

<https://helda.helsinki.fi>

Miten suomen ja ruotsin opettajat käyttäisivät puheen automaattiseen arviointiin kehitettyä työkalua?

von Zansen, Anna

2023-04-17

von Zansen , A & Heijala , M 2023 , ' Miten suomen ja ruotsin opettajat käyttäisivät puheen automaattiseen arviointiin kehitettyä työkalua? ' , AFinLA-teema , Vuosikerta. 2023 , Nro 15 ,
pö Sivut 124 141 . <https://doi.org/10.30660/afinla.124822>

<http://hdl.handle.net/10138/357226>

<https://doi.org/10.30660/afinla.124822>

cc_public_domain

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Mäkipää, T., R. Hilden & A. Huhta (toim.) 2023. Kielenoppimista tukeva arviointi – Assessment for supporting language learning. AFinLA-teema / n:o 15, 124–141.

Anna von Zansen & Maikki Heijala
Helsingin yliopisto

Miten suomen ja ruotsin opettajat käyttäisivät puheen automaattiseen arviointiin kehitettyä työkalua?

Nostot

- Automaattinen palaute tuo uusia mahdollisuuksia puhumisen harjoitteluun.
- DigiTala-työkalun automaattinen palaute konkretisoi taitotasoja ja puheesta mitattavia piirteitä kielten opettajille ja opiskelijoille.
- Analyttinen palaute tuo esiin puhujan vahvuuksia ja kehittämiskohteita.
- Haastatellut opettajat eivät vielä hyödynnä taitotasoasteikkoja suullisen kielitaidon arvioinnissa.

Abstract

In this study, we interviewed four language teachers in order to investigate their perceptions of an automated tool developed for assessing Finnish and Swedish learners' speech. We showed the first version of the Moodle plugin developed by the DigiTala research project to the teachers. In this version, after recording a speech sample, the system provides analytic and holistic feedback to the learner. The automated feedback was developed based on the rating criteria that human raters have used earlier in the project. The data consist of four semi-structured individual interviews conducted in Zoom. The interviews were transcribed and analysed using content analysis. We found that automated feedback has significant implication areas in language learning and teaching. Teachers found the feedback statements understandable and useful, even though they do not cover all dimensions relevant for the speaking construct. The results will benefit those working on automated assessment of oral language skills.

Keywords: automated assessment, language assessment, oral language skills, automated feedback

Asiasanat: automaattinen arviointi, kielitaidon arviointi, suullinen kielitaito, automaattinen palaute

1 Johdanto

Tämä soveltava tutkimus käsittelee automaattista palautetta kielitaidon formatiivisen arvioinnin kontekstissa. Automaattinen palaute on yleistynyt monenlaisissa oppimisympäristöissä (Deeva ym. 2021), joista esimerkkinä ovat kielten opiskelun virtuaaliset tutorit (*tutoring systems*, Golonka ym. 2014). Tutkimus liittyy monitieteiseen DigiTala-hankkeeseen, jossa kehitetään puheentunnistusta (*automatic speech recognition*, ASR) ja automaattista arviointia (*automatic scoring*) hyödyntävää verkkotyökalua suomen ja ruotsin puhumisen harjoitteluun (Kautonen & von Zansen 2020). Puheentunnistuksen ja automaattisen arvioinnin lisäksi työkalun (von Zansen ym. 2022a) kolmas keskeinen ominaisuus liittyy puhujalle näytettävään automaattiseen palautteeseen (*automated feedback*, ks. katsaus automaattiseen palautteeseen oppimiskonteksteissa Deeva ym. 2021, suullisen kielitaidon osalta Evanini & Zechner 2020; Gu & Davis 2020). Työkalun kehityksessä hyödynnetään foneettista (esim. Kallio ym. 2022; Kallio ym. arvioitavana), kielididaktista (esim. von Zansen ym. 2022b; von Zansen & Hildén 2022; von Zansen ym. 2022c) sekä kieliteknologiaan (esim. Al-Ghezi ym. 2021; Al-Ghezi ym. arvioitavana) liittyvää tutkimusta.

Työkalua voitaisiin käyttää osana formaalia koulutusta, mutta myös sen ulkopuolella, kun kielen oppijat harjoittelevat puheen tuottamista itsenäisesti. Formatiivisen arvioinnin lisäksi työkalu voisi toimia ihmisarvioinnin tukena isoissa ja tärkeissä kielikokeissa (ylioppilastutkinnon osalta ks. Vaarala ym. 2021). Työkalun ensisijaisena kohderyhmänä ovat suomalaiset lukiolaiset, jotka opiskelevat suomea tai ruotsia toisena tai toisena kotimaisena kielenä. Myöhemmin työkalun kohderyhmää ja kohdekieliä on tarkoituksena laajentaa – esimerkiksi akateemisten maahanmuuttajataustaisten

aikuisopiskelijoiden (n=115) suhtautuminen tietokoneella suoritettavaan puhumisen kokeeseen on ollut positiivista (von Zansen & Hildén 2022).

Puheen automaattiseen arviointiin perustuvat työkalut pystyvät tunnistamaan vain tiettyjä automaattisesti mitattavissa olevia piirteitä suullisesta kielitaidosta (Evanini & Zechner 2020; Xi 2010), joten tässä artikkelissa käytämme termejä puhuminen ja puheen arvioiminen. Kielten oppimisen kannalta puheen automaattinen arviointi nähdään siihen liittyvistä rajoituksista huolimatta lisäresurssina (Gu & Davies 2020: 159), sillä se lisää suullisen kielitaidon harjoittelumahdollisuuksia tuomalla joustoja ajan ja paikan suhteen. Lisäksi automaattinen arviointi yhdenmukaistaa arviointia ja säästää resursseja – kone arvioi väsymättä kaikki puhujat samalla tavalla (Evanini & Zechner 2020: 3).

Artikkelissa tutkimme kieltenopettajien ensivaikutelmia työkalusta (von Zansen ym. 2022a) haastatteleamalla neljää kieltenopettajaa. Arvioinnin validiustarkastelua on syytä rajata, sillä erilaiset validiusrakenteet ovat keskeisiä riippuen arviointikon-tekstistä (Xi 2010: 293). Tämän tutkimuksen lähtökohtana on formatiivinen arviointi, jossa oppija saa työkalun antaman analyttisen eli diagnostisen palautteen (*diagnostic feedback*) avulla tietoa vahvuuksistaan ja kehittämiskohteistaan, mikä puolestaan tukee oppijan suullisen kielitaidon kehittymistä ja kieliminän rakentumista. Analyttisen palautteen lisäksi työkalu antaa holistisen arvion puhujan taitotasosta. Jotta palautteesta olisi hyötyä, automaattisen palautteen tulisi kuitenkin olla oppijalle ymmärrettävää, kannustavaa, ja se tulisi saada sopivassa muodossa oikeaan aikaan.

2 Tutkimuksen tausta ja tutkimuskysymykset

Tämän soveltavan kielentutkimuksen alaan kuuluvan artikkelin tavoitteena on tuottaa uutta tietoa automaattisen arvioinnin ja palautteen hyödyntämismahdollisuuksista suullisen kielitaidon harjoittelussa ja opetuksessa (ks. tutkimuskysymykset 1 ja 2 myöhemmin tässä luvussa). Kuten soveltavalle tutkimukselle on luonteenomaista, osallistamme opettajia työkalun suunnitteluun näyttämällä ensimmäistä versiota kehittämästämme työkalusta. Tutkimuksessa yhdistämme sekä laadullisen haastattelututkimuksen (Hyvärinen ym. n.d.) että käytettävyydestauksen (Ovaska ym. 2005) menetelmiä. Käyttäjiltä kerätty palaute kuuluu keskeisesti erilaisten työkalujen ja palvelujen kehitys- ja validointityöhön.

Automaattisia puhumisen arviointityökaluja on olemassa englannin kielelle, mutta näissäkin järjestelmissä automaattinen palaute rajoittuu yleensä ääntämiseen (Golonka ym. 2014). Poikkeuksena voidaan kuitenkin mainita ETS:n SpeechRater (Gu & Davis 2020), Cambridge Englishin Linguaskill (Xu ym. 2020) -työkalut ja sekä Pearson Educationin (2018, 2022) Versant English ja Versant Arabic -testit. Lisäksi automaattista palautetta on hyödynnetty esimerkiksi hollannin sanajärjestyksen harjoittelussa (de Vries ym. 2014). Käyttäjiä haastatteleamalla saadaan konkreettisia kehitysehdotuksia,

joiden perusteella palautenäkymää voidaan muokata (Gu & Davies 2020). Käyttäjien näkemykset voivat myös auttaa tarkastelemaan laajemmin työkalun käyttökokemusta (de Vries 2014) sekä arvioimaan, missä määrin puhumisen tehtävät vastaavat arkielämän viestintätilanteita (Xu ym. 2020).

Suomalaisessa DigiTala-hankkeessa olemme keränneet suomen (n=308) ja ruotsin (n=178) puhujilta puheaineistoa (yli 3500 puhenäytettä, yhteensä noin 20 h puhetta). Puhenäytteiden ja ihmisiltä kerättyjen arviointien avulla olemme kehittäneet automaattista arviointia. Aiemmin olemme tutkineet eri käyttäjäryhmien suhtautumista puheen automaattiseen arviointiin sekä heidän näkemyksiään arviointikriteereistä ja tehtävistä sekä heidän kokemuksiaan puhetehtävien suorittamisesta tietokoneella (von Zansen ym. 2022b, von Zansen ym. 2022c, von Zansen & Hildén 2022). Olemme muun muassa havainneet, että opiskelijat (n=58) tuntevat tekoälyn mahdollisuuksia ja rajoituksia sekä kykenevät pohtimaan automaattiseen arviointiin liittyviä hyötyjä ja uhkia (von Zansen ym. 2022b). Puheen analyttinen arviointi tutkimushankkeen kriteereillä (von Zansen 2022) on osoittautunut arvioijille (n=37) haastavaksi etenkin liian lyhyistä puhenäytteistä (von Zansen ym. 2022c). Kouluttamamme arvioijat ja laatimamme arviointikriteerit ovat kuitenkin osoittautuneet luotettaviksi (von Zansen & Huhta 2022). Sekä opiskelijat että ihmisarvioijat pitävät vuorovaikutuksellisuutta tärkeänä ominaisuutena puhumisessa (von Zansen ym. 2022c, von Zansen & Hildén 2022), vaikka sen mittaaminen automaattisesti on haastavaa (Evanini & Zechner 2020; Xi 2010). Automaattiseen arviointiin liittyvien rajoitteiden takia sen nähdään soveltuvan etenkin ihmisarvioinnin rinnalle (hybridiarvioinnista Evanini & Zechner 2020: 11). Kuten seuraavassa luvussa kerrotaan tarkemmin, saimme kevään 2022 aikana luotua ensimmäisen version automaattisesta puheen arviointijärjestelmästä Moodleen. Toukokuussa 2022 esittelimme työkalua haastatellen sekä opettajia että opiskelijoita, ja tässä artikkelissa käsittelemme ensimmäistä palautenäkymän versiota opettajien näkökulmasta.

Tutkimus perustuu Xin (2010) laajempaan automaattiseen arviointia koskevaan validiusmalliin, jossa automaattisen palautteen (Xi 2010: 296) validiutta voidaan tarkastella sen arviointia (*evaluation*), yleistämistä (*generalization*), selittämistä (*explanation*), sekä käyttöä (*utilization*) koskevien kysymysten avulla. Tämän tutkimuksen kannalta keskeiset validiusnäkökulmat liittyvät kahteen jälkimmäiseen, nimittäin automaattisen palautteen seurauksiin (käyttö ja selittäminen, Xi 2010). Pyrimme tarkastelemaan automaattisen palautteen tuomia hyötyjä oppimisen ja puhumisen kehittymisen kannalta sekä arvioimaan, onko automaattisella palautteella positiivinen vaikutus opetukseen ja oppimiseen (Xi 2010: 296).

Tässä pilottivaiheen artikkelissa keskitymme kieltenopettajien näkemyksiin työkalun ensimmäisestä versiosta. Vastaamme kahteen tutkimuskysymykseen, joista ensimmäisen avulla voimme tehdä johtopäätöksiä automaattisen palautteen käytöstä ja pohtia arvioinnin mahdollisia vaikutuksia. Lisäksi tutkimme, kokevatko opettajat puhe-suorituksista annettavan automaattisen palautteen hyödylliseksi oppimisen

kannalta (tutkimuskysymys 2). Toisin sanoen tutkimme, ovatko hankkeen laatimien arviointikriteerien (von Zansen 2022) pohjalta tuotetut palauteväittämät kieltenopettajien mielestä ymmärrettäviä ja hyödyllisiä. Tutkimuskysymykset ovat:

1. Millaisia mahdollisuuksia ja rajoituksia haastatellut kieltenopettajat näkevät työkalun käytölle?
2. Miltä osin haastatellut kieltenopettajat pitävät työkalun antamaa analyttistä ja holistista palautetta hyödyllisenä?

Tämä laadullinen tutkimus tuottaa uutta tietoa automaattisen palautteen ensimmäisestä versiosta. Kieltenopettajien näkemykset työkalusta ovat keskeisiä oppijoille suunnatun automaattisen palautteen kehityksessä, sillä niiden tutkiminen lisää ymmärrystä siitä, millaista automaattista palautetta puhumisen suorituksista kannattaisi antaa. Tutkimustiedon avulla palautetta voidaan muotoilla tarkoituksenmukaisempaan ja ymmärrettävämpään muotoon, jotta se hyödyttäisi mahdollisimman erilaisia oppijoita ja olisi linjassa opetuksen tavoitteiden ja sisältöjen kanssa.

3 Aineisto ja menetelmät

Tutkimushanke toteutti Moodle-liitännäisen (*plugin*) yhteistyössä Helsingin yliopiston tietojenkäsittelytieteen opiskelijoiden kanssa (von Zansen ym. 2022a). Tässä artikkelissa analysoidaan neljän kieltenopettajan puolistrukturoidut haastattelut järjestettiin toukokuussa 2022 Zoom-videopuhelun avulla. Seuraavissa alaluvuissa esittelemme tutkimuksen aineiston ja menetelmät. Luvussa 3.1 esittelemme Moodlessa toimivaa työkalua, luvussa 3.2 tutkimuksen osallistajat, luvussa 3.3 haastattelujen toteutuksen ja lopulta luvussa 3.4 aineiston analyysimenetelmät.

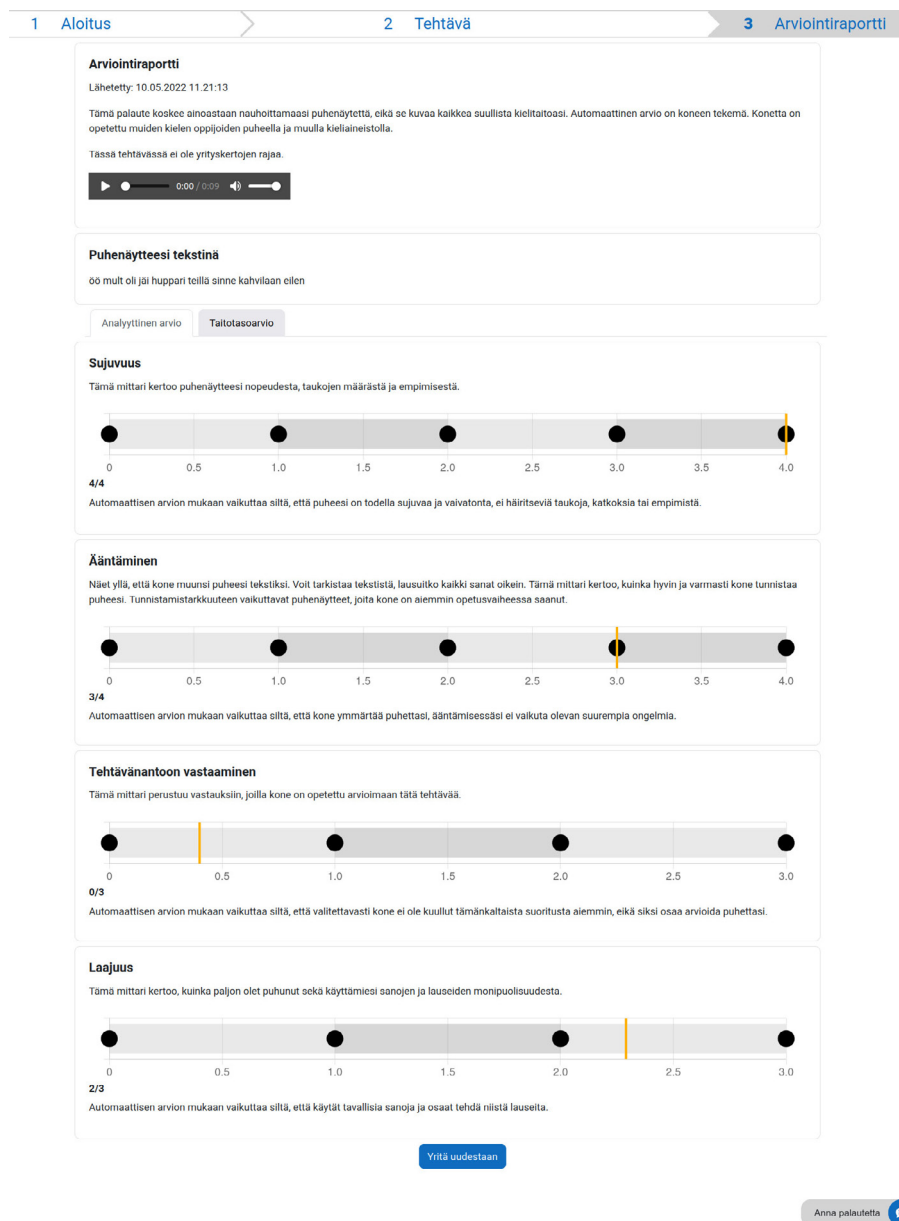
3.1 Palautetyökalu Moodlessa

Selaimessa toimivan Moodle-työkalun (von Zansen ym. 2022a) avulla oppija äänittää ensin puhe-suorituksensa, joka arvioidaan Aalto-yliopiston palvelimilla automaattisesti, minkä jälkeen työkalu esittää automaattisen palautteen oppijalle. Työkalun ominaisuuksia esitellään tarkemmin käyttöohjeessa (Alanen ym. 2022), ja automaattisen puheen tunnistuksen ja arvioinnin kehitystä käsittelevät Al-Ghezi ym. (arvioitavana).

Ensimmäisessä vaiheessa opiskelijoille annettava automaattinen palaute muotoiltiin arviointikriteereistä (von Zansen 2022), joita hankkeeseen osallistuneet ihmisarvioijat ovat käyttäneet automaattisen arvioinnin kehitysvaiheessa. Näitä kriteerejä verrattiin puheesta automaattisesti arvioitaviin piirteisiin (Al-Ghezi ym. arvioitavana), minkä pohjalta palauteväittämät muotoiltiin. Holistisen taitotasoarvion lisäksi automaattista palautetta annetaan opiskelijoille neljästä analyttisestä ulottuvuudesta, jotka

129 MITEN SUOMEN JA RUOTSIN OPETTAJAT KÄYTTÄISIVÄT PUHEEN AUTOMAATTISEEN ARVIOINTIIN KEHITETTYÄ TYÖKALUA?

ovat tehtävän suorittaminen, sujuvuus, ääntäminen sekä ilmaisun laajuus. Sanallisen palautteen lisäksi keltaisen viivan avulla visualisoidaan oppijan puhe-suorituksen sijoittuminen kullekin skaalalle (analyttisista skaaloista ks. myös von Zansen & Huhta 2022), ja skaalan portaiden kuvaukset tulevat näkyviin viemällä hiiri kunkin mustan pallon päälle (ks. kuvio 1). Lisäksi työkalun avulla opettaja tai tutkija voi lisätä oman arvionsa, kommentoida äänitettyny puhenäytettä tai automaattista palautetta sekä ladata äänitetyt puhenäytteet ja arviointitiedot (Alanen ym. 2022).



KUVIO 1. DigiTala-työkalun palautenäkymä.

3.2 Osallistujat

Tutkimuksen otantamenetelmänä käytettiin harkinnanvaraista näytettä, sillä tutkimuksen tavoitteena oli selvittää automaattisen palautteen käyttötarkoituksia ja hyödyllisyyttä kielten opetuksessa (ks. luku 2). Kuten laadullisessa tutkimuksessa on tavallista, osallistujat valittiin tarkoituksenmukaisesti edustamaan keskeisiä käyttäjäryhmiä eli sekä suomen että ruotsin opettajia eri kouluasteilta. Haastateltavat valittiin lähettämällä haastattelukutsu 20:lle 60:stä DigiTala-projektiin aiemmin osallistuneesta opettajasta, joiden tiesimme työskentelevän opettajana tällä hetkellä. Heistä tutkimusta varten haastateltiin neljää kieltenopettajaa, joihin viittaamme artikkelissa jatkossa tunnuksilla H1, H2, H3 ja H4. Osallistujista yksi opetti haastatteluhetkellä ruotsia (H2) ja kolme suomea (H1, H3, H4) eri koulutusasteilla yläkoulusta yliopistokursseihin. Haastatteluun osallistuneista opettajista yhdellä (H4) oli opetuskokemusta opetusaineessaan alle viisi vuotta ja muilla noin 20–30 vuotta. Haastatellut opettivat kotimaisten kielten eri oppimääriä: H1 finskan A-, B- ja mofi (*modersmålsinriktad*)-oppimääriä, H2 A- ja B-ruotsia yläkoulussa ja lukiossa, H3 suomi äidinkielenä ja kirjallisuus - ja suomi toisena kielenä -oppimääriä lukiossa sekä IB-lukion Finnish B -oppimäärää kahdella tasolla, H4 suomi toisena ja vieraana kielenä -oppimäärää yliopistossa.

3.3 Haastattelujen toteutus

Haastattelut toteutettiin Helsingin yliopiston Zoom-videoneuvottelutyökalulla toukokuun 2022 loppupuolella. Puolistrukturoidut haastattelut kestivät 50–65 minuuttia, ja haastattelun aikana haastattelijä esitteli opettajille DigiTala-työkalua (versio 0.1.40) Moodlella (versio 4.0) näytönjaon avulla. Opettajilta kerättävät taustatiedot minimoitiin, mikä helpotti laadullisen aineiston anonymisointia (Yhteiskuntatieteellinen tietoarkisto n.d.). Haastattelijä toi kaikissa haastatteluissa ilmi, että hän on ollut mukana rakentamassa haastattelun keskiössä olevaa Moodle-työkalua, ja pyrki tietoisesti vaikuttamaan haastateltavien vastauksiin mahdollisimman vähän.

Haastattelukysymykset (Heijala & von Zansen 2022) laadittiin kuuden aihepiirin ympärille. Tässä artikkelissa käsittelemme haastattelun osia I taustatiedot, III ensivaikutelmat palautenäkymästä, IV analyttinen palaute, V holistinen palaute ja VI työkalun ja automaattisen arvioinnin käyttömahdollisuudet. Haastattelun osa II käsitteli menetelmiä, joita opettajat tällä hetkellä käyttävät puhumisen opetuksessa ja arvioinnissa, ja rajasimme tämän osan artikkelin ulkopuolelle.

Hyödynsimme haastattelukysymysten laadinnassa aiempaa kokemustamme tutkijoina ja kieltenopettajina sekä tutkimuskirjallisuutta. Kysymysten laadinnassa kiinnitimme huomiota menetelmään liittyviin ongelmiin (ks. esim. Hyvärinen ym. n.d.). Pyrimme esimerkiksi välttämään liian monen asian kysymistä samalla kertaa ja jätimme tilaa haastateltavan vastauksille. Ennen aineistonkeruuta kokeilimme haastattelurunkoa projektitiimin opettajataustaisten tutkijoiden kesken. Pilotoinnin

perusteella muutimme esimerkiksi vaikeaselkoisia kysymyksiä ja järjestimme haastattelurungon kysymykset uudelleen.

Ennen haastattelua osallistujille lähetettiin työkalusta lyhyt esittelyvideo, jonka katsominen oli vapaaehtoista. Haastatteluissa keskityttiin palautenäkymään, jonka oppija näkee nauhoitettuaan äänitteen tuottamistehtävään. Haastattelija oli valmistellut näkymän etukäteen nauhoittamalla lyhyen puhe-suorituksen tehtävään, jossa puhujan tulee soittaa kahvilaan, jonne hän on jättänyt hupparin edellispäivänä. Etukäteen nauhoittamisella vältettiin automaattiseen arviointiin liittyvät odotusajat tai tekniset ongelmat, joita haastattelutilanteessa tehdyssä nauhoituksessa olisi voinut ilmetä. Palautenäkymässä analyttiset arviot näkyvät yhdellä välilehdellä ja holistinen taitotasoarvio toisella (ks. kuvio 1, luku 3.1).

3.4 Analyysimenetelmät

Haastattelut tallennettiin Zoom-videoneuvottelutyökalun avulla tutkijan tietokoneelle, jonka jälkeen litteroimme ne peruslitteroinnin (Yhteiskuntatieteellinen tietoarkisto 2022) periaatteita noudattaen. Esimerkiksi aiheeseen liittymättömät keskustelut jätettiin litteroimatta, ja haastattelijan selostamat käyttöliittymän toiminnot referoitiin. Seuraavaksi yhdistimme neljän haastattelun litteraattit haastattelun kuuden osan perusteella. Taustatietoja lukuun ottamatta kunkin osan (osat II–VI, ks. Heijala & von Zansen 2022) aineisto vietiin tekstitiedostona Atlas.ti -ohjelmaan (versio 22.0), jonka avulla toteutimme sisällönanalyysin (Vuori n.d.). Sisällönanalyysissa hyödynsimme myös aiempaa kokemustamme kieltenopettajina sekä aiempaa tutkimusta. Ensin koodasimme Atlas.ti:n avulla opettajien vastauksia 27 luokkaan, mikä auttoi hahmottamaan vastausten yhtäläisyyksiä.

Opettajahaastattelut olivat sisällöltään melko laajoja (artikkelista pois rajatun osan II litteraatti 16 sivua, osat III–VI yhteensä 26 sivua) ja vastaukset haastattelijan kysymyksiin perusteellisia, joten Atlas.ti:n tarjoama dekonstruoiva tyypittely ei tässä vaiheessa vaikuttanut jatkoanalyysin kannalta enää hedelmälliseltä. Havaitimme, että eri opettajien kommentit rinnakkain kysymyskohtaisesti analysoituna tarjoaisivat paremman kokonaiskuvan heidän näkemyksistään. Seuraavaksi etenimme sisällönanalyysissa tutkimuskysymysten ohjaamina. Nämä uudelleen ryhmitellyt litteraattit tiivistettiin luetteloiksi, jotka sisälsivät vastausten keskeisen sisällön, minkä perusteella voitiin luoda yleiskuva opettajien näkemyksistä. Lopulta jäsensimme sisällönanalyysin tulokset ja valitsimme aineistosta kuvaavia esimerkkejä havaintojemme tueksi.

4 Tulokset

Esittelemme seuraavissa alaluvuissa tulokset tutkimuskysymysten mukaisessa järjestyksessä. Alaluku 4.1 käsittelee haastateltavien tunnistamia mahdollisuuksia ja

rajoituksia työkalun käytössä. Luvussa 4.2 kuvaamme haastateltavien näkemyksiä analyttisen ja holistisen palautteen hyödyllisyydestä.

4.1 Työkalun mahdollisuudet ja rajoitukset

Haastattelujen perusteella automaattinen arviointi tuo lisäresurssin puhumisen harjoitteluun itsenäisesti tai opettajan johdolla. Opettajat käyttäisivät työkalua esimerkiksi kotitehtävissä, osana kieliprofilia sekä arvioinnin tukena (lainaus 1). Lisäksi työkalu soveltuisi osaksi suullisia kokeita. Työkalun pelillisuus voisi myös lisätä oppijan motivaatiota (lainaus 2).

- (1) – – mahdollisuus siis siinä [lukion] suullisessa syventävässä kurssissa hyödyntää tätä. – – Silloinhan sen vois nauhottaa ja käyttää Moodlea ja silloin ne sais palautetta ja opettaja sais nähdä sen palautteen ja tää auttais sitten meitä opettajia siinä arvioinnissa. H1
- (2) – – tietynlainen pelillisuus ja semmonen motivaatiotekijä joillekin eli harjotella sitä ja yrittää päästä vaikka sitten seuraavalle tasolle ja kokeilla mitkä asiat siellä se kone nyt sitten huomaa – – H2

Tunnistetut haasteet puolestaan liittyivät työkalun ja oppimisympäristöjen yhteensopivuuteen, tietosuojaan ja teknologiavastaisuuteen (lainaus 3 ja lainaus 4). Opettajatkaan eivät luottaisi pelkkään automaattiseen arviointiin isoissa ja tärkeissä tehtävissä (lainaus 5). Myös mahdolliset kielteiset vaikutukset opiskelijan itsetuntoon huolettivat opettajia (lainaus 6).

- (3) – – helppo tapa tallentaa – – miten tää on siirrettävissä sitten nää näytteet jonnekin muille alustoille – – H2
- (4) Kaikki opiskelijat ei ehkä pidä tämmösestä, meil on siis tämmösiä tietotekniikkavastaisiakin opiskelijoita, kumma kyllä, mutta mä uskon että useimmat kyllä tekisivät innostuneesti. H3
- (5) – – edut on tietysti tää palaute, jonka opiskelija saa, ja se tietysti auttaa myös opettajaa – – kielteinen voi olla, et ne luottaa liikaa siihen koneeseen – – ei saa jäädä vaan sen koneen varaan se arviointi – – isompi tehtävä tai tärkeämpi tehtävä, silloin myös opettaja joku antaa suullista palautetta. H1
- (6) Ehkä just se että se ei oo sataprosenttisesti luotettavaa tää palaute ja jos se toistuvasti sanoo jotain että sä et oo hyvä – – jos opiskelija yrittää ja yrittää ja saa jotain ei-niin-hyvää palautetta, niin ehkä siitä vois olla haittaakin. H4

133 MITEN SUOMEN JA RUOTSIN OPETTAJAT KÄYTTÄISIVÄT PUHEEN AUTOMAATTISEEN ARVIOINTIIN KEHITETTYÄ TYÖKALUA?

Haastateltavat toivat lisäksi esiin arvioinnin läpinäkyvyyteen liittyvän tarpeen ymmärtää, mistä automaattinen arvio koostuu (lainaus 7). Automaattisen arvioinnin ymmärtäminen on lähtökohta arvioinnin luotettavuuden ja mahdollisten haasteiden pohtimiselle. Esimerkiksi puhenäytteiden lyhyys (lainaus 8) ja erilaisten puhujien huomioiminen (lainaus 9) huolettivat, kuten olemme havainneet muissa tutkimuksissamme (von Zansen ym. 2022b, von Zansen ym. 2022c).

- (7) Mun on vaikee kun mä en tiedä mitä kaikkea se kone osaa [aurua], mitä sinne on syötetty, minkälaisia asioita se pystyy nyt sit sanoo eri tasoille – – H2
- (8) Kun nää näytteet on aika lyhyitä, aika usein tulee sellanen tilanne että se taitotaso-arvio se ei oo ihan oikea – – että siis tää ei oo ehkä luotettavin mittari. H4
- (9) Mä mietin vaan opiskelijoita, joilla on jotain puhevikaa tai jotain, mitä silloin tapahtuu – – miten tarkka tää on tää apuväline nyt sitten niitä ajatellen – – [joilla] on joku ärrävika tai jotain, onks tässä jotain mitä pitäis vielä miettiä? H1

Työkalan ulkonäkö herätti haastateltavissa ajatuksia, ja esimerkiksi värikyyttä toivottiin käyttöliittymään. Opettajat tuntevat kuitenkin saavutettavuusvaatimuksia ja pitivät tärkeänä työkalun soveltuvuutta eri käyttäjäryhmille (lainaus 10). Analyttisissä palauteväittämissä skaalojen eripituisuus kiinnitti huomiota (lainaus 11). Lisäksi haastateltavat toivat esiin tarpeen muuttaa työkalun sisältöä vastaamaan omiin tarpeisiinsa (lainaus 12). Palautteen näyttämistä eri kielillä (suomeksi, ruotsiksi tai englanniksi) pidettiin tärkeänä, ja joitakin palauteväittämiä voisi yksinkertaistaa (lainaus 13).

- (10) – – on aika iso prosentti ihmisiä jotka eivät näe värejä eli ovat värisokeita – – H4
- (11) – – se laajuus, siis oliko se kolme, se on maksimi? Eli näissä ylemmissä se oli nelonen, maksimi oli nelonen, skaala on vähän laajempi näissä. H4
- (12) Pystyykö opettaja vaihtaa tätä asteikkoa mitenkään? H3
- (13) – – semmoselle jolla on pienempi kielitaito nää tekstiosuudet voi olla pikkusen hankalia, esimerkiksi nyt tämmönen sana ku empimisestä. Sitä nyt ei ymmärrä ihan kaikkea. H3

Kaiken kaikkiaan haastateltavat pitivät koneen tarjoamaa analyttistä arviota kattavana. Opettajat toivat kuitenkin esiin joitakin puheen ulottuvuuksia, jotka nyt jäävät automaattisen arvioinnin ulkopuolelle. Esimerkiksi laajuuden lisäksi tulisi arvioida puheen oikeellisuutta (lainaus 14), ja vuorovaikutusta voitaisiin arvioida virtuaalisen keskustelukumppanin kanssa simuloitussa keskustelussa (lainaus 15). Puheen sisällön

ja luovuuden automaattinen arviointi tunnistettiin haastavaksi (lainaukset 16 ja 17). Tällä hetkellä työkalu ei myöskään arvioi nonverbaalista viestintää, kuten kehonkieltä (lainaus 18).

- (14) – – tää ei ota nyt ihan vielä kantaa sitten siihen tavallaan oikeellisuuteen. – – se nyt ei oo se ensimmäinen asia mitä me ruvetaan tarkastelemaan että onks kaikki rakenteet oikein, mutta kyllähän se yks osa tätä on. H2
- (15) – – joku videolla – – sinulta kysyy et hei hei mistäs olet kotoisin ja sitten on tietty aika vastata – – siellä olis tämmönen video, joka antaa sen oman repliikin jälkeen sitten tietyn vastausajan. H2
- (16) – – täällä on tää tehtävänantoon vastaaminen, se sisältöpuoli ei tässä tietenkään näy – – H3
- (17) – – videoissa, joita meillä oli käytössä nyt keväällä, arvioidaan myös kuinka luova opiskelija on. – – Tässä ei saa olla niin luovaa. – – mutta miten kone voi arvioida luovuutta? H4
- (18) – – kun puhutaan kommunikoinnista, niin siinä on muitakin tekijöitä, jotka on tärkeitä. Et opettaja näkee niin paljon muuta sit kuitenkin. Et nehän tulee nää kaikki kehonkieli ja kaikki muut [naurua]. Tää [koneen palaute] on aika raakaa, jos sanotaan näin. H1

Haastateltavat näkivätkin automaattisen arvioinnin soveltuvan etenkin opettajan työn tueksi, eivät sitä kokonaan korvaavaksi. Haastateltavat pitivät tärkeänä, että opettajat voivat kommentoida automaattisia arvioita ja tarjota puhujalle monipuolisempaa palautetta (lainaukset 19 ja 20).

- (19) – – kone antaa omat arvionsa ja sitten vielä kun mä korjaan ja kuuntelen näitä niin sitten mä vielä annan palautetta. H1
- (20) – – mun mielestä se on tärkeätä et opettajalla on se mahdollisuus lisätä tai jotenki laajentaa tätä koneen antamaa palautetta. H4

4.2 Analyyttisen ja holistisen palautteen hyödyllisyys

Automaattinen analyttinen palaute koettiin pääosin hyödylliseksi ja selkeäksi, joskin palautteen kannustavuus mietitytti haastateltavia (lainaukset 21 ja 22). Opettajien oli kuitenkin jossain määrin vaikea ottaa kantaa siihen, missä määrin puheesta annettavat

**135 MITEN SUOMEN JA RUOTSIN OPETTAJAT KÄYTTÄISIVÄT PUHEEN
AUTOMAATTISEEN ARVIOINTIIN KEHITETTYÄ TYÖKALUA?**

analyttiset arviot vastaavat opetusta tai heidän käsitystään suullisen kielitaidosta. Haastateltavat vaikuttivat mieltävän puhumisen osa-alueet eri tavalla. Esimerkiksi lainauksessa 24 opettaja kuvaa sujuvuutta enemmän ymmärrettävyyden kuin puheen taukoisuuden (lainaus 23) kannalta. Lisäksi kuten H1 kertoo, eritasoisten puhujien kohdalla saatetaan kiinnittää huomiota erilaisiin asioihin (lainaus 24). Puheen näyttämistä tekstinä pidettiin hyödyllisenä oppimisen kannalta (lainaus 25).

- (21) – – lähes jokainen opiskelija haluaa tietää yksityiskohdista eli mikä onnistui ja mikä ei onnistunut – – voiko koneelle opettaa että se poimisi sieltä näitä onnistuneimpia kohtia? H4
- (22) – – joku opiskelija ehkä vois – – miettiä, et oonks mä todellakin näin huono vai ja miettiä tätä palautetta. Et silloin voi olla, että opettajan palaute tulee tarpeen. H1
- (23) Mun mielestä niiden [analyttisten palautteiden] ei tarvitse olla pitempiä. Mut se on näitä keskeisiä asioita joita [sujuvuudessa] arvioidaan, et se on sujuvaa, vaivatonta, ei häiritseviä taukoja, ei empimistä. H4
- (24) – – oliko se sujuvaa, – – ymmärsikö kohderyhmä mistä sä puhuit, menikö sanoma perille, sehän on tietysti sitä juuri, et se tehtävänanto ja se sujuvuus, ja tietysti sitten kun et puhutaan, no se laajuus on tietysti aina mukana jossain määrin, et jos mä vaadin tietysti enemmän sitten just esimerkiksi mofeilta eli kakskielisiltä. H1
- (25) Opiskelija itsekin voi [litteraatista] katsoa, mitä mä halusin sanoa ja mitä tästä tuli. H4

Keskustelu holistisesta taitotasoarviosta toi esiin ristiriidan: kielten opetussuunnitelmat (Opetushallitus 2003, 2019) perustuvat taitotasoihin, mutta opettajat eivät vielä hyödynnä niitä opetuksessa tai arvioinnissa (lainaukset 26–28). Sen sijaan opettajat kertovat hyödyntävänsä numeerisia asteikkoja, joten opiskelijat eivät välttämättä tunne taitotasoasteikkoa entuudestaan.

- (26) Tarkotus olis [käyttää taitotasoasteikkoa opetuksessa], jos katsotaan nyt opetussuunnitelmaa taas [naurua] niin siinä on ne päämäärät. Mut mun täytyy kyl myöntää, et en mä hirveesti niitä kato. H1
- (27) Kyllähän ne aina taitotasot esitellään ja etenki tuolla toisella asteella lukiopuolella – – päättöarviointia [tehdessä] eli yhdeksännen luokan päättöarvioinnin kriteerit jotka nyt sit pohjautuu taitotasoihin – – mä kouluarvosanoin edelleen ja näin arvioin koheet ja yleisesti, et en mä nyt taitotasoja tosiaan käytä sitten ku tietyissä tilanteissa vaan. H2

- (28) Ei, opiskelijat ei tiedä näitä [taitotasoa], [eurooppalaisesta] viitekehystä tähän nää tulee, eikö tuukin? Niitä ei kyllä lukiomaailmassa käytetä missään, voi sanoo näin. – Et meillä on sit numeerinen arvostelu ja IB:llähän on siis ykkösestä seiskaan, siel on semmonen asteikko. H3

Lukion suullisen kielitaidon kokeen uudistuessa (Opetushallitus 2022) opettajien kuitenkin tulee antaa numeroarvosanan lisäksi taitotasoarvio, joten DigiTalan työkalu voisi konkretisoida taitotasoa niin opiskelijoille kuin opettajille (lainaukset 29 ja 30).

- (29) – – nythän [lukion suulliseen kokeeseen] tuli semmonen lisäys, tähän mennessä opiskelija on saanu arvosanan kasin, ysin, ihan tämmösen kouluarvosanan, nyt meidän pitää vielä juuri lisätä siihen vielä juuri tää taitotasoarvio. Eli tulee se kasi ja sitten vielä taitotasoarvio siihen. – – Silloin tää olis todella hyvä apuväline juuri siinä suullisessa. H1
- (30) – – tää olis opettajan työn tuki sitten, arvioinnin tuki, varsinkin sit jos ei oo itsellä kovasti arviointikokemusta – – H2

Haastateltavat eivät kokeneet ongelmallisena taitotasojen näyttämistä päätasolla (B1) välietappien (B1.1) sijaan. He tunnistivat, että välietappien määrittäminen on ihmisarvioijillekin haastavaa, ja toivat esiin opettajien mahdollisuuden täydentää automaattista arviota (lainaukset 31 ja 32).

- (31) No se voi olla ihan hyvä [ettei taitotasoiissa näytetä välietappeja kuten B1.1], koska joskus kun on kattonu noita [taitotasoa], niin vähän miettii itsekin, et [naurua] mikä sitten on se ero. – – nää taitotasoarviot voi olla aika vaikee monesti. H1
- (32) – – tässä voi aatella sitten varmaan että se opettaja määrittelee sen alatasen. Että jos tän on tarkoitus tukea sitä opettajan arviota päätason osalta niin voihan sitä sitten täydennellä. H2

5 Pohdinta

Tämä soveltava tutkimus käsittelee kieltenopettajien näkemyksiä suomen ja ruotsin oppijoille kehitetystä työkalusta, joka antaa oppijalle automaattisesti palautetta hänen puheasuorituksestaan. Automaattista arviointia ja palautetta hyödynnetään enenevässä määrin etenkin matemaattis-luonnontieteellisten ja teknisten alojen – mutta myös kielten – opetuksessa (Deeva ym. 2021; Golonka ym. 2014). Tutkimusalue on kuitenkin uusi, ja valtaosa puhumisen harjoitteluun suunnitelluista työkaluista rajoittuu ääntä-

137 MITEN SUOMEN JA RUOTSIN OPETTAJAT KÄYTTÄISIVÄT PUHEEN
AUTOMAATTISEEN ARVIOINTIIN KEHITETTYÄ TYÖKALUA?

misestä annettavaan palautteeseen (ks. kuitenkin Gu & Davis 2020, Pearson Education 2018, 2022; de Vries ym. 2014; Xu ym. 2020).

Suomessa automatiikkaa hyödynnetään kielten opetuksessa ja kielitaidon arvioinnissa toistaiseksi hyvin vähän, joten opettajilta kerätyt näkemykset ovat keskeinen osa työkalun kehitys- ja validointityötä. Automaattisen palautteen kehittelyn lisäksi tutkimustulokset hyödyttävät laajemmin kielten opetuksen, kielitaidon arvioinnin ja kieliteknologian parissa työskenteleviä kuten opettajia, opettajien kouluttajia sekä oppimispelien, oppimateriaalien ja kielikokeiden laatijoita.

Tutkimuksessa pyrimme Xin (2010: 296) validiusmalliin nojaten selvittämään automaattisen palautteen vaikutuksia opetukseen ja oppimiseen kartoittamalla työkalun käyttötarkoituksia (tutkimuskysymys 1) sekä automaattisen palautteen hyödyllisyyttä (tutkimuskysymys 2) opettajien näkökulmasta. Tutkimus toi esiin samankaltaisia näkökulmia, joita olemme havainneet muiden käyttäjäryhmien osalta (lukiolaiset ks. von Zansen ym. 2022b, ihmisarvioijat ks. von Zansen ym. 2022c, akateemiset aikuisopiskelijat ks. von Zansen & Hildén 2022). Opettajien näkemät hyödyt koskivat työkalun tarjoamaa lisäresurssia: kouluasteesta riippumatta opiskelijat saavat enemmän myös itsenäisiä harjoittelumahdollisuuksia ja palautetta puhesuorituksestaan (ks. Gu & Davies 2020). Opettajat tiedostivat, että opiskelijoiden olisi hyvä saada enemmän palautetta suullisesta kielitaidostaan, mutta opintojaksojen aikataulujen ja opetusryhmien koon takia opettajat eivät ehdi antaa riittävästi henkilökohtaista palautetta. Näin ollen opettajat hyödyntäisivät työkalua monipuolisesti opintojaksojen aikana esimerkiksi kotitehtävissä ja kokeissa.

Haastateltujen lukio-opettajien mukaan Moodle-työkalusta saatava jatkuva palaute auttaisi opiskelijoita arvioimaan omaa kielitaitoaan sekä sen kehittämiskohteita esimerkiksi osana kieliprofilia, joka otettiin lukioissa käyttöön uuden opetussuunnitelman (Opetushallitus 2019) myötä. Lisäksi työkalu konkretisoisi taitotasoa sekä opiskelijoille että opettajille, mikä on ajankohtaista ja tarpeellista, sillä lukion suullisen kielitaidon kokeen uudistuessa (Opetushallitus 2022) opiskelijat saavat numeroarvosanan lisäksi taitotasoarvion. Tällä hetkellä taitotasojen hyödyntäminen kielten opetuksessa on käytännössä vähäistä, vaikka Suomessa kielten opetussuunnitelma (Opetushallitus 2019) rakentuu Eurooppalaisesta viitekehuksesta (Euroopan neuvosto 2001) peräisin olevalle taitotasoajattelulle.

Taitotasoasteikon sijaan haastateltavat opettajat käyttivät erilaisia asteikkoja ja kriteerejä suullisen kielitaidon arvioimiseen, joista valtaosa oli numeerisia, kuten kouluarvosana. Opettajat vaikuttivat myös pitävän keskenään erilaisia puheen osa-alueita tärkeinä, ja heidän käyttämänsä kriteerit saattoivat vaihdella puhujien taitotason ja tehtävän perusteella. Puheesta automaattisesti mitattavissa olevien ominaisuuksien (Evanini & Zechner 2020; Xi 2010) lisäksi opettajat kertoivat puhetehtäviä arvioidessaan kiinnittävänsä huomiota kehonkieleen ja luovuuteen sekä antavansa myös näistä palautetta opiskelijoille.

Haastattelujen perusteella emme täysin pysty arvioimaan, kuinka hyvin DigiTala-työkalun analyttinen palaute vastaa opetuksessa annettavaa palautetta. Toisaalta opettajat kertoivat pitävänsä DigiTala-työkalussa näkemäänsä palautetta kattavana. Toisaalta havaitsimme, että he ymmärsivät arvioitavan käsitteen (esim. sujuvuus) keskenään eri tavalla, ja paikoitellen myös eri tavalla kuin se määriteltiin automaattisessa palautteessa.

Käsitteelliset erot ovat luonnollisia useastakin syystä. Esimerkiksi kansallisissa oppimistulosten arvioinneissa on havaittu, että arviointikriteereitä tulkitaan jo peruskouluissa eri tavalla yhteneväisistä päättöarvioinnin kriteereistä huolimatta (Kieliverkosto 2022). Tätä tutkimusta varten haastattelemamme kielten opettajat työskentelevät erilaisissa kouluissa, ja suomalaista kielikasvatusta ohjaa pedagoginen vapaus. Lisäksi työkalun palaute on peräisin ihmisarvioijien käyttämistä arviointikriteereistä (von Zansen 2022), jotka puolestaan on laadittu aiemman opetussuunnitelman (Opetushallitus 2003) perusteella. Aiempi opetussuunnitelma sopii paremmin tutkimushankkeen tekniisiin ja analyttisiin lähtökohtiin, sillä automaattinen arviointi rajoittuu puheesta mitattavissa oleviin ominaisuuksiin (Evanini & Zechner 2020; Xi 2010). Opettajat eivät kuitenkaan hyödynnä aiempien tai nykyisten opetussuunnitelmien (Opetushallitus 2003, 2019) taitotasosteikkoja suullisen kielitaidon arvioinnissa. Mikäli opettajien käsitystä suullisesta kielitaidosta halutaan yhtenäistää, ehdotamme ratkaisuksi kansallisten puhumisen arviointikriteerien laatimista sekä opettajien täydennyskoulutusta.

Opettajien näkemät rajoitteet työkalun käytössä koskivat arvioinnin oikeudenmukaisuutta ja paikkansapitävyyttä sekä opiskelijoiden mahdollista kriittistä suhtautumista tekniikkaan, mikä on linjassa Xun ym. tutkimuksen (2020) lisäksi aiempien käyttäjäryhmiä koskevien tutkimustemme (von Zansen ym. 2022b, von Zansen ym. 2022c, von Zansen & Hildén 2022) kanssa. Yhtenä uutena uhkana opettajat näkevät opiskelijoiden kyvyn vastaanottaa automaattista palautetta: he pelkäsivät liian suoran palautteen voivan lannistaa kielen oppijaa. Opettajalla nähtiinkin olevan tärkeä rooli myös työkalun antaman palautteen välittämisessä.

Tämän soveltavan tutkimuksen avulla onnistuttiin tuottamaan lisää tietoa puheen automaattisesta arvioinnista, ennustamaan sen mahdollisia vaikutuksia kielten opetuksessa sekä esittämään toimintasuosituksia. Harkinnanvarainen näyte sopi otantamenetelmäksi tähän pilottivaiheen haastattelututkimukseen. Tutkimuksen rajoitukset koskevat tulosten yleistettävyyttä, sillä haastatelimme neljää hankkeeseen aiemmin osallistunutta opettajaa. Aiommekin täydentää opettajien käsitysten tutkimista laajemmalla kyselytutkimuksella, jonka rajaamista tämä tutkimus tukee. Yksi jatkotutkimusmahdollisuus liittyy arviointityökalun erilaisten palauteversioiden kehittämiseen ja kokeilemiseen oppijoiden ja opettajien kanssa.

Yhteenvetona voidaan todeta, että haastateltavat käyttäisivät työkalua jo nyt tukemaan työtään ja täydentämään nykyisiä opetus- ja arviointimenetelmiä, vaikka tiedostivat työkaluun liittyviä kehitystarpeita. Tulosten perusteella automaattisen palautteen tuomat hyödyt ovat merkittäviä oppimisen ja opetuksen kannalta. Lisäksi

hankkeen laatimien arviointikriteerien (von Zansen 2022) pohjalta tuotetut palauteväittämät olivat kieltenopettajien mielestä ymmärrettäviä ja hyödyllisiä, vaikka ne eivät kata kaikkia puheen ulottuvuuksia. Kehitys- ja tutkimustyötä jatketaan yhteistyössä eri toimijoiden kanssa – haastateltavat toivoivat tekoälyn muun muassa edistävän suullisen kielitaidon mittaamista ylioppilastutkinnon kielikokeissa (Vaarala ym. 2021).

Kiitokset

Suomen Akatemia rahoittaa DigiTala-tutkimushanketta (2019–2023). Kiitämme yhteistyöstä projektin tutkijoita Helsingin yliopistosta (rahoituspäätös 322619), Aalto-yliopistosta (rahoituspäätös 322625) ja Jyväskylän yliopistosta (rahoituspäätös 322965).

Kirjallisuus

- Alanen, T., J. Erkkilä, T. Harjunpää & M. Heijala 2022. Digitala Moodle plugin user manual (1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.6535377>
- Al-Ghezi, R., Y. Getman, A. Rouhe, R. Hildén & M. Kurimo 2021. Self-Supervised End-to-End ASR for Low Resource L2 Swedish. Proc. *Interspeech 2021*, 1429–1433. <https://doi.org/10.21437/Interspeech.2021-1710>
- Al-Ghezi, R., K. Voskoboinik, Y. Getman, A. von Zansen, H. Kallio, C. Akiki, M. Kuronen, A. Huhta & R. Hilden (arvioitavana). Automatic speaking assessment of Spontaneous L2 Finnish and Swedish.
- Deeva, G., D. Bogdanova, E. Serral, M. Snoeck & J. De Weerd 2021. A review of automated feedback systems for learners: Classification framework, challenges and opportunities. *Computers & Education*, 162. <https://doi.org/10.1016/j.compedu.2020.104094>
- Euroopan neuvosto (2001). Common European framework of reference for languages: Learning, teaching, assessment. Council of Europe: Cambridge University Press.
- Evanini, K., & K. Zechner 2020. Overview of automated speech scoring. Teoksessa K. Zechner & K. Evanini (toim.) *Automated Speaking Assessment: Using language technologies to score spontaneous speech*. New York: Routledge, Taylor & Francis Group, 3–20.
- Golonka, E. M., A. R. Bowles, V. M. Frank, D. L. Richardson & S. Freynik 2014. Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer assisted language learning*, 27 (1), 70–105. <https://doi.org/10.1080/09588221.2012.700315>
- Gu, L., & L. Davis. 2020. Providing SpeechRater feature performance as feedback on spoken responses. Teoksessa K. Zechner & K. Evanini (toim.) *Automated Speaking Assessment: Using language technologies to score spontaneous speech*. New York: Routledge, Taylor & Francis Group, 159–175.
- Heijala, M., & A. von Zansen 2022. DigiTala’s teacher interview frame for teacher interviews (spring 2022). Zenodo. <https://doi.org/10.5281/zenodo.7082885>

- Hyvärinen, M., E. Suoninen & J. Vuori (n.d.). Haastattelut. Teoksessa J. Vuori (toim.) *Laadullisen tutkimuksen verkkokäsikirja*. Tampere: Yhteiskuntatieteellinen tietoaarkisto. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus/kvali/laadullisen-tutkimuksen-aineistot/haastattelut/> [luettu 20.06.2020]
- Kieliverkosto 2022. *Tiedote: Yhdeksäsluokkalaisten englannin taitoja arvioitiin keskellä koronaa – osaaminen jäi kouluarvosanaan* 7. <https://www.kieliverkosto.fi/fi/kuulumisia-kentalta/news/tiedote-yhdeksäsluokkalaisten-englannin-taitoja-arvioitiin-keskella-koronaa-2013-osaaminen-jai-kouluarvosanaan-7> [luettu 14.09.2022]
- Kallio, H., R. Suviranta, A. von Zansen & M. Kuronen 2022. Creaky voice and fluency measures in predicting perceived fluency and oral proficiency of spontaneous L2 Finnish. *Proceedings of Speech Prosody 2022*, 777–781. <https://doi.org/10.21437/SpeechProsody.2022-158>
- Kallio, H., M. Kautonen, & M. Kuronen (arvioitavana). Prosody and fluency of Finland Swedish as a second language: investigating global parameters for automated speaking assessment.
- Kautonen, M., & A. von Zansen 2020. DigiTala research project: automatic speech recognition in assessing L2 speaking. *Kieli, koulutus ja yhteiskunta*, 11 (4). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-kesakuu-2020/digitala-research-project-automatic-speech-recognition-in-assessing-l2-speaking>
- Opetushallitus 2003. *Lukion opetussuunnitelman perusteet*. Helsinki: Opetushallitus.
- Opetushallitus 2019. *Lukion opetussuunnitelman perusteet*. Helsinki: Opetushallitus.
- Opetushallitus 2022. *Suullisen kielitaidon arviointiin ja todistuskäytäntöihin muutoksia – koe mahdollista suorittaa myös suullisilla näytöillä*. <https://www.oph.fi/fi/uutiset/2022/suullisen-kielitaidon-arviointiin-ja-todistuskaytanteisiin-muutoksia-koe-mahdollista> [luettu 14.09.2022]
- Ovaska, S., A. Aula & P. Majoranta 2005. *Käytettävyyttutkimuksen menetelmät*. Tampere: Tietojenkäsittelytieteen laitos, Tampereen yliopisto.
- Pearson Education 2018. Versant™ Arabic Test. Test Description and Validation Summary. <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/SupportingDocs/Versant/ValidationSummary/Versant-Arabic-Test-Description-Validation-Summary.pdf>
- Pearson Education 2022. Versant™ English Test. Test Description and Validation Summary. <https://www.pearson.com/content/dam/one-dot-com/one-dot-com/english/SupportingDocs/Versant/ValidationSummary/Versant-English-Test-Description-Validation-Report.pdf>
- Tietoaarkisto 2022. *Kvalitatiivisen datan käsittely*. <https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/kvalitatiivisen-datan-kasittely/> [luettu 7.6.2022]
- Vaarala, H., S. Riuttanen, E. Kyckling & S. Karppinen 2021. *Kielivaranto. Nyt!: Monikielisyyden vahvuudeksi -selvityksen (2017) seuranta*. Jyväskylä: Soveltavan kielentutkimuksen keskus, Jyväskylän yliopisto. https://www.jyu.fi/hytk/fi/laitokset/solki/tutkimus/julkaisut/pdf-julkaisut/kielivaranto-nyt-_julkaisu_sivuittain-1.pdf
- de Vries, B. P., C. Cucchiari, S. Bodnar, H. Strik & R. van Hout 2015. Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*, 28 (6), 550–576. <https://doi.org/10.1080/09588221.2014.889713>
- Vuori, J. n.d. Laadullinen sisällönanalyysi. Teoksessa J. Vuori (toim.) *Laadullisen tutkimuksen verkkokäsikirja*. Tampere: Yhteiskuntatieteellinen tietoaarkisto. <https://www.fsd.tuni.fi/fi/palvelut/menetelmaopetus> [luettu 17.03.2022]
- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading?. *Language Testing*, 27 (3), 291–300. <https://doi.org/10.1177/0265532210364643>

- Xi, X., & Y. Sawaki 2017. Methods of test validation. Teoksessa E. Shohamy, I. Or & S. May (toim.) *Language Testing and Assessment*. Cham, Switzerland: Springer, 193–209. <http://doi.org/10.1007/978-3-319-02261-1>
- Xu, J., M. Brenchley, E. Jones, A. Pinnington, T. Benjamin, K. Knill, G. Seal-Coon & A. Geranpayeh 2020. *Linguaskill Building a validity argument for the Speaking test*. Cambridge: Cambridge Assessment English. <https://www.cambridgeenglish.org/Images/589637-linguaskill-building-a-validity-argument-for-the-speaking-test.pdf>
- Yhteiskuntatieteellinen tietokirasto. (n.d.). Tunnisteellisuus ja anonymisointi. <https://www.fsd.tuni.fi/fi/palvelut/aineistonhallinta/tunnisteellisuus-ja-anonymisointi/#minimointi-eli-miten-kerata-aineisto-niin-ettei-se-sisalla-turhia-tunnisteita> [luettu 20.06.2022]
- von Zansen, A. 2022. *DigiTala's rating criteria: Holistic and analytic scales for assessing L2 speaking*. Zenodo. <https://doi.org/10.5281/zenodo.6477089>
- von Zansen, A., T. Alanen, R. Al-Ghezi, J. Erkkilä, T. Harjunpää, M. Heijala & H. Kallio 2022a. DigiTala Moodle plugin. https://github.com/aalto-speech/moodle-mod_digitala
- von Zansen, A., R. Hildén & M. Sneck 2022b. Lukiolaisten käsitykset ja heidän antamansa palaute suullisen kielitaidon automaattisesta arvioinnista: Ainedidaktinen symposium 2021. Teoksessa R. Kantelinen, M. Kautonen & Z. Elgundi (toim.) *LINGUAPEDA 2021*. Konferenssijulkaisu. Ainedidaktisia tutkimuksia 21. Itä-Suomen yliopisto, Filosofinen tiedekunta, 176–205. <http://hdl.handle.net/10138/352128>
- von Zansen, A. & R. Hildén 2022. "It was cool and comfortable!" Akateemisten alkeistason S2-opiskelijoiden kokemuksia tietokoneella suoritettavasta puhumisen kokeesta. Routarinne, S., P. Heinonen, T. Kärki, A. Roiha, M-L. Rönkkö & A. Korkeaniemi (toim.). *Ainedidaktiikka ajassa: Laajenevat oppimisympäristöt ja eri-ikäiset oppijat*. Rauma: Suomen Ainedidaktinen tutkimusseura. Turun yliopiston opettajankoulutuslaitos, 72–90.
- von Zansen, A., H. Kallio, M. Sneck, M. Kuronen, A. Huhta & R. Hildén 2022c. Ihmisarvioijien näkemyksiä suullisen kielitaidon automaattisesta arvioinnista, digitaalisesta arviointiprosessista sekä puhe-suorituksista arvioitavista ulottuvuuksista. Teoksessa T. Seppälä, S. Lesonen, P. Ikkänen & S. D'hondt (toim.) *AFinLAN vuosikirja*.
- von Zansen, A. & A. Huhta 2022. Developing Automated Feedback on Spoken Performance: Exploring the Functioning of Five Analytic Rating Scales Using Many-facet Rasch Measurement. Teoksessa J. H. Jantunen, J. Kalja-Voima, M. Laukkarinen, A. Puupponen, M. Salonen, T. Saresma, J. Tarvainen & S. Ylönen (toim.) *Diversity of Methods and Materials in Digital Human Sciences : Proceedings of the Digital Research Data and Human Sciences DRDHum Conference 2022, December 1-3, Jyväskylä, Finland*, p. 211–229, University of Jyväskylä. <http://urn.fi/URN:ISBN:978-951-39-9450-1>