaan keskustelu etenee.

Haastatteluun liittyi kuitenkin arvioijien mukaan muita ongelmia, joilla voi olla vaikutusta arviointiin ja osallistujan suoriutumiseen osakokeessa. Yksi tällainen vaikuttava tekijä on haastattelijan ja hänen toimintansa ja persoonansa haastattelun aikana. Arvioijat totesivatkin, että haastattelijoiden on eroja, sillä osa heistä osaa viedä haastattelua osallistujaa tukien eteenpäin, kun taas osa esittää vain kysymyksiä innostamatta osallistujia näyttämään todellisen kielitaitonsa. Yhden arvioijan mukaan pahimmassa tapauksessa haastattelijan saattaa vaikuttaa osallistujan suoriutumiseen heikentävästi tai jopa estää kokonaan taidon esiin tuomisen. Osa haastattelijoiden myös puhui arvioijien mukaan itse haastattelun aikana liikaa tai toi liiaksi esiin omia mielipiteitään vieden siten osallistujalta tilaa osoittaa osaamistaan. Joskus haastattelu koettiin myös raskaaksi arvioida, jos osallistujalla oli häiritseviä manereita tai hän oli tullut tutkintoon liian heikolla kielitaidolla, jolloin keskustelun kuunteleminen oli pitkästyttävää.

Kysymykseen haastattelijan persoonan vaikutuksesta arviointiin arvioijat vastasivat eri tavoin. Kolmannes ($n = 05/16$) arvioijista oli sitä mieltä, että haastattelijalla on kohtalaisesti vaikutusta haastattelun arviointiin. Toiselle kolmannekselle ($n = 5/16$) haastattelijalla oli joko paljon tai hyvin paljon vaikutusta arviointiin. Loput arvioijista ($n = 6/16$) vastasivat joko *hieman* tai *ei lainkaan vaikutusta*. Se, että vastaukset erosivat toinen toisistaan, johtunee ehkä siitä, että arvioijat tulkitsivat kysymykset eri tavoin. Vaihtoehdon *hieman* tai *ei lainkaan vaikutusta* valinneista ehkä osa oli tulkinnut kysymyksen siten, että haastattelijan persoonalla ei ole suoraa vaikutusta arviointiin vaan vaikutus on epäsuora: haastattelijan toiminta vaikuttaa siihen, miten osallistuja suoriutuu haastattelutilanteesta ja miten hyvin hän pystyy osoittamaan kielitaitonsa.

Arvioijien vastauksista haastatteluun tehtävätyyppinä liittyy heikkous – erot haastattelijoiden välillä. Haastattelujen välisiä eroja pyritään tutkintojärjestelmässä vähentämään laatimalla haastattelijoiden apukysymyksiä haastatteluaiheista, joita he voivat käyttää haastattelujensa runkona, mutta

haastattelijan persoonaan tai tyyliin haastatella on vaikea vaikuttaa. Erot haastattelujen välillä ovat tutkinnon luotettavuutta heikentävä tekijä, koska haastattelutilanteen tulisi olla kaikille osallistujille samankaltainen, jotta jokainen osallistuja voisi osoittaa samalla tavoin kielitaitonsa. Arvioijien mukaan osa haastatteloista ei kykene luomaan haastattelutilanteeseen rentoa tunnelmaa tai antamaan osallistujalle rohkaisevaa palautetta, joka innostaisi haastateltavaa puhumaan: *"Haastattelijä voi persoonallaan saada arvioitavan 'jäätymään' tai aukeamaan. Joskus haastattelijat nostavat itsensä haastateltavan yläpuolelle, mikä lisää haastateltavan epävarmuutta. Empaattinen ja haastateltavasta kiinnostunut haastattelijä voi saada arvioitavan tuottamaa enemmän ja parempaa puhetta kuin jäykkä, kylmä ja etäinen. Jotkut haastattelijat ovat robottimaisia, eivätkä vaikuta olevan pätkäkään kiinnostuneita tehtävästään."* (Arvioija 9)

Kaikki haastattelijat eivät myöskään osaa haastaa osallistujia kielitaidollisesti, jotta he voisivat osoittaa parhaan taitonsa haastattelutilanteessa: *"Jos haastattelijä lukee annetut apukysymykset mekaanisesti läpi ilman omaa panosta haastatteluun, ei osallistuja pääse näyttämään kielitaitoaan. Haastattelun pitäisi muistuttaa normaalia epäsymmetristä keskustelua, jossa toinen auttaa ja tukee sitä osapuolta, jolla on heikommat valmiudet. Hyvä haastattelijä saa puhujasta monin verroin enemmän irti."* (Arvioija 2)

Arvioijat olivat sitä mieltä, että arvioinnin aikana hieman vaikutusta on myös sillä, että he näkevät osallistujan haastatteluvideolta (n = 10/16). Studio-osahan tehdään pelkästään puhetallenteen avulla, joten jos arvioija arvioi studio-osan suoritukset ensin, hänellä ei ole käsitystä siitä, miltä puhuja todellisuudessa näyttää. Todennäköisesti hän kuitenkin tekee äänitallennetta kuunnellessaan jonkinlaisia oletuksia osallistujasta, sillä aikaisemmissa tutkimuksissa on todettu (esim. tässä teoksessa Halonen; Ahola, Honko, Huhta), että arvioijat tekevät testattavan puheen perusteella päätelmiä hänen persoonastaan ja taustoistaan (ks. myös Ahola 2022; Ahola & Halonen 2022). Arvioijien vastausten perusteella osallistujan näkemisellä oli enemmän positiivisia kuin negatiivisia vaikutuksia, sillä se saattoi synnyttää hieman myötätuntoa tai ymmärrystä osallistujaa kohtaan: *"Jos oma mielikuvani hakijasta (joka on syntynyt kuunnellessani studio-osiota) vahvistuu tai muuttuu, saan yleensä hieman enemmän 'tietoa' hakijan kyvyistä, ja tämä tarkoittaa yleensä sitä, että saatan osoittaa hieman enemmän myötätuntoa arvioinnissani."* (Arvioija 10)

Vaikka arvioijat tunnistavat itse arvioinnin aikana osallistujan persoonan vaikuttavan hieman arviointiin, useimmiten osallistujan persoonan vaikutus arviointiin on kuitenkin tiedostamaton, kuten pari arvioijaa kommentoissaan totesivat: *"Tämä on niitä seikkoja, jotka tapahtuvat tiedostamattomasti"* (Arvioija 6): *"Joskus huomaan, että osallistuja saattaa ilmaista näkemyksen, jota en todellakaan jaa, mutta uskon silti pystyväni erottamaan omat mielipiteeni ja hänen puhetaitonsa arvioinnin aikana. En kuitenkaan voi sulkea pois tiedostamattomien ennakkoluulojen mahdollisuutta, joten tästä syystä olen merkinnyt 'hieman'."* (Arvioija 14). Hieman-vaihtoehdon vastanneet perustelivat vastaustaan myös sillä, että haastatteluun kuuluvat eleet ja ilmeet, joiden näkeminen on keskeinen osa vuorovaikutusta: *"Puhumista tukeva non-verbaalisuus voi vaikuttaa osallistujan kannalta positiivisesti; Osallistujan ilmeet, eleet, liikehdintä ja yleisesti läsnäolo ja reagointi vaikuttaa arviointiin. Kokonaiskuva kielitaidosta syntyy vuorovaikutuksessa – koen myös, että arvioija seuraa intensiivisemmin osallistujan puhumista, kun saa seurata kuvasta."* (Arvioija 4)

Arvioijista ainoastaan yksi sanoi osallistujan näkemisen vaikuttavan paljon hänen arviointiinsa. Paljon vastannut arvioija kuitenkin lisäsi kirjallisessa kommentissaan, että osallistujan jännittyneisyys voi vaikeuttaa arviointia, eli vaikutusta on lähinnä silloin, kun osallistuja on silminnähden jännittynyt, mutta muuta vaikutusta osallistujan näkemisellä ei ole. Muutama arvioija oli taas sitä mieltä, että osallistujan näkemisen ei saisi vaikuttaa lainkaan arviointiin, ja jos se jonkun osallistujan kohdalla vaikuttaisikin, on hyvä, että suoritusta arvioi aina useampi kuin yksi arvioija. Vaikka

osallistujan näkemisen merkitys arvioinnissa jakoi arvioijien mielipiteitä, suurempi merkitys arvioinnin näkökulmasta oli sillä, että he havaitsivat osallistujan pyrkivän tuottamaan aktiivisesti puhetta: "Arvioijana huomaan suhtautuvani lähtökohtaisesti positiivisemmin sellaisiin osallistujiin, jotka ovat haastattelussa avoimia, innokkaita, yrittävät puhua paljon jne." (Arvioija 3). Aktiivisuus ja yritteliäisyys ovat aikaisempienkin tutkimusten mukaan (esim. Ahola 2022; May 2006; Lumley 1998) piirteitä, joista arvioijat palkitsevat osallistujia.

Osallistujan näkeminen nähtiin kuitenkin tarpeellisena, vaikka siihen liittyvät riskit tunnistetaankin. Arvioijat toivat esille, että sanaton viestintä on olennainen osa vuorovaikutusta, ja siksi videoitu haastattelu vahvistaa arvioinnin luotettavuutta: "Osallistujan ilmeet, eleet, liikehdintä ja yleisesti läsnäolo ja reagointi vaikuttaa arviointiin. Kokonaiskuva kielitaidosta syntyy vuorovaikutuksessa." (Arvioija 4): "Ilmeet ja eleet tukevat puhuttua. Ovat tärkeä osa kommunikaatiota." (Arvioija 7). Arvioijien huomiot ovat siis linjassa aiemman tutkimuksen kanssa siitä, että vuorovaikutuksellinen, kasvokkainen puhetilanne lisää testattavan kielitaidon laajuutta ja sen myötä testin validiteettia (esim. Nakatsuhara ym. 2020; Galaczi 2010).

Studio-osan ja haastattelun vertailua

Kun arvioijilta kysyttiin, kumpi osio heidän mielestään kertoo osallistujan kielitaidosta paremmin, haastattelu koettiin jossain määrin paremmin kuvaavaksi osaksi:

TAULUKKO 1. KUMPI OSIO KERTOO OSALLISTUJAN KIELITAIIDOSTA MIELESTÄSI PAREMMIN?

Molemmat kertovat yhtä hyvin	7
Haastattelu	9
Studio-osa	0

56 % arvioijista (n = 9/16) on sitä mieltä, että haastattelu kertoo osallistujan kielitaidosta paremmin kuin studio-osuus, ja 43 % arvioijista (n = 7/16) on sitä mieltä, että molemmat osiot kertovat osallistujan kielitaidosta yhtä hyvin. Yksikään arvioija ei kokenut, että studio-osuus kertoo kielitaidosta paremmin kuin haastattelu.

Avovastauksissa monet arvioijat kuvaavat haastattelua luonnollisemmaksi kielenkäyttötilanteeksi, jossa vuorovaikutus on olennaista: "Pidän enemmän haastattelusta, koska se on luonnollisempi ja vuorovaikutteisempi viestintätilanne" (Arvioija 14), "Tässä haastateltava joutuu reagoimaan aidosti oikeassa tilanteessa eikä siihen harjoittelemisen ole mahdollista samalla tavalla kuin studio-osaan" (Arvioija 13). Yksi arvioija myös ajatteli, että studio-osa saattaa aiheuttaa jännitystä, kun taas haastattelun ollessa tavallisempaa kielenkäyttöä osallistujan jännitys voi laueta: "Studioissa osallistuja on testissä, haastattelijan kanssa hän saattaa unohtaa testitilanteen ja samalla jännittäminen ja liika yrittäminen unohtuu" (Arvioija 4).

He, jotka vastasivat molempien osien kertovan kielitaidosta yhtä hyvin, korostivat sitä, että eri osiot mittaavat erilaista kielitaitoa: "Minusta molemmat osat ovat yhtä tärkeitä, sillä niin osallistujat saavat erityyppisen tilaisuuden, joihin liittyy erityyppisiä muuttujia, joissa osoittaa osaamisensa" (Arvioija 10) ja "Eri puhujat ovat eri tilanteissa vahvoilla" (Arvioija 7). Toisaalta yksi tässä ryhmässä ollut arvioija huomautti, että "Haastattelu on kylläkin osallistujalle armollisempi ja lähempänä luonnollista vuorovaikutustilannetta" (Arvioija 6).

Vaikka 9 arvioijaa oli sitä mieltä, että haastattelu kuvaa osallistujien kielitaitoa, vastaajista 69 % (n = 16) oli sitä mieltä, että molemmat osat ovat tarpeellisia ylimmän tason kielitaidon mittaamisessa (taulukko 2). Syyksi annettiin se, että ne mittaavat erilaista kielenkäyttöä, ja siksi kaksiosainen puhumisen osakoe koetaan reilummaksi myös osallistujan kannalta: yhtäältä sen takia, että erilaiset tehtävät sopivat eri osallistujille paremmin, ja toisaalta siitä syystä, että ihmiset saattavat jännittää eri tilanteissa eri lailla: "Molemmat osat tarvitaan, että kaikilla on mahdollisuus osoittaa taitonsa itselleen paremmin sopivassa tilanteessa" (Arvioija 8), "Pidän siitä että on molemmat, koska erilaiset ihmiset kokevat stressiä tai jännitystä eri lailla eri tilanteissa" (Arvioija 13). Lisäksi eräs arvioijain nosti esiin sen, että pelkässä haastatteluun perustuvassa puhumisen osakokeessa haastattelijan vaikutus korostuisi: "Molemmat osat ovat tarpeen, koska haastattelijoita on niin monenlaisia ja heilläkin on hyviä ja huonoja päiviä" (Arvioija 5).

TAULUKKO 2. OVATKO MOLEMMAT OSAT TARPEELLISIA, VAI PYSTYISIKÖ VAIN JOMMANKUMMAN PERUSTEELLA ARVIOIMAAN OSALLISTUJAN KIELITAIKON? MIKSI?

Molemmat tarpeellisia	11
Haastattelu riittäisi	4
Studio-osa riittäisi	1

25 % vastaajista oli sitä mieltä, että pelkkä haastattelu olisi riittävä osallistujan kielitaidon mittaamiseen, mutta heistäkin yksi korosti haastattelijan merkitystä tällaisessa tilanteessa. Toinen arvioija puolestaan pohti, että siinä tapauksessa haastattelun kysymyksiin olisi kiinnitettävä huomiota: "*sopivalla kysymysten asetelulla voisi saada osallistujan osoittamaan melko luotettavasti osaamistaan*" (Arvioija 6).

Arviointikriteerien soveltuvuus tehtäviin

Arvioijat kokevat, että kaikissa tehtävissä arviointikriteereitä pystyy soveltamaan vähintään kohtalaisesti (Likert-skaalan 3). Muita vaikeammaksi koettiin kriteerien soveltaminen tilannetehtäviin. Syitä vaikeuteen mainittiin kaksi: suoritusten lyhyys ja kriteerien soveltumattomuus. Tilannetehtävissä enimmäissuoritus-aika on 30 sekuntia, ja usein vähemmänkin, jolloin arvioitavaa kielen-tuotosta voi olla niukasti. Toisaalta voi olla niin, että tilanne ei välttämättä vaadi kaikkien osakriteerien tason 5 tai 6 kuvausten mukaista kielitaitoa, jolloin arvioijan on tehtävä valinta siitä, mitä osakriteerejä soveltaa tilanteen arvioinnissa: "*Tilanteet ovat kovin lyhyitä ja tarjoavat niukasti materiaalia arvioinnin pohjaksi*" (Arvioija 6), "*Monissa tilannetehtävissä ei voi soveltaa kaikkia kriteerien kategorioita*" (Arvioija 14).

Johtopäätökset

Kyselyn tulokset osoittavat, että molemmat osiot ovat arvioijien mielestä tarpeellisia, jotta osallistujan puhumisen taidosta saadaan mahdollisimman kattava, monipuolinen ja luotettava kuva. Vaikka arvioijat mainitsivat molempaan osaan liittyvän sekä positiivisia että negatiivisia piirteitä ja tekijöitä, kokonaisuutena kahden erilaisen osan yhdistelmä nähtiin kuitenkin hyvänä tapana testata ylimmän tason puhumisen taitoa.

Haastattelulla ja studiokokeella mitataan erilaista kielenkäyttöä, mutta arvioijat eivät pitäneet niitä toistensa kilpailijoina vaan ennemminkin toisiaan täydentävinä osina. Haastattelu testaa arvioijien mukaan testattavan taitoa puhua käsitteellisistä teemoista ja käyttää kieltä osana vuorovaikutusta.

Haastattelun avulla arvioijilla oli mahdollisuus saada kattava kuva osallistujan kielitaidosta, mutta haastatteluun liittyi haastattelijasta ja osallistujan persoonasta johtuvia tekijöitä, jotka voivat vaikuttaa osallistujan saamaan taitotasoarvioon. Vaikka näiden tekijöiden vaikutusta pyritään minimoimaan koulutuksella ja laatimalla haastatteliijoille apukysymyksiä, niiden vaikutusta ei voida täysin poistaa haastattelutilanteesta. Studio-osuus taas on kaikille samanlainen tilanne, mutta sitä pidetään epäluonnollisena puhetilanteena kanavan vuoksi, koska normaalissa arjessa puhumme harvoin itseksemme pelkälle tallenteelle. Lisäksi tehtävissä syntyy haastatteluun verrattuna varsin vähän tuotosta, minkä vuoksi arvioijat eivät voi soveltaa arviointikriteereitä yhtä laajasti kuin haastatteluissa.

Kirjallisuus

- Ahola, S. 2022. *Rimaa hipoen selviää tilanteesta: yleisten kielitutkintojen suomen kielen arvioijien käsityksiä kielitaidon arvioinnista ja suullisesta kielitaidosta*. Jyväskylän yliopisto. JYU Dissertations, 489. <http://urn.fi/URN:ISBN:978-951-39-9005-3>
- Ahola, S., & Tossavainen, H. 2013. Tilannetehtävät Yleisten kielitutkintojen suomen keskitason tutkinnon puhumisen osakokeessa. *AFinLA-e: Soveltavan kielitieteen tutkimuksia*, 5, 5-23. <https://journal.fi/afinla/article/view/8736>
- Ahola, S., Lammervo, T., Neittaanmäki, R. & Ohranen, S. 2013. Comparing speaking situations in three different language tests. *Eesti Rakenduslingvistika Ühingu aastaraamat / Estonian Papers in Applied Linguistics* 9, 7-21. DOI: <http://dx.doi.org/10.5128/ERYa9.01>
- Galaczi, E D 2010. Face-to-face and computer-based assessment of speaking: Challenges and opportunities, in Araújo, L (Ed.) *Proceedings of the Computer-based Assessment (CBA) of Foreign Language Speaking Skills*, Brussels, Belgium, European Union, 29-51.
- Lumley, T. 1998. Perceptions of Language-trained Raters and Occupational Experts in a Test of Occupational English Language Proficiency. *English for Specific Purposes* 17(4), 347-367. <https://doi.org/10.1016/S0889-4906%2897%2900016-1>.
- Luoma, S. 1997. *Comparability of a Tape-Mediated and a Face-to-Face Test of Speaking. A triangulation study*. A Licentiate Thesis. University of Jyväskylä
- May, L. 2006. An examination of rater orientations on a paired candidate discussion task through stimulated verbal recall. *Melbourne Papers in Language Testing*, 29-51. https://arts.unimelb.edu.au/_data/assets/pdf_file/0006/3518682/11_1_3_May.pdf.
- Nakatsuhara, F., Inoue, C., & Taylor, L. 2020. Comparing Rating Modes: Analysing Live, Audio, and Video Ratings of IELTS Speaking Test Performances. *Language Assessment Quarterly*, 18(2), 83-106. <https://doi.org/10.1080/15434303.2020.1799222>
- Opetushallitus 2011. *Yleisten kielitutkintojen perusteet 2011*. Määräykset ja ohjeet 2011:33. Opetushallitus. <https://www.oph.fi/sites/default/files/documents/yleistenkielitutkintojen-perusteet-2011.pdf>
- Stansfield, C., Kenyon, D. M., Paiva, R. & F. Doyle 1990. The Development and Validation of the Portuguese Speaking Test. *Hispania* 73(3), 641-651. DOI:10.2307/343942
- Wigglesworth, G. & O'Loughlin, K. 1993. An investigation into the comparability of direct and semi-direct versions of an oral interaction test. *Melbourne Papers in Language Testing*, 61-71.