

# **Lexize-testin jatkovalidointi: vastaustarkkuuden ja reaktioajan yhteydet yleiseen kielitutkintoon**

Turun Yliopisto  
Yhteiskuntatieteellinen tiedekunta  
Psykologian ja logopedian laitos  
Logopedia  
Pro gradu -tutkielma

Laatija: Tiina Manninen  
Ohjaajat: Raymond Bertram, Minna Lehtonen, Rosa Salmela

10.10.2023

Turku

Pro gradu -tutkielma

**Oppiaine:** Logopedia

**Tekijä:** Tiina Manninen

**Otsikko:** Lexize-testin jatkovalidointi: reaktioajan tarkastelu ja yhteydet yleiseen kielitutkintoon

**Ohjaajat:** Raymond Bertram, Minna Lehtonen, Rosa Salmela

**Sivumäärä:** 41 sivua, 5 liitesivua

**Päivämäärä:** 10.10.2023

Tässä tutkielmassa tarkasteltiin Lexize-testin vastaustarkkuuden ja reaktioaikojen yhteyttä yleisessä kielitutkinnossa (YKI-tutkinto) suoriutumiseen. Kansainvälistyneessä maailmassa on ilmeinen tarve ihmisten kielitaidon tutkimiselle. Tämä taas vaatii kielitaidon mittaamista. Nykyiseltään kielitaidon mittaaminen suomalaisessa kokeellisessa kielipsykologisessa tutkimuksessa pohjautuu pitkälti itsearviointeihin. Maailmalla on kuitenkin jo jonkin aikaa käytetty rinnalla kyllä/ei-sanastotestiksi nimitettyä konseptia, jossa tutkittava erottelee oikeita sanoja kohdekieltä läheisesti muistuttavista epäsanoina (esim. *relo*). Tällaisen testin kyky arvioida kielitaitoa perustuu ajatukseen siitä, että sanaston laajuus on vahvasti yhteydessä muihin kielenkäytön osa-alueisiin. Näin ollen kartoittamalla laajasti sanastoa kyllä/ei-sanastotestin avulla saadaan itsearviointeja objektiivisempi kuva testattavien kielitaidosta. Tutkielmassa käsiteltävä Lexize-testi on suomenkielinen kyllä/ei-sanavarastotesti.

Lexizen on alun perin kehittänyt Salmelan ja kumppaneiden työryhmä vuonna 2021. Tällöin testi validoitiin vertaamalla S1- ja S2-puhujien suoriutumista testissä. Lisäksi kartoitettiin Lexize-pisteiden korrelaatiota erinäisten taustatekijöiden kanssa. Tämän tutkielman tarkoitus on jatkaa Lexizen validointia vertaamalla tutkittavien suoriutumista Lexizessä heidän itseraportoituun suoriutumiseensa suomen kielen YKI-tutkinnossa. YKI-tutkinto on kansallinen kielitutkinto, jossa mitataan tutkittavan kykyä käyttää kieltä luontevasti erilaisissa kielen tuottoa ja ymmärtämistä vaativissa tehtävissä. YKI-tutkinnossa arvioidaan kirjoittamisen, lukemisen, puhumisen ja kuullun ymmärtämisen modaliteetteja. Tässä tutkielmassa tarkasteltiin Lexize-pisteiden korrelaatioita YKI-pisteiden kanssa niin kokonaissuoriutumisen näkökulmasta kuin modaliteettikohtaisesti. Hypoteesina oli, että Lexize-pisteet korreloivat YKI-tulosten kanssa, erityisesti luetun ja kuullun ymmärtämisen tulosten kanssa, sillä aiemmat tutkimukset ovat esittäneet meta-analyysitason todisteita sanaston laajuuden (jota Lexize mittaa) ja näiden modaliteettien välisestä yhteydestä. Lisäksi selvitettiin, antaako tutkittavien reaktioaikojen mittaaminen lisäarvoa Lexizelle. Aiempien tutkimustulosten pohjalta voidaan olettaa, että nopeammat vastaajat saisivat parempia tuloksia YKI-testistä, ja reaktioajan tarkkailu voisi näin ollen lisätä Lexizen kykyä erotella tutkittavien suoriutumista erityisesti korkeilla taitotasolla.

Tutkielmaan kerättiin aineisto aikuisilta S2-puhujilta verkkokyselylomakkeen ja verkossa suoritettavan Lexizen tulosten pohjalta ( $N=65$ ). Tutkielmassa verrattiin itseraportoituja YKI-tuloksia pisteytyskaavan mukaiseen Lexize-pistemäärään ja Lexizessä mitattuun keskimääräiseen reaktioaikaan. Tuloksina havaittiin, että Lexize korreloi YKI-tulosten kanssa vahvasti tai vähintään keskitasolla. Korrelaatio oli tilastollisesti merkitsevä kaikissa modaliteetissa. Vahva korrelaatio löytyi kirjoittamisen ja luetun sekä kuullun ymmärtämisen taitojen kanssa. Puhumisen taidon kanssa korrelaatio oli hieman heikompi. Reaktioajan suhteen löydös taas ei tukenut hypoteesia: Lexize-reaktioaika ei ennustanut suoriutumista YKI-tutkinnossa.

Tutkielman tulokset tukevat Lexizen validiteettia. Toisaalta tulokset viittaavat siihen, ettei reaktioajan tarkastelu paranna testin erottelukykyä. Tulosten perusteella voidaan ajatella, että Lexize on toimiva tapa mitata tutkittavan kielitaitoa silloin, kun kyse ei ole tutkittavalle erityisen painokkaista seurauksista. Esimerkiksi kielentutkimuksen kontekstiin Lexize on pätevä työkalu. Tämän tutkielman tuottamat tulokset ovat osittain erilaisia kuin aiempien samankaltaisia testejä tarkastelleiden tutkimusten tulokset. Reaktioajan mittaamisen hyötyjä kielitaidon testaamiselle tulisi kartoittaa tarkemmin eri konteksteissa ja kohderyhmissä.

**Avainsanat:** sanastotesti, kielitaito, L2, sanasto, suomen kieli, leksikaalinen päätöksenteko, reaktioaika, S2

# Sisällysluettelo

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Johdanto</b>   | <b>5</b>  |
| 1.1      | Sanaston laajuuden yhteydet kielitaitoon eri modalityteeteissa                | 6         |
| 1.2      | Kyllä/ei-sanastotestien validiteetti  | 8         |
| 1.3      | Lexize-testi  | 9         |
| 1.4      | Reaktioaika kyllä/ei-testeissä  | 11        |
| 1.5      | Yleinen kielitutkinto (YKI-tutkinto)  | 12        |
| 1.6      | Lexize-testin validointi YKI-tutkinnon avulla                                 | 13        |
| 1.7      | Tutkimuskysymykset ja hypoteesit  | 14        |
| <b>2</b> | <b>Menetelmät</b>   | <b>16</b> |
| 2.1      | Tutkittavat   | 16        |
| 2.2      | Tutkimusmenetelmät  | 18        |
| 2.2.1    | Tutkimusmateriaalit   | 18        |
| 2.2.2    | Tutkittavat muuttujat   | 19        |
| 2.3      | Tutkimuksen kulku   | 20        |
| 2.4      | Tilastolliset menetelmät  | 20        |
| 2.5      | Tutkimuksen eettisyys   | 21        |
| <b>3</b> | <b>Tulokset</b>   | <b>22</b> |
| 3.1      | Datan valmistelu  | 22        |
| 3.2      | Datan tarkastelu  | 23        |
| 3.3      | Korrelaatioanalyysit  | 25        |
| 3.3.1    | Lexize-pisteiden ja itseraportoitujen YKI-tulosten välinen yhteys             | 25        |
| 3.3.2    | Lexize-reaktioaikojen ja YKI-tulosten väliset yhteydet                        | 28        |
| <b>4</b> | <b>Pohdinta</b>   | <b>30</b> |
| 4.1      | Lexize-testin ja YKI-tutkinnon yhteyden merkitys Lexize-testin validiteetille | 31        |
| 4.2      | Modalityteettien väliset erot   | 33        |
| 4.3      | Lexize-reaktioaika ja YKI-tutkinto  | 35        |
| 4.4      | Tutkielman rajoitteet   | 37        |
| 4.5      | Yhteenvedo ja jatkotutkimus   | 40        |

|                 |           |
|-----------------|-----------|
| <b>Lähteet</b>  | <b>41</b> |
| <b>Liitteet</b> | <b>45</b> |
| <b>Liite 1</b>  | <b>45</b> |
| <b>Liite 2</b>  | <b>46</b> |
| <b>Liite 3</b>  | <b>47</b> |
| <b>Liite 4</b>  | <b>48</b> |
| <b>Liite 5</b>  | <b>49</b> |

# 1 Johdanto

Lukuisat ihmiset tarvitsevat päivittäisessä elämässään kieltä, joka ei ole heidän äidinkieltensä. Kieltä saatetaan käyttää niin opinnoissa, työelämässä kuin vapaa-ajallakin vaihtelevissa tarkoituksissa. Tätä myöhemmin opittua kieltä kutsutaan L2-kieleksi. Myös suomen kieli on L2-kieli monelle ihmiselle. Suomen kielestä puhuttaessa voidaan puhua myös S2-kielestä (suomi toisena kielenä). Kielen käyttö ja ymmärtäminen vaativat laajaa sanastoa. Nation (2006) esitti tutkimuksessaan, että ymmärtääkseen sujuvasti luettua tekstiä tulisi L2-puhujan tuntee 8000–9000 sanaperhettä eli sanasta ja sen taivutusmuodoista (esim. *sana* => *sanassa*) ja johdoksista (esim. *sana* => *sanasto*) koostuvaa kokonaisuutta. Puhutulle kielelle vastaava määrä Nationin mukaan olisi 6000–7000 sanaperhettä, Zeelandin ja Schmittin (2013) mukaan taas 2000–3000 sanaperhettä. Koska sanasto on yhteydessä lukemisen, kirjoittamisen, kuullun ymmärtämisen ja puhutun kielen taitoihin, mahdollistaa sanaston testaaminen osittain myös yleisemmän kielitaidon kehittymisen seuraamisen (Read, 2000). Lisäksi sanaston testaamista tarvitaan silloin, kun pyritään kehittämään entistä parempia menetelmiä kielen oppimisen tukemiseen ja kielen poikkeavuuksien seulontaan. Tässä tutkielmassa keskitytään suomenkieliseen sanastotestiin, Lexizeen (Salmela ym., 2021). Lexize-testin kehittäminen ja validointi on tärkeää, sillä kyseessä on nopea, ilmainen ja itsearviointeja luotettavampi testi, joka edustaa testiformaattia, joka on todettu maailmalla toimivaksi useilla kielillä (esim. Lemhöfer & Broersma, 2012, Brysbaert, 2013). Tällaisesta käytännöllisestä sanavaraston mittaamenetelmästä on apua suomenoppijoihin kohdistuvassa tutkimuksessa, jossa on tarpeen saada nopea mutta luotettava kuva tutkittavien kielitaidosta. Tutkimuksen tärkeyttä korostavat maahanmuuton myötä lisääntyvä suomen kielen L2-oppijoiden määrä ja kansainvälistymisen aiheuttamat paineet entistä laajemmalle kielitaidolle ja sujuvalle kielitaidon osoittamiselle.

Tässä pro gradu -tutkielmassa käsiteltävä testi Lexize (Salmela ym., 2021) on kyllä/ei-tyypin sanavarastotesti. Tämäntyyppisessä testissä testattava erottelee epäsanoja oikeista sanoista. Lexize on validoitu tutkittavien kielitaidon itsearvioita ja tutkittavien taustatietoja hyödyntämällä (Salmela ym., 2021). Tämän tutkielman tarkoituksena on kuvata Lexize-testin jatkovalidointiprosessi, jossa osallistujien tulosta Lexize-testissä verrataan heidän itseraportoituun suoriutumiseensa suomen kielen yleisessä kielitutkinnossa (YKI-tutkinto). Yleinen kielitutkinto on aikuisten kielitutkinto, jossa mitataan yleiskielen osaamista käytännön tilanteissa (Opetushallitus, 2011). Vertaamalla Lexizea YKI-tutkintoon pyrittiin selvittämään Lexizen kykyä ennustaa osallistujan suoriutumista kussakin YKI-tutkinnon modaliteetissa (lukeminen, puhuminen, kuunteleminen, kirjoittaminen) ja näin saamaan tukea Lexizen rinnakkaisvaliditeetille. Lisäksi tarkasteltiin Lexize-testissä mitatun

reaktioajan yhteyttä YKI-tutkinnosta saatuihin pisteisiin ja pyrittiin näin selvittämään, voiko reaktioajan avulla tehdä päätelmiä testattavan osaamistasosta ja tuoko reaktioaikojen seuraaminen näin testille lisäarvoa.

### 1.1 Sanaston laajuuden yhteydet kielitaitoon eri modaliteeteissa

Sanaston laajuutta eli oppijan tunnistamien sanojen määrää on tutkittu yleisen kielitaidon ennustajana eniten reseptiivisten modaliteettien eli lukemisen ja kuullun ymmärtämisen osalta (Qian & Lin, 2020). Esimerkiksi Laufer ja Ravenhorst-Kalovski (2010) totesivat, että jo pieni lisäys L2-oppijan sanastossa auttaa huomattavasti luetun sisäistämässä ja että yhteys sanaston ja luetun ymmärtämisen välillä on näin ollen vahva. Myös Stæhrin (2008) tutkimuksessa sanaston koon yhteys L2-oppijoiden lukemisen taitoihin oli suuri ( $r = .83$ ). Ibrahim (2016) tutkimusryhmineen taas löysi keskisuuren yhteyden ( $r = .64$ ). Keskisuureen yhteyteen päätyivät myös Gungör ja Yayli (2016) ( $r = .41$ ). Myös Jeon ja Yamashita (2014) esittivät meta-analyysissään todisteita sanaston yhteydestä lukemisen taitoon: korrelaatio osoittautui suureksi erityisesti aikuisilla L2-oppijoilla ( $r = .79$ ).

Kuullun ymmärtämisen taitojen osalta yhteys ei välttämättä ole ollut aivan yhtä yksiselitteinen. Mehrpour ja Rahimi (2010) eivät löytäneet yhteyttä yleisen sanastotietämyksen ja kuullun ymmärtämisen taitojen välille. Van Zeeland ja Schmitt (2013) taas totesivat tutkimuksessaan, että erityisesti L2-puhujilla sanaston koon ja kuullun ymmärtämisen välisessä yhteydessä oli huomattavasti varianssia silloin, kun heidän sanasto-osaamisensa oli suppeaa. Toisin sanoen alhaisella tasolla kieltä osaavat henkilöt, joilla oli suurin piirtein yhtä suuri sanasto, saattoivat saada kuullunymmärtämistehtävästä hyvin vaihtelevia pistemääriä suhteessa toisiinsa. Kun sanaston koko kasvoi, efekti pieneni. Yhteys L2-oppijoiden sanaston laajuuden ja kuullun ymmärtämisen välillä oli kuitenkin selkeä. Samaan tulokseen päätyivät Stæhr (2009) sekä Wang ja Treffers-Daller (2017), joiden tutkimuksissa yhteys oli vahva tai keskitasoinen ( $r = .70$  ja  $r = .44$ ). Toisaalta kuullun ymmärtämisen taitojen yhteydestä sanaston laajuuteen on esitetty myös meta-analyysitason todisteita; laajassa meta-analyysissään Zhang ja Zhang (2020) totesivat kuullun ymmärtämisen ja sanaston välisen yhteyden vahvaksi ( $r = .56$ ).

Kielen tuottavista modaliteeteista eli puhumisesta ja kirjoittamisesta sekä niiden yhteydestä sanaston laajuuteen on tehty vähemmän tutkimusta (Qian & Lin, 2000). Stæhr (2008) tutki lukemisen ja kuuntelemisen ohella myös L2-oppijoiden kirjoittamisen taitojen yhteyttä sanaston kokoon. Hänen löydöksensä osoittivat, että yhteys kirjoittamisen taitoihin oli vahva ( $r = .73$ ). Lekin ja Carsonin (1994) tutkimien L2-opiskelijoiden raporteissa oli nähtävissä, että sanasto oli useimmin

mainittu edellytys koherenttien tekstien kirjoittamisessa. Mirapleixin ja Muñozin (2018) löydökset osoittivat, että korkeilla taitotasolla L2-oppijoiden kirjoittamisen taidot ovat vahvasti yhteydessä sanaston kokoon ( $r = .57$ ). He löysivät myös keskiuuret yhteydet muihin kielen modaliteetteihin. Crossley ja kumppanit (2015) havaitsivat, että kyky tunnistaa kollokaatioita eli yhdessä esiintyviä sanoja selitti 84 % varianssista kirjoittamisen arvioinneissa ja 89 % varianssista puhumisen arvioinneissa. Tällainen sanasto-osaaminen osoittautui siis hyväksi ennustajaksi molemmissa tuottavissa modaliteetissa. Sanaston roolia L2-oppijoiden puhetaidossa tukevat myös de Jongin ja työryhmän (2012) tulokset.

Näiden tuottavien modaliteettien kannalta on olennaista huomata, että ne asettavat korkeammat vaatimukset sanaston syvyydelle kuin vastaanottavat modaliteetit (Qian & Lin, 2000). Sanaston syvyydellä tarkoitetaan oppijan kykyä ymmärtää käyttämiään sanoja syvällisesti ja soveltaa niitä eri konteksteissa (Read, 2000). Sanaston syvyyden ja laajuuden käyttöä yleisen kielitaidon mittareina onkin vertailtu paljon. Esimerkiksi Fengin (2014) tutkimustulokset viittaavat siihen, että sanaston syvyys olisi sanaston laajuutta parempi ennustaja arvioitaessa kuullun ymmärtämisen taitoja. Dabbagh (2016) päätyi saman suuntaisiin tuloksiin, joiden mukaan vain sanaston syvyydellä, ei sanaston laajuudella, pystyttäisi ennustamaan kuullun ymmärtämistä. Esimerkiksi Qian (2002) kuitenkin toteaa tutkimuksessaan, että sanaston koolla ja syvyydellä on yhtäläiset ja hyvät mahdollisuudet ennustaa suoriutumista akateemisessa lukemistehtävissä. Syvyys- ja laajuuskysymys näkyy myös sanaston testaamisessa. Syvyyteen keskittyvillä testeillä saadaan testattua pieni määrä sanoja syvällisesti, kun taas laajuuteen keskittyvällä testillä sanoja voidaan samassa aikaikkunassa testata enemmän, mutta pinnallisesti. Näin ollen on pohdittava, kumpi testityyppi on kuhunkin tilanteeseen tarkoituksenmukainen. Tässä tutkielmassa tarkasteltavan Lexize-testin kannalta sanaston laajuuden tarkastelu on tarkoituksenmukaista, sillä tavoitteena on ollut kehittää nopea tapa saada kuva testattavien kielitaidosta, ja sanaston laajuuden testaaminen on syvyyden testaamista nopeampaa. Syvyyden ja laajuuden erottelun ohella myös oppilaan ominaisuudet ovat yhteydessä sanaston potentiaaliin ennustaa kokonaiskielitaitoa. Han (2017) toteaa väitöskirjassaan, että sanaston laajuus on parempi ennustaja silloin, kun oppijan taitotaso on matala, kun taas sanaston syvyys oli muuten parempi ennustaja. Qian ja Lin (2000) kuitenkin huomauttavat, että tälle seikalle olisi hyvä saada lisää näyttöä, ennen kuin siitä voidaan tehdä vankempia johtopäätöksiä. Lexizen voidaan kuitenkin ajatella sopivan laajemmalle joukolle eri taitotasoja edustavia testattavia kuin sanaston syvyyttä mittaavan testin. Vielä viimeksi on myös todettava, että sanaston syvyyden ja laajuuden on osoitettu korreloivan keskenään vähintään keskiuurin korrelaatioin (Kirby & Li, 2015).

Koska sanasto on joka tapauksessa vahvasti yhteydessä yleiseen kielitaitoon, on sanastoa syytä testata myös siksi, että saataisiin nopea käsitys oppijan kokonaisosaamisesta. Tässä tutkielmassa pyritään selvittämään, kykeneekö Lexize-testi tuomaan esille nämä sanaston ja laajemman kielitaidon väliset yhteydet ja miltä yhteydet voisivat näyttää suhteessa kielitaidon eri modaliteetteihin eri tasoilla oppijoilla. Tässä tapauksessa modaliteettien osaamisen mittarina käytetään Yleisen kielitutkinnon (YKI-tutkinnon) osakokeita.

## 1.2 Kyllä/ei-sanastotestien validiteetti

Sanastotestien kehitys vaatii aina kompromisseja ajan, sanojen määrän ja sanojen ymmärtämisen syvyyden mittaamisen välillä (Read, 2000). Tässä tutkielmassa kyllä/ei-testeillä tarkoitetaan testejä, jotka hyödyntävät leksikaalisen päätöksenteon tehtävää (“visual lexical decision”), jossa tutkittavaa ohjeistetaan vastaamaan ”kyllä” oikeisiin sanoihin ja ”ei” epäsanoihin, eli kirjoitusasultaan kohdekieltä muistuttaviin sanoihin, jotka eivät kuitenkaan tarkoita mitään. Ne ovat tällä hetkellä todennäköisesti yksinkertaisin ja nopein tapa mitata L2-oppijoiden sanastoa. Niiden etuna on, että lyhyessä ajassa kyetään käymään läpi suuri määrä sanoja. Heikkoutena taas voidaan nähdä se, ettei testi mittaa testattavien taitoa soveltaa sanastotietämystään kontekstissa. Kyllä/ei-testit siis mittaavat sanaston laajuutta, eivät sen syvyyttä. Kyllä/ei-testit perustuvat checklist-testeihin, jotka eivät sisällä epäsanvoja, vaan tutkittavan on yksinkertaisesti raportoitava mitkä sanat hän tunnistaa (Read, 1988). Zimmerman tutkimusryhmineen (1977) huomasi, että tutkittavat yliarvoivat näin osaamistaan, ja lisäsi testiin epäsanat tämän kontrolloimiseksi. Tätä näin syntyneitä testityyppejä kyllä/ei-testiksi alkoivat ensimmäisinä nimittää Anderson ja Freebody (1983) sekä Meara ja Buxton (1987).

Kyllä/ei-testiformaattia on sovellettu useille kielille ja sitä on tutkittu useissa eri ympäristöissä. Esimerkkejä nykyisille kyllä/ei-formaatin testeille ovat englannin-, hollannin- ja saksankielinen LexTALE (Lemhöfer & Broersma, 2012), espanjankielinen LexTale-esp (Izura ym., 2014) sekä ranskankielinen LexTale\_FR (Brysbaert, 2013). Lisäksi kehitteillä on kiinankielinen toistaiseksi julkaisematon LexTale\_CH (Chang & Chang, 2017). Näistä testeistä tehdyn tutkimuksen perusteella kyllä/ei-testit korreloivat kielitaitojen itsearviointeja paremmin muiden sanastotestien kanssa (Nakata ym., 2020; Lemhöfer & Broersma, 2012; Mochida & Harrington, 2006). Korrelaatiot ovat olleet keskisuuria tai vahvoja. Tuottavan sanaston testien kanssa kyllä/ei-testien korrelaatiot ovat tosin olleet kielen vastaanottoa eli reseptiivisiä taitoja mittaavien testien kanssa mitattuja korrelaatioita heikompia (Eyckmans ym., 2007; Cameron, 2002).



Yleisen kielitaidon tehtävien ja kyllä/ei-testien väliset yhteydet ovat olleet kooltaan vaihtelevia. Nakata ja kumppaneiden (2020) sekä Lemhöferin ja Broersman (2012) löytämät korrelaatiot englanninkielisen LexTalen ja yleisen kielitaidon mittareiden kanssa vaihtelivat pienestä keskisuureen. Harrington ja Carey (2009) taas totesivat nimeämättömän englanninkielisen kyllä/ei-testin kykenevän luokittelemaan oppilaita lähes yhtä hyvin kuin laajempi lähtötasotesti ja yksinkertaisuuteensa nähden erinomaisesti. Faircloughin (2011) löydökset taas viittaavat hänen kehittämänsä nimettömän espanjankielisen kyllä/ei-testin melko suuriin korrelaatioihin Cloze-testin (jossa testattavan on täydennettävä joka viides sana tekstiin) ja Multiple Task -testin (Faircloughin tutkimusryhmän kehittämä neljä tuottavaa tehtävää sisältävä kokonaisuus, ks. Fairclough, 2011) kanssa. Kyllä/ei-testit toki eroavat toisistaan myös esimerkiksi siinä, kuinka tarkasti niiden sanat ja epäsanat on valittu. Testejä ei siis voi tarkastella täysin yhtenä kokonaisuutena. Ei siis ole täysin selvää, kuinka hyvä yleisen kielitaidon ennustaja kyllä/ei-testi formaattina on, mutta yhteys on joka tapauksessa joidenkin testien kohdalla havaittavissa.

Kyllä/ei-testi on kustannustehokas, nopea ja helppokäyttöinen. Lisäksi se on helposti muunnettavissa tietokoneella tehtäväksi, mikä tekee siitä soveltuvan työkalun testaajille. Kontekstisidonnaisuuden puutteen lisäksi sillä on kuitenkin myös muutama muu heikkous. Kyllä/ei-testit voivat toimia eri tavoin kulttuurikontekstista riippuen (Lemhöfer & Broersma, 2012; Meara, 1990). Epäsanojen muodostamisessa ei myöskään ole yhtenäisiä periaatteita, ja tutkijat joutuvat usein tasapainoilemaan sen välillä, etteivät sanat ole niin helppoja, että ne voisi tunnistaa epäsanoina ilman riittävää kielitaitoa ja toisaalta etteivät sanat ole niin vaikeita, että L1-puhujat valitsevat niitä todennäköisemmin kuin L2-puhujat (Brysbaert, 2013). Myös testin pisteytyksestä on esitetty eriäviä näkökulmia (esim. Beeckmans ym., 2001). Kyllä/ei-formaatin pulmakohtiin ja niihin esitettyihin ratkaisuihin palataan tarkemmin työn pohdintaosioissa.

Kyllä/ei-testiformaatilla on siis etunsa ja ongelmansa. Tässä tutkielmassa tarkasteltiin lähemmin kyllä/ei-testin suhdetta yleiseen kielitaitoon vertaamalla Lexize-testiä suomen kielen YKI-tutkintoon. Samalla tarkasteltiin reaktioajan ja suoriutumisen yhteyttä. Näin pyrittiin luomaan näkökulmia yleisen kielitaidon mittaamiseen ja reaktioajan soveltamiseen liittyviin aiheisiin kyllä/ei-testeissä ja tuomaan lisävarmuutta sille, että Lexize-testi sopii suomen kielen kielitaidon arviointiin.

### 1.3 Lexize-testi

Lexize (Salmela ym., 2021) on suomenkielisen sanavaraston mittaamiseen tarkoitettu verkkoympäristössä toteutettava kyllä/ei-testi. Lexize on kehitetty LexTALen (Lemhöfer &

Broersma, 2012) pohjalta. Ennen tässä tutkielmassa kuvattavaa jatkovalidointia Lexizen luotettavuutta on tarkasteltu yhdessä aikaisemmassa tutkimuksessa (Salmela ym., 2021). Tutkimuksessa verrattiin suomen kielen S1- ja S2-puhujien suoriutumista Lexize-testissä ja tarkasteltiin suoriutumisen yhteyttä itsearvointeihin sekä taustatietoihin. Tulokset osoittivat, että Lexize pystyi erottelemaan S1- ja S2-puhujat toisistaan. Lexizen antamat tulokset myös korreloivat vahvasti S2-puhujien raportoimien itsearvointien ja suomen kielelle altistumisen määrän kanssa. S1-puhujien osalta Lexize korreloi itsearvointien, äidinkielen arvosanan, koulutusasteen sekä iän kanssa. Lexizen voidaan siis ajatella kykenevän erottelemaan suoriutumista niin matalilla kuin korkeilla suomen kielen taitotasolla, joskin paremmin matalammilla taitotasolla. (Salmela ym., 2021).

Lexizen luomisesta voi lukea tarkemmin alkuperäisartikkelista, jossa kuvataan sanojen luokat sekä esiintymistiheys kielessä sekä epäsanojen muodostaminen (Salmela, 2021). Lopullinen Lexize koostuu 102 kirjainjonosta. Nämä kirjainjonot on valittu mukaan alkuperäisestä, suuremmasta ärsykejoukosta niiden erottelukyvyn perusteella. Arvaamisen vaikutus huomioidaan Lexizen pisteytysosiossa, jossa käytettävä pisteytyskaava sakottaa arvaamisesta. (Salmela ym., 2021). Osio-vaste teorian mukaisen analyysin jälkeen lopulliseen testiin valikoitui 68 sanaa ja 34 epäsanaa. Tämä sanojen ja epäsanojen välisen suhteen epätasaisuus on tyypillistä kyllä/ei-testeille silloin, kun pyritään mittaamaan L2-oppijoiden osaamista, sillä voidaan olettaa, etteivät korkeatasoisetkaan L2-puhujat tunnista kaikkia oikeita sanoja (esim. Lemhöfer & Broersma, 2011). Itemien helppouden takia Lexize erottelee parhaiten matalan osaamistason suorittajia, tässä tapauksessa L2-puhujia. (Salmela ym., 2021).

Lexize pisteytetään noudattamalla seuraavaa kaavaa:

$$N_{oikein\ vastatut\ sanat} - \left( \frac{N_{sanat}(68)}{N_{pseudosanat}(34)} \right) * N_{väärin\ vastatut\ pseudosanat}$$

Oikein vastatuista oikeista suomen kielen sanoista vähennetään siis väärin vastattujen epäsanojen määrällä kerrottu kaikkien oikeiden sanojen määrä jaettuna kaikkien epäsanojen määrällä. Kaavan tarkoitus on rankaista arvauskäyttäytymisestä. Mikäli vastaaja arvaa epäsanoina virheellisesti sanoiksi, vähentävät nämä vastaukset kokonaispistemäärää. Samoin oikeiden sanojen arvaaminen epäsanoina vähentää pistekertymää. Mikäli vastaaja vastaa täysin satunnaisesti, oletetaan hänen arvioivan puolet sanoista oikeiksi sanoiksi ja puolet epäsanoina, jolloin tulos oheisen pisteytyskaavan mukaisesti olisi 0. Mikäli vastaaja vastaa yli puoleen sanoista tai epäsanoina

väärin, on hänen testituloksensa negatiivinen. Maksimitulos kaikki oikein saavalle vastaajalle on 68 pistettä. (Salmela ym., 2021).

Lexize-testin voi suorittaa ilmaiseksi verkkoympäristössä osoitteessa <https://lexize.org/>.

#### 1.4 Reaktioaika kyllä/ei-testeissä

Reaktioaikaa kyllä/ei-testeissä ovat tarkastelleet Harrington (2006), Harrington ja Carey (2009) sekä Pellicer-Sánchez ja Schmitt (2012). Kaikki kolme tutkimusta viittaavat siihen, että reaktioajan tarkastelu voi tuoda lisäarvoa perinteisiin kyllä/ei-testeihin. Harrington (2006) vertaili englanninoppijoiden ja natiivipuhujien reaktioaikoja kyllä/ei-testeissä ja huomasi, että vastaustarkkuuden ja reaktioaikojen välillä oli negatiivinen korrelaatio – enemmän oikeita vastauksia saavien reaktioaika oli pienempi. Toisin sanoen nopeat vastaajat olivat myös tarkempia. L1-puhujat olivat L2-puhujia nopeampia ja edistyneemmät L2-puhujat nopeampia kuin matalan tason puhujat. Harrington kuitenkin huomauttaa, että kaikki hänen tutkittavansa olivat melko korkeatasoisia.

Harrington ja Carey (2009) taas tarkastelivat kyllä/ei-testin reaktioaikojen potentiaalia jakaa tutkittavia erään kielikoulun lähtötasotestin mukaisesti taitotasoihin. Reaktioaika oli vastaustarkkuutta selkeästi huonompi erottelija, mutta tutkijat näkevät reaktioajan hyödyntämisessä potentiaalia, sillä vastaustarkkuus ja reaktioaika yhdistettynä antoi paremman arvion tutkittavan taitotasosta kuin perinteinen vastaustarkkuus yksinään.

Pellicer-Sánchez ja Schmitt (2012) lähestyivät reaktioajan ja kyllä/ei-testin suhdetta hiukan eri näkökulmasta esittäen kysymyksen siitä, voisiko kyllä/ei-testien epäsanat ja pisteytyskaavat korvata reaktioaikaan pohjautuvalla lähestymistavalla, tai toisiko reaktioaikojen lisääminen epäsanajoja sisältäviin kyllä/ei-testeihin lisäarvoa. Tutkimuksessa lähdettiin liikkeelle oletuksesta, jonka mukaan ihmiset vastaavat kyllä/ei-testeihin hitaammin silloin, kun ovat epävarmoja vastauksestaan. Varmojen vastausten taas voitaisiin olettaa olevan nopeita. Toisin sanoen, jos tutkittava vastaisi huomattavan hitaasti, voitaisiin tätä pitää todisteena siitä, ettei hän tiedä sanaa vaan arvaa vastauksensa. Tutkimus osoitti, että reaktioaika olisi perinteisiä pisteytyskaavoja parempi tapa pisteyttää kyllä/ei-testi silloin, kun tutkittava valitsee vain vähän tai ei ollenkaan epäsanajoja testissä. Tutkijat esittävät tulostensa perusteella, että kyllä/ei-testien pisteyttämiseen voitaisiin soveltaa erilaisia menetelmiä riippuen tutkittavan taitotasosta. Yksi näistä menetelmistä olisi reaktioajan perusteella pisteyttäminen, jonka tutkijat esittävät sopivan hyvin nykymuotoiseen tietokoneistettuun kyllä/ei-testaamiseen.

## 1.5 Yleinen kielitutkinto (YKI-tutkinto)

Opetushallituksen (2011) mukaan Yleisellä kielitutkinnolla (YKI-tutkinto) viitataan Suomessa toteutettavaan aikuisen henkilön toiminnallista kielitaitoa mittaavaan tutkintoon. (Opetushallitus, 2011). Opetushallitus kuvaa, että tutkintoon sisällytettyjen tehtävien tarkoitus on mitata osallistujan kykyä toimia luontevasti erilaisissa kielen ymmärtämistä ja tuottoa vaativissa tehtävissä. Härmälän ja kumppaneiden (2010) mukaan tutkinto on mahdollista suorittaa yhdellä kolmesta tasosta: perustasolla, keskitasolla ja ylimmällä tasolla. Perustason saavuttamiseen riittää lyhyt mutta ymmärrettävä tuotos, keskitasolla vaaditaan luontevampaa suoriutumista ja ylimmällä tasolla kielen käytön on jo oltava sujuvaa ja tarkoituksenmukaista. Tutkinon suorittaja valitsee itse taitotasonsa sen mukaan, miksi hän on suorittamassa tutkintoa tai minkä tasoiseksi hän itse arvioi taitonsa (Härmälä ym., 2010). Kukin näistä tasoista jakautuu vielä kahteen taitotasoon (perustasot A1–A2, keskitasot B1 ja B2 sekä ylimmät tasot C1 ja C2). Näin ollen YKI-tutkinnoissa kielitaitoa arvioidaan eurooppalaiseen viitekehykseen verrattavissa olevalla kuusiportaisella taitotasoasteikolla (Opetushallitus, 2011). YKI-tutkinnoissa on neljä modaliteettia: kirjoitettu tuotto, luetun ymmärtäminen, puheen tuotto ja puheen ymmärtäminen (Ahola & Lebay, 2014).

Aholan ja Leblayn (2014) mukaan tutkinon suorittamiseen motivoi esimerkiksi tarve saada todiste kielitaidosta esimerkiksi kansalaisuuden, opiskelupaikan tai työpaikan hakemista varten. Tutkinon käyttötarkoitukset vaihtelevat kuitenkin eri kielten välillä. Suomen kielessä yleisimmät syyt ovat työnhaku ja kansalaisuuden hakeminen. YKI-tutkinon keskitason suorittaminen onkin osoitus kansalaisuuden hakemiseen vaadittavasta taitotasosta, minkä vuoksi keskitason suorittaminen suomen kielessä on erityisen suosittua (Ahola & Lebay, 2014). Suomen kielessä vahvimmat taidot ovat yleensä ymmärtämisessä ja heikoimmat kirjoittamisessa (Härmälä ym., 2010). YKI-tutkinto on avoin kaikille, eikä näin ollen vaadi virallisia kieliopintoja. YKI-tutkinon tehtävissä ei myöskään vaadita yleistietämystä tai taustakoulutusta (Ahola & Lebay, 2014).

Kielen asiantuntijat arvioivat YKI-tutkinnot. Yleisten kielitutkintojen asetuksen 2004/1163 mukaan arvioijilta vaaditaan kohdekielen opintoja niin, että heillä on vaadittu kelpoisuus opettaa arvioimaansa kieltä yliopistossa tai ammattikorkeakoulussa, tai ylempi korkeakoulututkinto kyseisestä kielestä. Lisäksi heiltä vaaditaan Opetushallituksen hyväksymä arvioijakoulutus (Yleisten kielitutkintojen asetus 2004/1163, 7 §). Lisäksi Ahola ja Lebay (2014) kuvaavat, että koulutuksen ohella laadukkaan arvioinnin periaatteita pyritään edistämään hyödyntämällä tutkintojärjestelmään kertyvää tilastotietoa tehtävien erottelukyvystä, toimivuudesta ja vaikeudesta. YKI-tutkinon tehtäville ei suoriteta esitestausta, mutta niistä kerätään tilastotietoa jokaisen

testikerran jälkeen. Tilastollisten analyysien jälkeen tehtävät siirtyvät osiopankkiin, josta niitä voidaan hyödyntää tulevien testikertojen suunnittelussa. Osiopankin avulla pyritään varmistamaan, että YKI-tutkinnon vaikeustaso säilyy samana eri toteutuskertojen välillä. Tilastotietoa hyödynnetään tehtävien arvioinnin ohella myös YKI-suoritusten arviointien toteuttajien toiminnan seuraamiseen. Tilastotiedon avulla arvioijat voivat saada palautetta johdonmukaisuudestaan ja mahdollisesta taipumuksesta liialliseen ankaruuteen tai lempeyteen arvioinnissa (Ahola & Lebay, 2014).

Tilastotiedon hyödyntämisen ohella vertailukelpoisten tutkintokertojen toteuttamiseen pyritään tutkintokielen asiantuntijoiden arvioilla tehtävien haastavuudesta ja pisterajojen asettamisesta (Ahola, Hirvelä & Neittaanmäki, 2016). Aholan, Hirvelän ja Neittaanmäen (2016) mukaan asiantuntijat määrittelevät kussakin tehtävässä vaadittavan osaamisen tason suhteessa taitotasokuvaimiin ja pyrkivät löytämään sopivat pisterajat suhteessa kielitaidon kehittymistä kuvaavaan taitoasteikkoon.

Standardointiin osallistuvilta asiantuntijoilta edellytetään ensisijaisesti vahvaa asiantuntemusta heidän arvioimistaan kielistä, mutta tämän lisäksi valinnassa huomioidaan ikä, sukupuoli sekä alueellisuus (Ahola, Hirvelä & Neittaanmäki, 2016). Luotettavan tuloksen saamiseksi asiantuntijoita ajatellaan tarvittavan vähintään kymmenen. YKI-tutkinnon sisällä kerätyn tilastotiedon ja asiantuntija-arvioiden ohella YKI-testin suomen kielen keskitason tutkinnon laadukkuutta valvotaan eurooppalaisen kielitutkintojärjestön Association of Language Testers in Europe (ALTE) toimesta (Ahola & Lebay, 2014). Järjestö tarkkailee viiden vuoden välein tutkinnon yhdenmukaisuutta, yhdenvertaisuutta, tarkkuutta ja luotettavuutta. Kokonaisuudessaan YKI-tutkintoa voidaan siis pitää varsin luotettavana.

## **1.6 Lexize-testin validointi YKI-tutkinnon avulla**

Vertaamalla Lexize-testiä ja YKI-tutkintoa tässä tutkielmassa pyrittiin varmistamaan Lexizen kyky välillisesti arvioida suomen kielen taitoa. Korrelaatio YKI-testin kanssa antaisi osoituksen Lexize-testin rinnakkasvaliditeetista. Lisäksi se viittaisi Lexizen kykyyn arvioida kielitaitoa myös sanastoa laajemmin. YKI-tutkinto on luotettu ja laadukas testikokonaisuus, joten osoitus siitä, että Lexize-antaisi samansuuntaisia tuloksia olisi vahva osoitus siitä, että Lexizeä voitaisiin käyttää objektiivisena välineenä nopeaa kielitaidon arvioita vaativissa tilanteissa esimerkiksi tutkimuskontekstissa.

Ilmestyessään Lexize vastasi suureen tarpeeseen, sillä Suomessa ei ennen sitä ollut nopeaa standardoitua sanastotestiä aikuisille S2-puhujille eli suomen kielen oppijoille. S2-puhujien taitoja voitiin mitata Kielo-testillä, mutta Kielo on tarkoitettu ensisijaisesti opettajien työkaluksi (Tani, 2008). Tämä näkyy siinä, että Kielo on pidempi ja monipuolisempi kuin on usein tarpeen psykologisen kielentutkimuksen kontekstissa. Lexize pyrkii paikkaamaan tämän aukon suomalaisen kielipsykologisen tutkimuksen kentällä. Lisäksi Lemhöfer ja Broersma (2012) tuovat LexTale-testinsä kehityksen yhteydessä esille standardoinnin mahdollisuuden, mikäli testiformaatista tulee laajalti levinnyt. Tämä pätee myös Lexizeen: suomalainen tutkimus on sekä maan sisäisesti että kansainvälisesti vertailukelpoisempaa, mikäli käytetyt menetelmät ovat mahdollisimman yhtenäisiä.

Modaliteettien välisiä eroja tutkimalla pyrittiin selvittämään, kattaako Lexizen kyky arvioida kokonaiskielitaitoa kaikki modaliteetit. Tämä on olennaista, mikäli Lexizeä halutaan käyttää laajasti kaikkia modaliteetteja ja yleistä kielenosaamista käsittelevässä tutkimuksessa. Reaktioaikoja tarkastelemalla taas pyrittiin selvittämään, onko reaktioaikojen lisäämisellä testiin parantava vaikutus testin erottelukykyyn. Näin pyrittiin luomaan mahdollisimman tehokas työkalu kielitaidon testaamiseen. Reaktioaikojen tutkiminen voi myös auttaa pohtimaan vaihtoehtoisia pisteytystapoja nykyisten pisteytyskaavojen oheen.

## 1.7 Tutkimuskysymykset ja hypoteesit

YKI-tutkinto on laaja ja luotettava suomalainen yleisen kielitaidon mittari. Lexize-testi taas on kehitetty nopeaksi ja objektiiviseksi tavaksi mitata tutkittavien sanastoa. Lexize-testin ja YKI-tutkinnon välisistä yhteyksistä esitettiin tutkielmassa seuraavat kysymykset ja hypoteesit:

1. Missä määrin Lexize-testistä saatu pistemäärä on yhteydessä tutkittavien itseraportoitujen YKI-tulosten kanssa?

Hypoteesina oli, että Lexize-pisteet ovat yhteydessä YKI-tuloksiin. Tämä perustuu aiempaan Lexizeä käsittelevään tutkimukseen (Salmela ym., 2021), jossa Lexize kykeni erottelemaan suomen kielen L1- ja L2-puhujat toisistaan sekä korreloi kielitaitoa ennustavien taustatekijöiden ja tutkittavien itsearviointien kanssa. Koska nämä taustatekijät vaikuttavat myös YKI-tutkinnoissa suoriutumiseen ja myös YKI-tutkinto pyrkii erottelemaan eri kielitaitoryhmiä, voidaan esittää oletus siitä, että Lexize on yhteydessä YKI-tutkinnoissa suoriutumiseen. Lisäksi vastaavat kyllä/ei-testit ovat kansainvälisesti korreloineet muiden yleisen kielitaidon mittareiden kanssa (esim. Nakata ym., 2020; Lemhöfer & Broersma, 2012; Mochida & Harrington, 2006).

1.1 Alakysymys: Millainen korrelaatio on suhteessa YKI-tutkinnon eri modaliteetteihin (lukeminen, kirjoittaminen, puhuminen ja kuullun ymmärtäminen)?

Hypoteesina oli, että Lexize-pisteet korreloivat parhaiten lukemisen ja kuullun ymmärtämisen modaliteettien kanssa. Tämä perustuu aiempaan tutkimukseen siitä, että tuottavat modaliteetit (kirjoittaminen ja puhuminen) asettavat enemmän vaatimuksia sanaston syvyydelle (Qian & Lin, 2000), kun taas Lexize on sanaston laajuutta mittaava testi. Lisäksi lukemisen ja kuullun ymmärtämisen yhteydestä sanaston laajuuteen on tehty enemmän tutkimusta kuin vastaanottavista modaliteetista, ja tulokset ovat olleet yhdensuuntaisempia. Oletus oli kuitenkin, että Lexize on yhteydessä myös tuottaviin modaliteetteihin.

Reaktioajan merkitystä kyllä/ei-testissä on tutkittu vähän, lähtökohtaisesti korkeatasoisilla oppijoilla ja kirjoittajan tietämyksen mukaan vain englanninkielisillä tutkittavilla. Tässä tutkielmassa pyrittiin vastaamaan seuraaviin Lexize-testin reaktioaikoihin liittyviin kysymyksiin:

2. Onko Lexize-testissä mitattu keskimääräinen reaktioaika yhteydessä YKI-tuloksiin?

Hypoteesina oli, että Lexize-testissä mitattu keskimääräinen reaktioaika korreloi YKI-tulosten kanssa. Tämän hypoteesin pohjalla olivat aiemmat tutkimukset kyllä/ei-testien ja muiden sanastotestien reaktioaikojen yhteydestä sanastotesteissä suoriutumiseen (Harrington, 2006; Harrington ja Carey, 2009; Pellicer-Sánchez ja Schmitt, 2012).

2.1 Alakysymys: Onko pelkistä oikeista tai vääristä vastauksista mitattu reaktioaika yhteydessä YKI-tuloksiin?

Hypoteesina oli, että oikeista ja vääristä vastauksista mitatut reaktioajat korreloivat YKI-testin kanssa päinvastaisesti. Aiemman tutkimuksen perusteella tutkittavat vastaavat hitaammin vastatessaan väärin ja nopeammin vastatessaan oikein (Pellicer-Sánchez & Schmitt, 2012).

2.2 Alakysymys: Onko reaktioaika yhdistettynä vastaustarkkuuteen Lexizessä parempi YKI-tulosten ennustaja kuin pelkkä Lexize-vastaustarkkuus?

Hypoteesina oli, että vastaustarkkuus yhdistettynä reaktioaikaan on parempi YKI-tulosten ennustaja kuin pelkkä vastaustarkkuus. Tämä perustuu aiempaan tutkimukseen (Harrington & Carey, 2009), jossa reaktioajan huomioiminen testin pisteytyksessä kasvatti testin erottelukykä.

## 2 Menetelmät

### 2.1 Tutkittavat

Koehenkilöt tutkimukseen rekrytoitiin verkossa jaettavan ilmoituksen avulla (ks. liite 1) erilaisista suomenoppijoiden verkkoyhteisöistä. Ilmoitukseen oli sisällytetty lyhyt esittelyteksti ja linkki lisätietoihin, Lexize-testiin sekä taustatietolomakkeeseen. Ilmoituksen jakamisen alustoina käytettiin Facebookia suomen kielen opiskeluun liittyvän blogin kommenttikenttää ja Discord-viestintäsovellusta Tarkemmat tiedot käytetyistä verkkosivuista ovat luettavissa liitteessä 2. Lisäksi oltiin yhteydessä sähköpostitse eri maiden ystävyysseuroihin, mutta näiden kautta ei saatu yhtään osallistujaa. Kaiken kaikkiaan ilmoitukseen kautta saatiin 657 vastausta, joista sisäänottokriteerit ja poikkeavat arvot huomioiden tutkimukseen soveltui 65.

Sisäänottokriteerit tutkimukseen osallistuvalla olivat seuraavanlaiset: 1) vähintään 18 vuoden ikä, 2) äidinkieli muu kuin suomi, 3) suomen kielen S2-oppija, 4) suomen kielen yleinen kielitutkinto (YKI-tutkinto) suoritettu hyväksytysti ja 5) Lexize-testin pisteytyskaavan mukainen tulos yli 0. Mikäli henkilön havaittiin tehneen Lexize-testin kahdesti, vain ensimmäinen suoritus sisällytettiin tutkimusaineistoon. Näillä kriteereillä pyrittiin rajaamaan tutkittavat tutkimuksen tarkoitusta vastaavaksi ryhmäksi. Vaatimalla Lexize-testistä yli 0:n pisteen mukaista tulosta pyrittiin minimoimaan ne suoritukset, joissa ohjeistusta ei ollut ymmärretty tai vastaukset koostuivat pelkistä arvauksista.

Osallistujat puhuivat äidinkielenään ainakin 23:a eri kieltä. Luettelo kielistä on liitteessä 3. Suurimmat kieliryhmät olivat englanti (18 osallistujaa), venäjä (6 osallistujaa), vietnam (6 osallistujaa), kiina (5 osallistujaa) sekä espanja (4 osallistujaa). Kaksi osallistujista oli valinnut taustatietolomakkeella vaihtoehdon ”muu kieli”. Tarkemmin osallistujien äidinkieliä voi tarkastella kohdassa liite 5. Keskimääräinen itsearvio suomen kielen taitotasosta aineistossa asteikolla 1–10 oli 5.8 ( $VV = 3-5$ ,  $KH = 1,4$ ). Keskimääräinen arvio oman äidinkielen osaamisesta asteikolla 1–10 oli 9.3 ( $VV = 7-10$ ,  $KH = 0.91$ ). Naisia oli aineistossa huomattavasti enemmän kuin miehiä. Suurin osa osallistujista oli korkeakoulutettuja. Tarkempi kuvaus kyselyyn vastanneista löytyy taulukosta 1.



Taulukko 1

*Lukumäärätietoa tutkittavien sukupuolesta, koulutusasteesta, iästä ja kielitaitotasosta*

| Taustatieto                                  | <i>n</i> | %  |
|--|----------|----|
| Sukupuoli                                    |          |    |
| Mies   | 11       | 17 |
| Nainen                                       | 54       | 83 |
| Koulutusaste                                 |          |    |
| Ammattikoulu                                 | 2        | 3  |
| Lukio  | 3        | 6  |
| Ammattikorkeakoulu                           | 14       | 21 |
| Yliopisto                                    | 38       | 58 |
| Tohtorintutkinto                             | 8        | 12 |
| Ikäjakauma                                   |          |    |
| 20–29 v.                                     | 14       | 22 |
| 30–39 v.                                     | 30       | 46 |
| 40–49 v.                                     | 15       | 23 |
| 50–59 v.                                     | 6        | 9  |
| Itsearvio suomen kielen taidosta (1–10) *    |          |    |
| Tasot 3–4                                    | 10       | 15 |
| Tasot 5–6                                    | 36       | 55 |
| Tasot 7–8                                    | 15       | 23 |
| Tasot 9–10                                   | 4        | 6  |
| Itsearvio oman äidinkielen tasosta (1–10) ** |          |    |
| Taso 7                                       | 3        | 5  |
| Taso 8                                       | 11       | 17 |
| Taso 9                                       | 17       | 26 |
| Taso 10                                      | 34       | 52 |

\* Ei vastauksia alle tason 3

\*\* Ei vastauksia alle tason 7

Tutkittavat olivat suorittaneet YKI-tutkintoja vuosina 2005–2022. Pieni enemmistö tutkittavista ( $n = 37$ ) oli suorittanut tutkinnon vuosina 2018–2022. Keskimääräinen suoriutuminen kaikilla YKI-testin osa-alueilla asteikolla 1 – 6 oli 3.4 ( $VV = 1,25-6$ ,  $KH = 0.94$ ). Lukemisen osalta keskimääräinen suoriutuminen oli 3.6 ( $VV = 1-6$ ,  $KH = 1,0$ ), puhumisen osalta 3.3 ( $VV = 1-6$ ,  $KH = 0.95$ ), kuuntelemisen osalta 3.4 ( $VV = 1-6$ ,  $KH = 0.98$ ) sekä kirjoittamisen osalta 3.3 ( $VV = 1-6$ ,  $KH =$

= 1,0). Kaikki osa-alueet edustavat siis keskiarvoiltaan YKI-testin tasoja B1 ja B2. Tarkempi kuvaus osallistujien YKI-testituloksista löytyy taulukosta 2.

**Taulukko 2**

*YKI-arvosanojen määrä ja suhteellinen osuus aineistossa*

|                          | A1       |     | A2       |      | B1       |      | B2       |      | C1       |     | C2       |     |
|--------------------------|----------|-----|----------|------|----------|------|----------|------|----------|-----|----------|-----|
|                          | <i>n</i> | %   | <i>n</i> | %    | <i>n</i> | %    | <i>n</i> | %    | <i>n</i> | %   | <i>n</i> | %   |
| YKI Lukeminen            | 1        | (2) | 4        | (6)  | 29       | (44) | 24       | (37) | 1        | (2) | 6        | (9) |
| YKI Kuullun ymmärtäminen | 2        | (3) | 3        | (5)  | 36       | (55) | 18       | (28) | 2        | (3) | 4        | (6) |
| YKI Puhuminen            | 2        | (3) | 6        | (9)  | 35       | (54) | 16       | (25) | 4        | (6) | 2        | (3) |
| YKI Kirjoittaminen       | 2        | (3) | 7        | (11) | 33       | (51) | 16       | (24) | 3        | (5) | 4        | (6) |
| YKI Keskiarvo            | 2        | (3) | 2        | (3)  | 34       | (52) | 20       | (31) | 2        | (3) | 5        | (8) |

## 2.2 Tutkimusmenetelmät

### 2.2.1 Tutkimusmateriaalit

Tutkimusmateriaali koostui verkossa jaettavasta suostumuslomakkeesta, taustatietolomakkeesta sekä Lexize-testistä (Salmela ym., 2021). Osallistuja pääsi näkemään tutkimusmateriaalin seuraamalla linkkiä <https://psyk.abo.fi/LexizeWeb/#/> joko tietokoneella tai kosketusnäyttölaitteella.

Suostumuslomakkeessa (ks. liite 4) vastaajalle kerrottiin lyhyesti testin kulusta, tulosten käsittelystä ja luottamuksellisuudesta, osallistujan oikeuksista, testin suorittamisen vapaaehtoisuudesta sekä osallistumisen yhteydessä suoritetusta arvonnasta. Arvonnassa osallistujilla oli mahdollisuus voittaa 40 euron arvoinen lahjakortti GoGift-verkkokauppaan.

Taustatietolomakkeessa (ks. liite 5) kerättiin osallistujista seuraavat tiedot: sähköpostiosoite, sukupuoli, äidinkieli, itsearvioitu äidinkielen kielitaito, suomen opiskelun aloittamisen ajankohta, koulutusaste, ikä, Suomessa asuttu aika, YKI-testin arvosanat ja suoritusajankohta sekä tieto siitä, suoritettiinko testi kosketusnäyttölaitteella. Taustatietolomakkeen täytettyään osallistuja pääsi siirtymään Lexize-testiin.

Tutkimuksessa sovellettiin Salmelan ja kumppaneiden tutkimuksen (2021) mukaista Lexize-testiä pienin muunnoksien. Itse testissä ainoa muutos oli sanojen ajastaminen. Siinä missä

alkuperäistutkimuksessa tutkittavat saivat käyttää kuhunkin sanaan vastaamiseen niin pitkään kuin halusivat, tässä tutkielmassa sovellettiin versiota, jossa vastaajalla oli maksimissaan 4000 millisekuntia aikaa vastata, ennen kuin testi siirtyi automaattisesti eteenpäin seuraavaan sanaan. Itse testin ohella muutoksia tehtiin taustatietolomakkeeseen, joka oli alkuperäistä lomaketta suppeampi. Taustatietolomakkeen supistamisen ajateltiin laskevan kynnystä tutkimukseen osallistumiseen. Lomakkeessa kerättiin siis vain edellä mainitut tiedot, eikä esimerkiksi tietoa kielihäiriöstä taikka käätisyydestä.

### 2.2.2 Tutkittavat muuttujat

Tutkielman tutkimuskysymyksiin pyrittiin vastaamaan tutkimalla yhdeksää muuttujaa: pisteytyskaavan mukaiset Lexize-pisteet, kunkin osallistujan keskimääräinen reaktioaika, kunkin osallistujan keskimääräinen reaktioaika oikeista vastauksista, kunkin vastaajan keskimääräinen reaktioaika vääristä vastauksista sekä kunkin osallistujan taustatietolomakkeessa raportoidut YKI-pisteet keskimääräisesti sekä neljässä eri modaliteetissa. Ohjelma laski Lexize-pisteet automaattisesti osallistujan suorittaessa testin omalla laitteellaan. Laskettu pistesaldo tallentui tiedostoon, josta salasanasuojattu data oli noudettavissa analyysia varten. Myös kunkin osallistujan sanakohtainen reaktioaika tallentui suorituksen tallentumisen myötä kyseiseen tiedostoon. Keskimääräinen reaktioaika laskettiin näiden tietojen perusteella kullekin osallistujalle laskemalla yhteen hänen kaikkiin vastauksiin kuluttamansa aika ja jakamalla tämä vastausten määrällä. Muut reaktioaikamuuttujat laskettiin samalla periaatteella, mutta toisessa huomioitiin vain oikeisiin vastauksiin kulunut reaktioaika ja toisessa väärin vastauksiin kulunut reaktioaika. Maksimiajan (4000 ms) vaatineet vastaukset jätettiin pois aineistosta. YKI-pisteiden osalta tarkasteltiin keskimääräistä suoriutumista sekä suoriutumista kussakin neljässä modaliteetissa erikseen. YKI-pisteet ovat osallistujien itse raportoimia. Tässä tutkielmassa arvosanat muunnettiin analyysien helpottamiseksi numeeriseen muotoon niin, että alin arvosana A1 oli 1, toiseksi alin arvosana A2 = 2, korkein arvosana C2 = 6 ja niin edelleen. Varsinaisiin tutkimuskysymyksiin vastaamiseen liittyvien muuttujien ohella kerättiin tietoa muuttujista, joilla saattaa olla vaikutusta analyysien lopputulokseen. Tällaisia olivat reaktioajan osalta vastaamiseen käytetty laite (kosketusnäyttölaite vai muu) ja YKI-tulosten osalta vuosi, jolloin tutkittava oli suorittanut YKI-tutkinnon. Tutkielmassa käsiteltävien muuttujien ohella osallistujista kerättiin myös yleistä demografista tietoa ja kielitaustaan liittyvää tietoa, jota voidaan hyödyntää myöhemmässä tutkimuksessa.

### 2.3 Tutkimuksen kulku

Tutkimuksen datankeruu toteutettiin verkkoympäristössä ajanjaksolla 14.04,26.06.2022.

Tutkimukseen kerättiin osallistujia ilmoituksella (liite 1), jonka luettuaan osallistujat saattoivat siirtyä suoraan tutkimukseen linkkiä seuraamalla. Linkkiä seurattuaan osallistujat pääsivät tutustumaan oikeuksiaan, riskejä, suostumusta ja arvontaa käsittelevään suostumussivuun (liite 4), jonka jälkeen he pääsivät siirtymään taustatietolomakkeelle (liite 4). Taustatietolomakkeelta pääsi siirtymään Lexize-testiin klikkaamalla ”Done”-nappulaa näytöllä. Tämän jälkeen osallistujat pääsivät suorittamaan Lexize-testin, jossa heitä kehoitettiin vastaamaan mahdollisimman nopeasti ja oikein. Lexize-testissä osallistujalle esitettiin oikeita suomen kielen sanoja ja epäsanoja. Osallistujan tuli kunkin sanan kohdalla valita joko ”kyllä” tai ”ei” -painike näytöltä joko sormellaan (kosketusnäyttölaitteet) tai kursorilla (tietokone). Ennen kutakin sanaa näytölle ilmestyi plusmerkki 1.5 sekunniksi. Testi siirtyi seuraavaan sanaan joko osallistujan valitessa vastauksensa tai automaattisesti viimeistään neljän sekunnin kuluttua. Testi päättyi automaattisesti, jonka jälkeen osallistujalla oli mahdollisuus tarkastella suoriutumistaan palkkimuotoisesta kaaviosta. Näkyvässä tutkittava saattoi myös päättää, lähettääkö tulokset tutkimusryhmälle analysoitavaksi ja osallistuuko arvontaan.

### 2.4 Tilastolliset menetelmät

Tässä tutkielmassa hyödynnettiin tilastollisena menetelmänä Pearsonin osittaiskorrelaatiokertoimia. Korrelaatiota hyödynnettiin yhteyksien tarkasteluun, sillä se on yleisesti sovellettu käytäntö sanavarastotestien rinnakkaisvaliditeetin selvittämisessä. Tämä on nähtävissä muun muassa Lexizen edeltäjiä käsittelevissä tutkimuksissa (Lemhöfer & Broersma, 2012; Nakata, ym., 2020). Lisäksi kätevyystään huolimatta regressioanalyysin kaltaiset ennustavat tilastolliset menetelmät on perinteisesti nähty testien validointiin sopimattomina tai vähintäänkin tarpeettoman monimutkaisina (Shmueli, 2010). Pearsonin korrelaatiokerrointa hyödynnettiin, sillä aineiston koko mahdollistaa keskeisen raja-arvolauseen nojalla normaalijakauman olettamisen. Lisäksi aineisto on mitattu vähintään välimatka-asteikolla ja sirontakaavioiden perusteella tehdyn päättelyn osalta voitiin todeta, että muuttujien väliset yhteydet ovat lineaarisia lukuun ottamatta reaktioaikamuuttujien ja YKI-tulosten välisiä sirontakaavioita, joista lineaarisia yhteyksiä ei ollut kuvioista havaittavissa. Tuloksia tarkasteltiin kuitenkin myös Spearmanin korrelaatiokertoimen valossa. Nämä tulokset eivät eronneet Pearsonin korrelaatiokertoimella saaduista tuloksista. Lexize-pisteiden ja YKI-tulosten välisten korrelaatioiden tarkastelussa otettiin huomioon se, että tutkittavat olivat suorittaneet YKI-tutkintoja eri vuosina. YKI-testin ja Lexize-testin suorittamisen välillä oli

vaihtelevia aikajaksoja, minkä ajateltiin voivan vaikuttaa yhteyksiin. Tämä huomioitiin hyödyntämällä osittaiskorrelaatiota, jossa kontrolloituna muuttujana toimi YKI-tutkinnon suorittamisvuosi.

Reaktioajan ja tarkkuuden yhteisvaikutusta vertailtiin standardoimalla reaktioaikamuuttujat ja Lexize-pisteet Z-pisteiksi SPSS:n descriptives-toimintoa hyödyntämällä ja yhdistämällä ne summamuuttujiksi. Standardoidut Lexize-pisteet yhdistettiin siis standardoituun kaikista vastauksista laskettuun reaktioaikaan, standardoituun oikeista vastauksista laskettuun reaktioaikaan sekä standardoituun vääristä vastauksista laskettuun reaktioaikaan. Tämän jälkeen tarkasteltiin, korreloivatko nämä summamuuttujat paremmin YKI-testin kanssa kuin yksittäinen Lexize-pistemäärä. Tämäkin analyysi kontrolloitiin YKI-vuodella. Tähän tarkasteluun on saatu inspiraatiota Harringtonin ja Careyn (2009) tutkimuksesta, jossa myös vertailtiin reaktioajan ja tarkkuuden yhdistelmää pelkkään tarkkuuteen. Yhdistelmämuuttujiin pohjautuva menetelmä poikkeaa kuitenkin Harringtonin ja Careyn (2009) hyödyntämästä regressiomallista, eivätkä tutkimusten menetelmät näin ollen ole vertailtavissa.

## **2.5 Tutkimuksen eettisyys**

Tutkimukseen osallistuminen oli suostumusperusteista. Tutkimus ei ollut pitkäkestoinen tai invasiivinen. Tutkittavat suorittivat tutkimuksen kotiympäristössä omilla laitteillaan. Tutkittavilla oli mahdollisuus keskeyttää tutkimus milloin tahansa ja olla lähettämättä tuloksiaan tutkimusryhmälle. Tutkittavia informoitiin heidän tietojensa käsittelystä, säilytyksestä ja jakamisesta. Lisäksi tutkittaville annettiin tutkijoiden yhteystiedot, mikäli he halusivat myöhemmin poistaa osansa aineistosta tai tiedustella jotain muuta tutkimukseen liittyvää seikkaa. Aineistoa käsiteltäessä tiedot anonymisoitiin heti kun aineistosta oli saatu seulottua pois saman henkilön useammat suorituskerrat poistamalla ainoana tunnistemuuttujana toiminut sähköpostiosoite. Tutkittavista kerättiin vain käytänteiden mukaisia yleisiä tietoja ja tutkimuksen kysymysten kannalta olennaisia tietoja. Tutkimuksen haitat arvioitiin minimaalisiksi, ja niiden ajateltiin koskevan lähinnä silmien väsymistä näyttötyöskentelyssä. Silmien lepuuttaminen ei ollut mahdollista ajastetun testin aikana. Eettisen toimikunnan kantaa tutkimukseen ei vaadittu, sillä tutkimuksen riskit ja invasiivisuus ovat pieniä.

### 3 Tulokset

Tässä tutkielmassa tarkasteltiin 65:n aikuisen suomen kielen L2-oppijan Lexize-testin tulosten ja itseraportoitujen suomen kielen YKI-tutkinnon tulosten välistä yhteyttä. Lisäksi tarkasteltiin näiden tutkittavien Lexize-testin reaktioaikojen ja YKI-tulosten välistä korrelaatiota. Tässä osiossa esitellään aluksi aineiston analysoinnin kulku ja tämän jälkeen taulukoituja tunnuslukuja. Lopussa esitetään tutkimuskysymyksiin vastaavat Pearsonin korrelaatiokertoimiin pohjautuvat tulokset.

Aineiston analysointi

#### 3.1 Datan valmistelu

Ennen aineiston analysointia data käsiteltiin analysoitavaan muotoon Excel-ohjelmalla (Versio 2211, Microsoft Corporation, 2022). Tämä prosessi aloitettiin poistamalla datasta sisäänottokriteerejä vastaamattomat suoritukset. Tämän jälkeen datasta tunnistettiin ne henkilöt, jotka olivat tehneet Lexize-testin useampaan kertaan. Näiltä henkilöiltä datassa huomioitiin vain ensimmäinen suorituskerta harjoitusvaikutusten ja huijaamisen välttämiseksi. Näin syntyneestä otoksesta poistettiin tunnistetiedot (sähköposti) ja vastaamatta jääneet arvot (vastaukset, joiden reaktioaika oli 4000 ms). Tämän jälkeen tunnistettiin keskimääräisen reaktioajan perusteella yli kaksi keskihajontaa poikkeavat ääriarvot IBM SPSS Statistics-ohjelmistoa (versio 28.0.0.0, IBM Corp., 2021) hyödyntämällä. Tämä johti kolmen nopeimman (keskimääräinen vastausaika alle 300 ms) tutkittavan poistamiseen, sillä ei voida pitää realistisena, että tässä ajassa vastaaja olisi todella keskittynyt vastaamiseen. Tätä väitettä voidaan perustella esimerkiksi sillä, etteivät äidinkielenäkään kieltä käyttävät vastaajat ole läheskään näin nopeita (esimerkiksi Keuleers, Lacey, Rastle & Brysbaert, 2012). Myös se, että väri- ja oikeiden vastausten suhde oli lähellä satunnaista vastaamista (50 %, 57 % ja 60 % vastauksista oikein) puolsi näiden osallistujien aineiston ulkopuolelle jättämistä. Yhden heistä kohdalla kaikki vastaukset eivät myöskään olleet tallentuneet oikein. Hitaampien vastaajien osalta yksi vastaaja jätettiin pois, sillä hänen suoriutumisessaan oli nähtävissä maksimaalinen 4000 sekunnin reaktioaika koko testin alkuosan ajan, josta voidaan päätellä, ettei osallistuja ole tällöin suorittanut testiä oikein. Muut hitaammat vastaajat päätettiin pitää mukana aineistossa, sillä heidän suoritustensa voidaan ajatella edustavan populaatiossa esiintyvää todellista suoriutumista. Tätä voidaan perustella esimerkiksi sillä, että Pellicer-Sánchezin ja Schmittin (2012) tutkimuksessa 3000 millisekuntia pidettiin rajana liian hitaalle vastaamiselle vastaavassa tilanteessa, ja kaikkien näiden osallistujien keskimääräinen reaktioaika oli reilusti tämän alle. Näin syntyi lopullinen 65:n tutkittavan näyte. Näistä 41 henkilöä suoritti kokeen kosketusnäyttölaitteella ja 24 tietokoneella.

Koska Lexize-testissä käytettävä verkkosivusto oli suunniteltu laajempaa tutkimusta varten, kertyi dataan myös paljon ylimääräisiä muuttujia. Nämä muuttujat poistettiin aineistosta. Jäljelle jäävät muuttujat olivat koehenkilönumero, modaaliteettien mukaan jaotellut itseraportoidut YKI-arvosanat (A1-C2), testin suorittamiseen käytetty laite (kosketusnäyttö vai ei), Lexize-testin pisteytyskaavan mukainen pistemäärä, YKI-testin suorittamisajankohta (vuosina) sekä vastauskohtainen reaktioaika (ms). YKI-arvosanat muutettiin analyysia varten numeeriseen muotoon (A1=1, A2=2, B1=3 ja niin edelleen). Aineistoon kertyneiden arvojen perusteella aineistoon laskettiin myös yksilön YKI-kokonaissuoriutuminen (kaikkien modaaliteettien arvosanojen keskiarvo) sekä kolme reaktioaikamuuttujaa: yksilön keskimääräinen reaktioaika, yksilön keskimääräinen reaktioaika oikeissa vastauksissa ja yksilön keskimääräinen reaktioaika väärissä vastauksissa.

### 3.2 Datatarkastelu

Aineistoa tarkasteltiin tutkimalla sen tunnuslukuja sekä jakaumaa. Aineiston tarkasteluun ja analysointiin käytettiin IBM SPSS Statistics –ohjelmistoa (versio 28.0.0.0) (IBM Corp., 2021). Aineistossa ei ollut puuttuvia arvoja. Aineiston tunnuslukuja esitellään seuraavissa taulukoissa.

**Taulukko 3**

*Lexize-pisteiden ja YKI-arvosanojen keskiarvo, keskiarvon luottamusväli, keskihajonta sekä jakauma*

| Testi              | <i>ka</i> | <i>ka:n lv</i> | <i>kh</i> | <i>vv</i> | <i>Jakauman normaalius</i>                            |
|--------------------|-----------|----------------|-----------|-----------|---|
| Lexize             | 16.51     | 13.89–19.13    | 10.58     | 1–41      | Noudattaa normaalijakaumaa<br>D(65) = 0.11, p = .95   |
| YKI-kokonaistulos  | 3.40      | 3.17–3.64      | 0.94      | 1.25–6.00 | Ei noudata normaalijakaumaa<br>D(65) = 0.17, p < .001 |
| YKI-lukeminen      | 3.57      | 3.31–3.83      | 1.03      | 1–6       | Ei noudata normaalijakaumaa<br>D(65) = 0.25, p < .001 |
| YKI-kirjoittaminen | 3.34      | 3.08–3.61      | 1.05      | 1–6       | Ei noudata normaalijakaumaa<br>D(65) = 0.29, p < .001 |
| YKI-puhuminen      | 3.31      | 3.07–3.54      | 0.95      | 1–6       | Ei noudata normaalijakaumaa<br>D(65) = 0.29, p < .001 |
| YKI-kuunteleminen  | 3.40      | 3.16–3.64      | 0.98      | 1–6       | Ei noudata normaalijakaumaa<br>D(65) = 0.30, p < .001 |

*ka*=keskiarvo, *lv*=luottamusväli (95%), *kh*=keskihajonta, *vv*=vaihteluväli

Taulukosta 3 esitellään Lexize-pisteiden ja YKI-tulosten tunnuslukuja. Aineiston jakauma oli Lexize-pisteiden osalta melko litteä. Huipukkuuden arvo lähentelee keskivirhe huomioiden

itseisarvoltaan yhtä ( $-0.99$ ,  $S=0.58$ ). Logaritmi- tai neliöjuurikorjauksilla ei ollut vaikutusta tähän. Kaikkien YKI-muuttujien huipukkuus oli arvoltaan yli yhden (kokonaistulos:  $1.58$ ,  $S=0.59$ ; lukeminen:  $1.20$ ,  $S=0.59$ ; kirjoittaminen:  $1.22$ ,  $S=0.59$ ; puhuminen:  $1.40$ ,  $S=0.59$ ; kuunteleminen:  $1.98$ ,  $S=0.59$ ). Suurin osa YKI-muuttujista oli myös melko vinoja (kokonaistulos:  $0.74$ ,  $S=0.30$ ; lukeminen:  $0.78$ ,  $S=0.30$ ; kirjoittaminen  $0.70$ ,  $S=0.30$ ; kuunteleminen  $0.75$ ,  $S=0.30$ ). Otoskoko oli kuitenkin yli 30 ( $N=65$ ), jota suositellaan rajaksi, jonka perusteella aineiston voidaan olettaa olevan keskeisen raja-arvolauseen mukaan normaalisti jakautunut (esim. Kwak & Kim, 2017). Oletetaan siis aineiston normaalijakautuneisuus. Hajontakaaviota (ks. Kuvaaja 1 kohdassa 3.2 Korrelaatioanalyysit) tarkastelemalla nähdään Lexize-pisteiden ja YKI-tulosten välisen yhteyden olevan lineaarinen. Näiden perusteella hyödynnetään tarkastelussa parametrista korrelaatiota.

**Taulukko 4**

*Lexize-reaktioaikojen keskiarvo, keskihajonta, vaihteluväli sekä Shapiro-Wilkin mukainen normaalijakaumatestitulokset käytetyn laitteen mukaan*

| Käytetty laite      | Laskentatapa      | Yksilön keskimääräinen reaktioaika (ms) |                |           |              | Jakauman normaalisuus                                 |
|---------------------|-------------------|---|----------------|-----------|--------------|---|
|                     |                   | <i>ka</i>                               | <i>ka:n lv</i> | <i>kh</i> | <i>vv</i>    |   |
| Kosketusnäyttölaite | Kaikki vastaukset | 1423                                    | 1300–<br>1547  | 391       | 400–<br>2476 | Ei noudata normaalijakaumaa<br>$W(41)=0.94$ , $p=.04$ |
|                     | Oikeat vastaukset | 1426                                    | 1301–<br>1550  | 395       | 393–<br>2384 | Ei noudata normaalijakaumaa<br>$W(41)=0.94$ , $p=.03$ |
|                     | Väärät vastaukset | 1422                                    | 1297–<br>1546  | 394       | 408–<br>2546 | Ei noudata normaalijakaumaa<br>$W(41)=0.94$ , $p=.02$ |
| Tietokone           | Kaikki vastaukset | 1352                                    | 1206–<br>1497  | 344       | 907–<br>1989 | Ei noudata normaalijakaumaa<br>$W(24)=0.91$ , $p=.04$ |
|                     | Oikeat vastaukset | 1340                                    | 1196–<br>1484  | 341       | 919–<br>2037 | Ei noudata normaalijakaumaa<br>$W(24)=0.91$ , $p=.03$ |
|                     | Väärät vastaukset | 1372                                    | 1224–<br>1520  | 351       | 886–<br>2055 | Noudattaa normaalijakaumaa<br>$W(24)=0.93$ , $p=0.8$  |

*ka*=keskiarvo, *lv*=luottamusväli (95%), *kh*=keskihajonta, *vv*= vaihteluväli

Taulukosta 4 nähdään, että reaktioaikamuuttujien ominaisuudet vaihtelevat laskentatavan ja tutkittavan käyttämän laitteen mukaan. Tietokoneella vastanneiden osallistujien vastausnopeuden



keskiarvot vaikuttaisivat olevan hiukan nopeampia kuin kosketusnäyttölaitteella vastanneiden henkilöiden.

Koska tietokoneella testin tehneiden henkilöiden määrä ( $N=24$ ) ei ollut tarpeeksi suuri täyttämään normaalijakaumaoletusta ja Shapiro-Wilkin normaalijakaumatesti osoitti, ettei tietokoneella kokeen tehneiden henkilöiden aineisto ollut normaalisti jakaantunut muuten kuin väärin vastausten kohdalla, verrattiin ryhmiä toisiinsa Mann-Whitneyn U-testin avulla. Mann-Whitneyn U-testin mukaan ryhmien välillä ei ollut tilastollisesti merkitsevää eroa yhdessäkään laskentatavassa (kaikki vastaukset:  $U=1.10$ ,  $p=.27$ ; oikeat vastaukset:  $U=1.28$ ,  $p=.20$ ; väärät vastaukset:  $U=0.83$ ,  $p=.41$ ). Koska kahden laitteen välillä tehtyjen suoritusten reaktioajoissa ei näin ollen kuitenkaan ollut merkitsevää eroa, tarkasteltiin reaktioaikamuuttujia yhdessä laitteesta riippumattomana kokonaisuutena. Koska tällöin tutkittavien määrä oli 65, voitiin myös reaktioaikojen osalta keskeisen raja-arvolauseen nojalla hyödyntää parametrissa korrelaatiota.

### 3.3 Korrelaatioanalyysit

#### 3.3.1 Lexize-pisteiden ja itseraportoitujen YKI-tulosten välinen yhteys

Ensimmäisenä tutkimuskysymyksenä tarkasteltiin YKI-testin kokonaistuloksen ja modaliteettien mukaisten tulosten yhteyttä pisteytyskaavan mukaisiin Lexize-pisteisiin. Osittaiskorrelaatio, jossa YKI-tutkinnon suorittamisvuosi oli huomioitu kontrolloituna muuttujana, antoi suuremmat korrelaatiot Lexize-pisteiden ja YKI-tulosten välille kuin pelkkä Pearsonin korrelaatiokerroin. Yleiskuva osittaiskorrelaation mukaisista tuloksista on nähtävissä taulukosta 5.

**Taulukko 5**

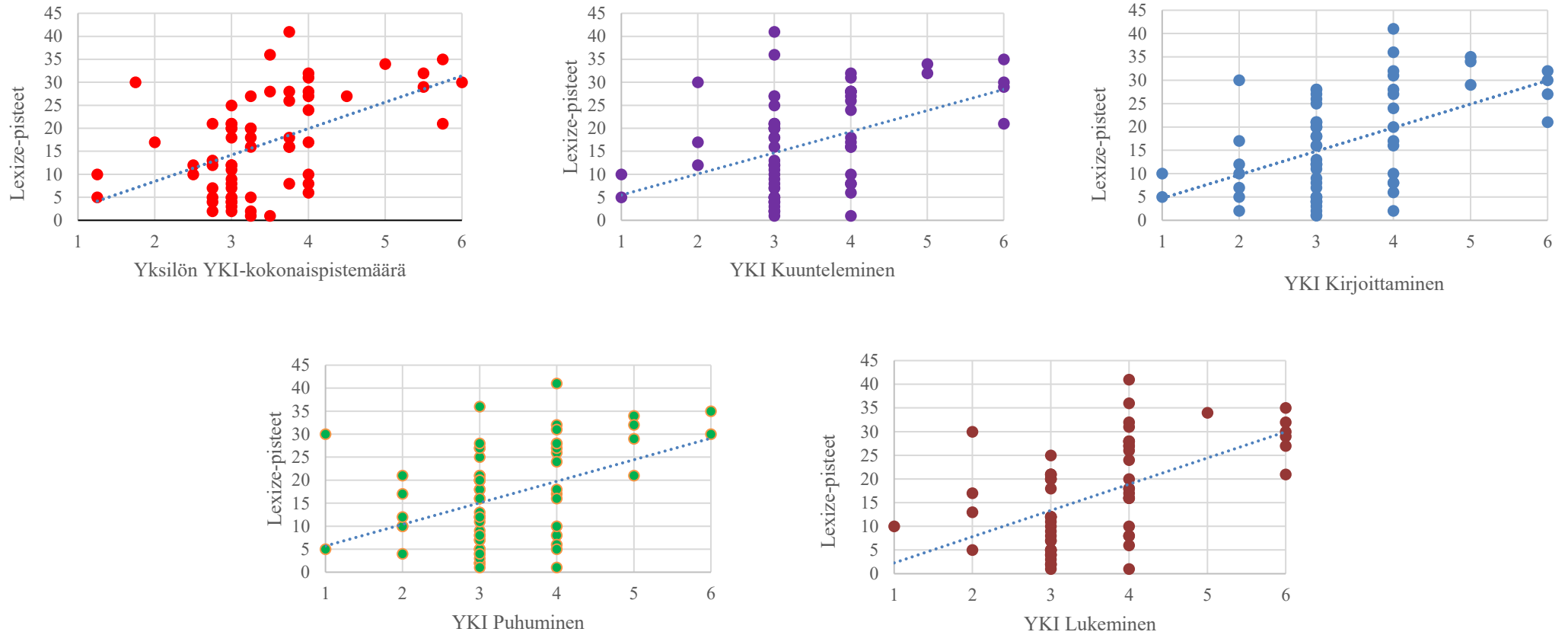
*YKI-tulosten ja Lexize-pisteiden välinen parametrinen osittaiskorrelaatio kontrolloituna YKI-vuodella, sen tilastollinen merkitsevyys sekä selitysaste*

| YKI-modaliteetti | Pearsonin korrelaatiot Lexize-pisteiden kanssa |                            |                       |
|------------------|--|----------------------------|-----------------------|
|                  | r  | Tilastollinen merkitsevyys | Selitysaste ( $r^2$ ) |
| Kaikki           | .56  | $p<.001$                   | .31                   |
| Lukeminen        | .60  | $p<.001$                   | .36                   |
| Kirjoittaminen   | .53  | $p<.001$                   | .28                   |
| Puhuminen        | .48  | $p<.001$                   | .23                   |
| Kuunteleminen    | .50  | $p<.001$                   | .25                   |

Taulukon 5 mukaiset tulokset viittaavat siihen, että Lexize-pistemäärän ja YKI-tuloksen välillä oli merkitsevä yhteys. Kaikki korrelaatiot olivat tilastollisesti merkitseviä. Korrelaatiot olivat kaikilla osa-alueilla myös suuruudeltaan melko korkeita. Parhaiten korreloivat Lexize-pistemäärä ja lukemisen arvosana YKI-tutkinnosta, huonoiten taas Lexize-pistemäärä ja puhumisen arvosana YKI-tutkinnosta. Tulosten perusteella havaitun korrelaation voidaan ajatella tukevan Lexize-testin luotettavuutta kielitaidon arvioinnin välineenä. Visuaalisesti tilannetta havainnollistaa kuvaaja 1.

## Kuvaaja 1

Lexize-pisteiden ja YKI-tulosten välistä korrelaatioita kuvaavat hajontakaaviot ja trendiviivat



### 3.3.2 Lexize-reaktioaikojen ja YKI-tulosten väliset yhteydet

Seuraavaksi tarkasteltiin Lexize-reaktioaikojen yhteyttä itseraportoituihin YKI-tuloksiin kolmesta eri näkökulmasta. Reaktioaikoihin liittyvä päätutkimuskysymys keskittyi vertaamaan kokonaisreaktioajan kykyä ennustaa suoriutumista YKI-testissä. Asiaa lähestyttiin tarkastelemalla Pearsonin korrelaatioita yksilökohtaisen kokonaisreaktioajan ja YKI-tulosten välillä. Myös tällöin YKI-tutkinnon suoritusvuosi asetettiin kontrolloivaksi muuttujaksi. Nämä korrelaatiot olivat arvoiltaan hyvin pieniä eivätkä osoittautuneet tilastollisesti merkitseväksi yhdenkään YKI-modaliteetin kohdalla (YKI-kokonaissuoriutuminen:  $r(62) = .023, p=.86$ ; lukeminen:  $r(62) = -.05, p=.73$ ; kuunteleminen:  $r(62) = .02, p=.90$ ; puhuminen:  $r(62) = -.00, p=.99$ ; kirjoittaminen:  $r(62) = .11, p=.38$ ). Näin ollen vastaus kysymykseen YKI-tulosten ja keskimääräisen kokonaisreaktioajan yhteydestä ei tukenut hypoteesia: keskimääräinen reaktioaika Lexize-testissä ei ole yhteydessä suoriutumiseen YKI-tutkinnossa.

On kuitenkin mielenkiintoista tarkastella asiaa myös tarkemmin. Ensimmäinen alakysymys keskittyi pohtimaan, millainen yhteys voisi löytyä tarkastelemalla pelkästään oikeita tai vääriä vastauksia. Tarkasteluun hyödynnettiin jälleen Pearsonin korrelaatiokerrointa YKI-vuodella kontrolloituna. Tulokset myötäilivät kuitenkin kaikista vastauksista saatua kuvaa; merkitsevää korrelaatiota ei löytynyt YKI-kokonaistuloksen tai eri modaliteettien välille pelkästään oikeita vastauksia tarkastelemalla: kokonaistulos:  $r(62) = .03, p=.84$ ; lukeminen:  $r(62) = -.04, p=.73$ ; kuunteleminen:  $r(62) = .02, p=.90$ ; puhuminen:  $r(62) = .01, p=.97$ ; kirjoittaminen:  $r(62) = .11, p=.37$  tai vääriä vastauksia tarkastelemalla: (kokonaistulos:  $r(62) = .02, p=.88$ ; lukeminen:  $r(62) = -.04, p=.74$ ; kuunteleminen:  $r(62) = .01, p=.91$ ; puhuminen:  $r(62) = -.01, p=.92$ ; kirjoittaminen:  $r(62) = .11, p=.39$ ). Näin ollen tämän tutkielman perusteella vaikuttaisi, ettei Lexize-testin reaktioajalla ole merkitsevää roolia YKI-testissä suoriutumisen ennustajana.

Toinen reaktioaikoihin liittyvä alakysymys keskittyi vastaamaan siihen, ovatko Lexize-pisteet yhdessä reaktioajan kanssa parempi YKI-suoriutumisen ennustaja kuin pelkät Lexize-pisteet. Tuloksiin voi tutustua lähemmin taulukosta 6.

**Taulukko 6**

*Pelkkien Lexize-pisteiden ja reaktioaikamuuttujiin yhdistettyjen Lexize-pisteiden korrelaatiot YKI-tulosten kanssa kontrolloituna YKI-suoritusvuodella*

| YKI-modaliteetti | Laskentatapa   |          |  |          |  |          |  |          |
|------------------|----------------|----------|--|----------|--|----------|--|----------|
|                  | Lexize-pisteet |          | Lexize-pisteet ja RT <sub>kaikki</sub> |          | Lexize-pisteet ja RT <sub>oikeat</sub> |          | Lexize-pisteet ja RT <sub>väärät</sub> |          |
|                  | <i>r</i>       | <i>p</i> | <i>r</i>                               | <i>p</i> | <i>r</i>                               | <i>p</i> | <i>r</i>                               | <i>p</i> |
| Kaikki           | .56            | <.001    | .41                                    | <.001    | .42                                    | <.001    | .41                                    | <.001    |
| Lukeminen        | .60            | <.001    | .39                                    | .002     | .39                                    | .001     | .38                                    | .002     |
| Kirjoittaminen   | .53            | <.001    | .45                                    | <.001    | .45                                    | <.001    | .45                                    | <.001    |
| Puhuminen        | .48            | <.001    | .33                                    | .007     | .34                                    | .006     | .32                                    | .01      |
| Kuunteleminen    | .50            | <.001    | .36                                    | .003     | .36                                    | .003     | .36                                    | .004     |

Kuten taulukosta 6 on havaittavissa, osoittautui, ettei millään laskentatavalla lasketun reaktioaikamuuttujan lisääminen Lexize pisteisiin nosta Lexize-pisteiden korrelaatiota YKI-suoriutumisen kanssa. Vaikka myös summamuuttujien korrelaatiot YKI-testin kanssa ovat suurilta osin merkitseviä, jää korrelaatioiden suuruus aina pelkkää Lexize-pistemäärää selvästi pienemmäksi. Näin ollen voidaan päätellä, ettei reaktioajan mittaaminen tuo tämän tutkielman perusteella lisäarvoa Lexize-testille.

## 4 Pohdinta

Tässä tutkielmassa tarkasteltiin tutkittavien Lexize-testistä saatuja pistemääriä ja reaktioaikoja suhteessa YKI-tutkinnosta saatuihin tuloksiin. Tähän liittyen esitettiin kaksi tutkimuskysymystä, joista ensimmäinen käsitteli Lexizessä ja YKI-tutkinnossa suoriutumisen välistä yhteyttä. Alakysymyksenä tarkasteltiin modaliteettikohtaisia eroja testien välisissä yhteyksissä. Toinen kysymys tarkasteli Lexize-testin reaktioaikojen yhteyttä YKI-tutkinnossa suoriutumiseen. Etsimällä vastauksia näihin kysymyksiin pyrittiin jatkamaan Lexize-testin validointia ja selvittämään, voidaanko testiä kehittää ottamalla siihen mukaan tutkittavien reaktioaikaa mittaava komponentti. Tutkielman tulokset osoittavat hypoteesin mukaisesti, että Lexize-pistemäärä korreloi vahvasti tai vähintään kohtalaisesti YKI-tutkinnon eri osa-alueiden kanssa. Korrelaatio YKI-tutkinnon tulosten kanssa on suuntaa antava osoitus Lexize-testin rinnakkaisvaliditeetista. Lisäksi korrelaatio YKI-testin kanssa viittaa Lexizen yhteyksiin eri kielenkäytön modaliteettien kanssa, erityisesti kuullun ymmärtämisen, lukemisen ja kirjoittamisen taitojen osalta. Täten vaikka Lexize on ensisijaisesti sanavarastotesti, voidaan sen ajatella kykenevän ennustamaan myös kokonaisvaltaisemmin eri kielenkäytön osa-alueita. Tätä tukevat myös monet aiemmat tutkimukset, joissa on tarkasteltu sanaston roolia kielen oppimisessa ja käytössä (esimerkiksi Laufer & Ravenhorst-Kalovski, 2010; Stæhr, 2008; Lin & Morrison, 2010) - sanasto on kokonaiskielitaidon pohja.

Toisaalta tulokset osoittavat myös, että Lexize-testissä mitatun reaktioajan tarkastelusta ei ole hyötyä YKI-suoriutumisen ennustamisessa. Reaktioajalla ei ole merkitystä laskettiinpa se kaikista vastauksista, vain oikeista vastauksista tai vain vääristä vastauksista. Reaktioajan yhdistäminen Lexize-pisteisiin ei myöskään paranna pisteiden kykyä ennustaa YKI-suoriutumista. Tulos ei siis tue hypoteesia yhdessä kolmesta reaktioaikaan liittyvästä kysymyksestä ja on näin ristiriidassa johdannossa esiteltyjen aiempien tutkimustulosten kanssa. Tätä ristiriitaa käsitellään tarkemmin myöhemmissä kappaleissa.

Yllätyksellisyydestään huolimatta tulos on tärkeä, sillä se antaa viitteitä siitä, ettei reaktioajan lisäämisestä testiin ole hyötyä testin kehittämisessä. Lexize vaikuttaisi siis sopivan paremmin sovellettavaksi yksinkertaisemmassa muodossa ilman reaktioaikojen mittaamista. Toisaalta kyllä/ei-testien reaktioajan tarkkailu on tulevaisuuden tutkimuksen kannalta mielenkiintoista, sillä aiheesta on toistaiseksi tehty kovin vähän tutkimusta.

#### 4.1 Lexize-testin ja YKI-tutkinnon yhteyden merkitys Lexize-testin validiteetille

Tutkielman tulokset osoittavat, että YKI-tutkinnon itseraportoitujen arvosanojen ja Lexize-testin tuloksen välillä on merkitsevä korrelaatio kaikissa modaliteeteissa. Tämä myötäilee hypoteesia näiden kahden testin välisestä yhteydestä ja tukee tutkielman tarkoitusta viedä Lexize-testin validiteettia entistä pidemmälle. Suhteessa korrelaatioihin yleisen kielitaidon mittarin kanssa Lexize myös myötäilee edeltäjiään. Esimerkiksi LexTalen ja laajempien kielitaidon mittareiden TOEIC-testin sekä TOEFL-testin välillä on havaittu merkitseviä korrelaatioita (Lemhöfer & Broersma, 2021; Nakata ym., 2020). Nakata ja tutkimusryhmä (2020) kuitenkin kirjoittavat artikkelissaan, että vaikka LexTalen voidaan useissa tilanteissa todeta tuottavan parempia korrelaatioita muiden kielitaidon mittareiden kanssa kuin itsearvioiden, tulisi mieluummin suosia muita testejä niissä tilanteissa, joissa niille on aikaa. Lisäksi he kirjoittavat LexTalen olevan merkitsevistä korrelaatioista huolimatta parempi sanastomittari kuin yleisen kielitaidon mittari, ja suosittavat näin osaltaan pidättäytymään testin alkuperäisessä tarkoituksessa. He muistuttavat, että vaikka LexTalen korrelaatiot yleisen kielitaidon mittareiden kanssa ovat merkitseviä, ovat ne suuruudeltaan vain keskikokoisia. Myös Lexize-pistemäärän ja YKI-tutkinnon arvosanojen välisiä korrelaatioita on siis syytä tarkastella kriittisesti. YKI-tutkintoa voidaan pitää luotettavana kielitaidon mittarina suuren osallistujamääränsä, laajuutensa ja siihen kohdistuvan tutkimusmäärän ja asiantuntijoiden valvonnan myötä, joten YKI-tutkintoon vertaaminen on perusteltua. Kuitenkin vaikka tässä tutkielmassa osoitetut korrelaatiot Lexizen ja YKI-tutkinnon välillä olivat suuruudeltaan vahvoja Cohenin (1988) yleisesti hyödynnetyllä asteikolla ja suurempia kuin LexTalen korrelaatiot yleisen kielitaidon mittareiden kanssa Lemhöferin ja Broersman (2012) sekä Nakatan ja kumppaneiden (2020) tutkimuksessa, on muistettava korrelaatiota koskevat rajoitteet.

Vaikka Cohenin (1988) asteikko on laajalti käytössä, Sadeghi (2013) huomauttaa, että korrelaatioiden kokojen raportointi vaihtelee tutkimusten välillä. Sadeghi itse suosittaa rajaksi .80-suuruista korrelaatiokertoimen arvoa, mikäli korrelaation perusteella on tarkoitus korvata aiempi testi tai ennustaa sen tuloksia. Tähän rajaan Lexizen korrelaatiot YKI-tutkinnon kanssa eivät yllä. Sadeghi muistuttaa myös, ettei tilastollisesti merkitsevä korrelaatio ole aina merkityksellinen korrelaatio, ja viittaa Brownin ja Rodgersin (2002) kehoitukseen korottaa kaikki korrelaatiot toiseen potenssiin, jotta voidaan nähdä korrelaation selitysaste. Näin tehtäessä Lexize-pisteet selittävät noin kolmasosan YKI-arvosanoissa havaitusta vaihtelusta. Sadeghin mukaan hyväksyttävän selitysasteen suuruus vaihtelee riippuen esimerkiksi siitä,

mihin tarkoitukseen korrelaatiota ollaan käyttämässä. Lexize-testin tarkoitus ei ole korvata YKI-tutkintoa esimerkiksi kansalaisuuden hakemiseen tarvittavana kielitutkintona, vaan se on ensisijaisesti luotu tutkimuskäyttöön antamaan objektiivisempi vaihtoehto kielitaidon itsearvioinneille tutkimuskontekstissa. Näin ollen Lexizen antama tulos ei vaikuta raskauttavasti tutkittavan tulevaisuuteen, ja selityksastetta voidaan pitää riittävänä. Testiä jaettaessa on toki myös tehtävä testin käyttäjille selväksi se, millaiseen käyttöön testi on tarkoitettu. Schmitt ja tutkimusryhmä (2019) muistuttavat, että testin kehittäjät ovat ne, jotka ovat lopulta vastuussa testin oikeanlaisesta käytöstä. Näin he peräänkuuluttavat testien tarkkaa ja totuudenmukaista kuvaamista: Lexizen tapauksessa tulisi siis selkeästi tuoda esille, että kyseessä on vastaanottavan sanaston laajuuden testi aikuisille suomen kielen puhujille niihin tilanteisiin, joissa on tarve nopealle ja pintapuoliselle arviolle kielitaidosta. Lisäksi voidaan mainita, että testi toimii paremmin L2-puhujilla, eikä kykene enää erottelemaan L1-puhujia korkeimmilla koulutustasoilla (Salmela ym., 2021). On painotettava, ettei testituloksella saa olla raskauttavia seurauksia testattavalle. Toisaalta koska Lexizen pääasiallinen tarkoitus on toimia tutkimuksen apuvälineenä, voidaan esittää kysymys siitä, kuinka luotettava testi tulisi olla, jotta se ei vääristäisi tutkimuksen tuloksia.

Korrelaation kyseenalaistamisen jälkeen syytä muistaa, että Lexizen validiteettia on tarkasteltu sen kehittäjien osalta jo aikaisemmin (Salmela ym. 2021). Salmelan ja kumppaneiden tutkimuksessa (2021) toteutettiin ja kuvattiin huolellinen itemien valinta, testin korrelaatiot erilaisiin kriteerimuuttujiin kuten ikään, itsearviointeihin, kouluarvosanoihin ja koulutustasoon sekä testin kyky erotella L1- ja L2-puhujat toisistaan ryhmien välillä ja sisällä. Nämä antavat itsessään tukea testin validiteetille, ja on perusteltua olettaa, että jo näiden perusteella Lexize on tarkoituksiinsa nähden melko validi testi. Kuitenkin kuten Fulcher ja Davidson (2009) toteavat, testit eivät koskaan voi olla täysin validoituja ja testin validointi on jatkuva prosessi, jota tulisi tarkastella ajoittain uudelleen testin julkaisun jälkeen. Koska kyllä/ei-testit voivat toimia eri tavoin kielitaustasta ja kulttuuriympäristöstä riippuen (Lemhöfer & Broersma, 2012; Meara, 1990), olisi Lexizenkin kohdalla eduksi tarkastella korrelaatiota erilaisilla ryhmillä. Tietty kulttuuriryhmä voi suosia varovaisempaa vastaustapaa, jolloin olleessaan epävarmoja tiedostaan he valitsevat sanan epäsanaksi, kun taas jokin toinen kulttuuriryhmä voi tyypillisemmin valita epävarmat sanat oikeiksi sanoiksi. Sama testi toimii myös eri tavalla riippuen siitä, minkä tasoisia testattavat ovat (Pellicer-Sánchez & Schmitt, 2012), esimerkiksi jokin tapa pisteyttää testi voi antaa realistisen kuvan matalatasoisten oppijoiden osaamisesta mutta suosia edistyneitä oppijoita tai päinvastoin.



Suomen kielen L2-puhujien suhteellisen pieni määrä voi kuitenkin vaikeuttaa erilaisten alaryhmien tarkastelua käytännössä, sillä pienen kieliryhmän sisällä myös erilaiset ryhmät ovat pienempiä ja näiden saavuttaminen voi olla tutkimukselle haaste. Toisaalta myös esimerkiksi käyttäjäkokemusten kerääminen ja testin seurauksien tarkastelu ovat olennainen osa validiteetin tarkastelua (D'Este, 2012).

Muiden kaltaistensa testien tavoin Lexize kärsii vastausprosesseihin liittyvästä arvaamisen ongelmasta – myöskään tämän tutkielman kohdalla emme voi olla täysin varmoja siitä, moniko oikea vastaus on ollut arvaus ja onko pisteytyskaava kohdellut osallistujia oikeudenmukaisesti, vaikka laskennallisesti pisteytyskaava huomiokin arvaamisen mahdollisuuden. Ohessa voidaan pohtia, onko Lexizen pisteytyskaava paras mahdollinen. Aiemmassa tutkimuksessa kyllä/ei-testien kehittäjät eivät ole yksimielesisiä siitä, mikä olisi paras tapa pisteyttää testit (esim. Beeckmans ym., 2001). Toisaalta tutkimuskirjallisuudessa vertailtujen kyllä/ei-testien pisteytyskaavojen tuottamat tulokset myös korreloivat vahvasti keskenään (Nakata ym., 2020; Lemhöfer & Broersma, 2012), joten kaikkien käytössä olevien kaavojen voidaan ajatella antavan samansuuntaisia tuloksia.

Pohtia voidaan myös sitä, että Lexize-testissä testattava ei voi jättää vastaamatta ärsykkeisiin. Kyllä/ei-testien pakotetun valinnan on pelätty vaikuttavan heikompien suorittajien motivaatioon (Perea ym., 2012). Pakotetun valinnan motivaatioon tuomien vaikutusten lieventämiseksi taas on esitetty esimerkiksi nopeutettua go/nogo-asetelmaa, jossa vastataan mahdollisimman nopeasti ”kyllä” oikeaan sanaan ja jätetään vastaamatta epäsanoihin (Perea ym., 2002). Tällainen asetelma johti Peraan ja tutkimusryhmän (2002) tutkimuksessa parempiin tuloksiin. Vastaavanlaista asetelmaa voisi testata myös Lexizen toteutuksessa. Lisäksi kyllä/ei-testi kertoo vain, että ihminen tunnistaa sanan. Tämä tarkoittaa sitä, ettei ihmisellä välttämättä ole tietoa sanan käytöstä tai merkityksestä (McLean ym., 2015). Tätä asiaa voisi kartoittaa esimerkiksi haastatteleamalla testin tehneitä ihmisiä kuten Pellicer-Sánchez ja Schmitt (2012) tekivät omassa kyllä/ei-testiä tarkastelevassa tutkimuksessaan.

## **4.2 Modaliteettien väliset erot**

Ensimmäisen alakysymyksen osalta tulokset vastasivat pitkälti hypoteesia. Silmämääräisesti tarkasteltuna erot eri modaliteettien välisissä korrelaatioissa olivat varsin pieniä. Tämä on ymmärrettävää, sillä Lexizen ennustuskyky pohjautuu reseptiivisen sanaston arvioimiseen, ja sanasto on tärkeä osa kielen käyttöä kaikissa modaliteeteissa. Aiemman tutkimuksen pohjalta ennustettiin kuuntelemisen ja lukemisen modaliteettien olevan vahvimmin yhteydessä Lexize-

pisteisiin. Näin myös tapahtui. Kuitenkin myös kirjoittamisen kanssa löytyi vahva yhteys, joka ylsi yli kuuntelemisen modaliteetin kanssa löydetyn yhteyden. Tulosten johtopäätöksiin on kuitenkin suhtauduttava varauksella, sillä ne perustuvat pelkkään korrelaatioiden silmämääräiseen vertailuun. Jotta tulokset eri modaliteettien välisistä eroista olisivat luotettavia, tarvittaisiin tarkasteluun menetelmiä, joilla on enemmän tilastollista voimaa.

Kuten jo johdannossa tulee ilmi, Lexizen ennustaman sanaston ja kirjoittamisen välinen vahva korrelaatio ei ole ainutlaatuinen löytö, vaan vahvoja yhteyksiä on löydetty muissakin tutkimuksissa (Stæhr, 2008; Mirapleixin & Muñozin, 2018; Lek & Carson, 1994). Toisaalta löydös antaa viitteitä siitä, että vaikka tuottava kieli vaatii enemmän sanaston syvällistä ymmärtämistä kuin reseptiivinen kielitaito (Qian & Lin, 2000), on sanaston laajuutta mittaava Lexize kykenevä antamaan arvioita myös kirjoittamisen, ja – joskin jonkin verran huonommin – puhumisen taidoissa. Tämä on selitettävissä sanaston syvyyden ja sanaston laajuuden välisillä yhteyksillä (esim. Li & Kirby, 2015; Vermeer, 2001). Toisin sanoen henkilöillä, joilla on mitattu olevan laaja sanasto, on havaittu olevan myös syvällisempi tietämys tuntemista sanoista. Lisäksi korkeaa korrelaatiota kirjoittamisen kanssa voi selittää myös se, että Lexize on visuaaliseen havaintoon perustuva testi, ja myös kirjoittaminen tapahtuu visuaalisessa modaliteetissa. Tätä tukee Coltheartin ja Rastlen (1994) teoria mentaalisen leksikon jakautumisesta fonologiseen ja ortografiseen varastoon. Mikäli sana on opittu ortografisessa muodossa, se tunnustetaan vain tässä muodossa, kunnes se prosessoidaan eteenpäin fonologiseen muotoon ja päinvastoin. Tälle teorialle tukea antaa Miltonin ja kumppaneiden (2010) tutkimus, jossa ortografinen kyllä/ei-testi korreloi paremmin kirjoitetun kielen taitojen kanssa ja kuulohavaintoon perustuva kyllä/ei-testi kuuntelemisen ja puhumisen taitojen kanssa. Tämä siis osoittaa, ettei havainto Lexizen ja YKI-tutkinnon kirjoittamisen arvosanan kanssa ole ainutlaatuinen löydös vaan tukee aiempia löydöksiä. Erona tutkimusten välillä on, että Lexize korreloi hyvin myös kuullun ymmärtämisen kanssa. Näin ollen voidaan ajatella, että sekä tuotto-vastaanottoakseli että ortografis-fonologinen akseli vaikuttavat siihen, kuinka hyvin Lexizen kaltaiset testit ennustavat kielitaitoa. Toisaalta ortografisen ja fonologisen sanaston koon välillä on vahva korrelaatio, vaikka ortografinen sanasto onkin usein laajempi (Milton ym., 2010). Myös huonompi korrelaatio tuottavan puheen kanssa on yleinen löytö ortografisessa muodossa esitettyjen sanastotestien saralla (esim. Milton ym., 2010; Uchihara & Clenton, 2020), vaikka tässä tutkielmassa puhumisen korrelaatio ylsi lähes samalle tasolle muiden modaliteettien kanssa. Yleisesti ottaen puheen ja sanastotestien välisten yhteyksien muista modaliteeteista eroavaa luonnetta on selitetty muun muassa sillä,

että puheessa käytetään vain tietyn vaativuustason sanastoa (esimerkiksi 2000 tai 3000 yleisintä sanaa), kun taas sanastotesti usein mittaa kielitaitoa laajemmalla skaalalla (Uchihara & Clenton, 2020). Tässä tutkielmassa saadut tulokset siis myötäilevät aiempien tutkimusten tuloksia. Käytännössä tulokset tarkoittavat, että Lexize on melko pätevä ennustamaan suoriutumista kaikissa modaliteeteissa, mikä antaa tukea väitteelle siitä, että Lexizen kaltainen kyllä/ei-testi voisi toimia yleisen kielitaidon mittarina tilanteissa, jossa seuraukset eivät ole tutkittavalle raskauttavat. Erityisesti puhumisen taitojen ja sanaston koon välinen yhteys vaatii kuitenkin lisätarkastelua.

### 4.3 Lexize-reaktioaika ja YKI-tutkinto

Tämän tutkielman perusteella Lexize-testissä mitatut reaktioajat eivät ennusta suoriutumista YKI-tutkinrossa. Korrelaatioita reaktioajan ja YKI-tutkinnon välillä ei havaittu yhdelläkään tässä tutkielmassa tarkastellulla laskentatavalla (reaktioaika kaikista vastauksista, reaktioaika oikeista vastauksista, reaktioaika vääristä vastauksista). Reaktioajan yhdistäminen Lexize-pisteisiin ei myöskään parantanut pisteiden kykyä ennustaa YKI-tuloksia. Näin ollen tutkielman tulos on, ettei reaktioajan tarkastelu tuo lisäarvoa testille.

Reaktioaikaa kielitaidon mittarina on tutkittu varsin vähän, mutta tämän tutkielman tulos on osittain ristiriidassa näiden aiempien tulosten kanssa. Johdannossa esitelty Harringtonin (2006) tutkimus esittää reaktioaikojen olevan merkitsevä sanasto-osaamisen ennustaja kyllä/ei-testeissä, ja myös Pellicer-Sánchezin ja Schmittin (2012) tutkimus antaa tukea reaktioaikojen käytölle ainakin edistyneimpien oppijoiden kohdalla. Eroa Pellicer-Sánchezin ja Schmittin (2012) tulokseen saattaa selittää osin se, että heidän tutkimuksessaan oli tarkasteltu pelkkiä ”kyllä”-vastauksien reaktioaikoja, kun tässä tutkielmassa tarkasteltiin niin ”kyllä” kuin ”ei”- vastauksia. Pellicer-Sánchezin ja Schmittin perustelevat valintaansa sillä, että nimenomaan ”kyllä”-vastaukset ovat alttiita yliarvioinnille (ihminen raportoi tietävänsä sanan vaikkei tiedä). Pellicer-Sánchezin ja Schmittin tutkimuksen lähtökohta oli tarkastella reaktioajan kykyä kontrolloida vastaajien yliarviointitaijumusta, joten juuri ”kyllä”-vastauksien kohdalla reaktioaikojen tarkastelu oli heidän tutkimuksessaan tarkoituksenmukaista.

Sanaston koon ja vastausnopeuden yhteyttä on tutkittu myös muutamassa muussa kielitaitoa mittaavassa testissä. Esimerkiksi Tanaben (2016) tutkimuksessa monivalintatestin reaktioaikoja mitattaessa edistyneempien oppijoiden reaktioajat olivat merkittävästi nopeampia kuin matalamman tason oppijoiden, ja lisäksi yleisempiin sanoihin vastattiin

nopeammin kuin harvinaisiin. Issa ja tutkimusryhmä (2022) taas tutkivat reaktioaikaa kuvien nimeämistehtävässä, ja päätyivät Tanaben tavoin toteamaan edistyneempien L2-puhujien olevan nopeampia. Myös Laufer ja Nation (2001) löysivät Vocabulary Levels Testin reaktioaikoja tarkasteltaessa merkitsevän yhteyden vastaustarkkuuden ja reaktioajan välille. Tämä yhteys vahvistui, kun oppijan taitotaso kasvoi – alemmilla taitotasolla tällä ei ollut niin paljon painoarvoa. Tämä yhteys oli nähtävissä 3000 yleisimmän sanan osaamistasosta ylöspäin, joskin jo 2000 sanan kohdalla oli havaittavissa joitain viitteitä ilmiöstä.

Oppijan taitotaso saattaakin olla yksi niistä selittävästä tekijöistä, miksi yhteyttä ei tässä tutkielmassa havaittu. Esimerkiksi Miralpeixin ja Mearan (2013) tutkimus osoitti tätä tutkielmaa vastaavasti, ettei sanaston koon ja kyllä/ei-testin vastausnopeuden välillä ollut yhteyttä. Kirjoittajat kuitenkin huomauttavat, että korkeammalla taitotasolla tilanne olisi voinut mahdollisesti olla erilainen. Tämän tutkielman tutkittavista suurin osa oli keskitasoisia (YKI-tutkinnon arvosanat B1 ja B2), ja korkeatasoisia oppijoita oli vain muutama. Saattaa siis olla, ettei osaaminen ollut tarpeeksi edistynyt niin, että reaktioaikojen tuottama efekti olisi voitu havaita. Edellä mainituissa tutkimuksissa on lisäksi pohdittu yksinomaan sanastoa, kun taas tässä tutkielmassa vertailukohtana toimiva YKI-tutkinto on sanastoa laajempi kokonaisuus. Myös tällä voi olla vaikutusta reaktioajan yhteydessä suoriutumiseen. Kuitenkin esimerkiksi Sahiruddin (2019) hyödynsi kyllä/ei-testin reaktioaikoja L2-oppijoiden lukemisen taitojen mittarina ja totesi reaktioajan kykenevän ennustamaan lukemissujuvuutta. Toisaalta myös hänen tutkimuksessaan reaktioaika selitti vain 2 % lukemissuoriutumisen varianssista. Liun tutkimuksessa (2020) sanojen kääntämistehtävässä mitattu suoritusaika taas korreloi puheen sujuvuuden kanssa. Syntaksiosaamisen kanssa tämä aika ei korreloinut. Koska YKI-tutkinnossa puhetta arvioidaan kokonaisuutena, leksikaalisen haun nopeuden olisi voinut ajatella olevan yhteydessä suoriutumiseen puheen osakokeessa, mutta Lexizen reaktioajat eivät nostaneet esiin tällaisia yhteyksiä. Tärkeää on huomioida myös se, että reaktioaika on vastaustarkkuutta häiriöherkempi mittari ja tässä tutkielmassa käytetyssä asetelmassa tutkittavat suorittivat kokeen vaihtelevissa ympäristöissä, jossa häiriöitä ei voitu kontrolloida.

Myös tutkielmassa hyödynnetty yhdistelmämuuttujiin pohjautuvaan tilastolliseen vertailuun on suhtauduttava kriittisesti. Reaktioaika ja vastaustarkkuus on yhdistetty myös aiemmassa tutkimuksessa (Harrington & Carey, 2009), mutta menetelmä on ollut tässä tutkielmassa käytettyä menetelmää monimutkaisempi. Aiempi tutkimus tutkielmassa hyödynnetyn menetelmän soveltamisesta vastaavissa konteksteissa on kirjoittajan tiedon mukaan vähäistä. Chignella ja tutkimusryhmä (2015) tarkastelivat tutkimuksessaan case-tapausta, jossa

hyödynsivät nopeuden ja tarkkuuden Z-summamuuttujia vertaillessaan suoriutumista Stroop-tehtävässä ja myyrälyöntipelissä. Tästä tutkimuksesta tämä tutkielma kuitenkin poikkeaa, sillä tutkielman toteutuksessa ei tarkkaan harkittu nopeuden ja tarkkuuden erilaisia painotuksia summamuuttujaa laskettaessa esimerkiksi iän perusteella. Reaktioajan ja vastaustarkkuuden välillä ei tutkielmassa myöskään löydetty korrelaatioita, mikä tekee Z-pisteiden käyttämisestä ennestään kyseenalaisen menetelmän. Ohella pelkkä korrelaatioiden vertailu ei ole tilastolliselta voimaltaan riittävä osoittamaan eroja summamuuttujien ja yksittäisten muuttujien välillä. Reaktioajan ja tarkkuuden summamuuttujien perusteella tehtyihin tilastollisiin päätelmiin on siis tutkielman osalta syytä suhtautua varauksella, ja niitä on pidettävä korkeintaan suuntaa antavina.

Kaiken kaikkiaan vaikuttaisi siltä, että kielitaidon modaaliteettien ja reaktioajan väliseen yhteyteen liittyvät kysymykset jäävät vielä avoimeksi. Tämän tutkielman tulosten varjolla testin ajallistamista voi pohtia lähinnä huijaamisen kontrolloimiseksi, sillä Lexizen voi toteuttaa myös kotiympäristössä, jossa testaajan on vaikea muuten valvoa esimerkiksi hakukoneen käyttöä.

#### **4.4 Tutkielman rajoitteet**

Tämän tutkielman tuloksia on syytä tarkastella muutaman rajoittavan seikan valossa. Kaikki tutkielman tulokset liittyvät itseraportoitujen YKI-tutkinnon arvosanojen pohjalta tehtyihin laskelmiin ja päätelmiin. Tähän seikkaan liittyy kaksi heikkoutta: itseraportoidut arvosanat voidaan muistaa tai raportoida väärin, ja toisaalta arvosanoja kertyy näin vähemmän kuin silloin, jos ne olisi saatu suoraan Yleisiä kielitutkintoja järjestäviltä tahoilta, mikä ei kuitenkaan yrityksistä huolimatta onnistunut. YKI-tutkinnon arvosanat ovat aineistona myös siinä mielessä ongelmallisia, että keskitason tutkintoja suoritetaan huomattavan paljon enemmän kuin korkean tason tutkintoja. Tämä näkyi myös tämän tutkielman aineistossa, ja johti siihen, että aineisto oli jakaumaltaan huipukas. Lisäksi aineiston koko oli suhteellisen pieni siihen nähden, että tutkielman tarkoitus oli validoida testiä. Testin validoimiseksi suurempi tutkittavien joukko olisi suositeltava tilastollisen voiman lisäämiseksi. Esimerkiksi Lexizen esimerkkinä toiminut LexTale on validoitu 155:llä tutkittavalla (Lemhöfer & Broersma, 2011), ja taas mahdollisesti maailman käytetyimmän sanastotestin Vocabulary Levels Testin uusioversiot peräti 1463:lla tutkittavalla (Webb ym., 2017). Tässä aineistossa suuri osa kerätyistä vastauksista (90,2 %) jäi tutkielman ulkopuolelle. Pääasiallinen syy tähän oli, ettei suurin osa tutkimukseen osallistuneista ollut suorittanut YKI-tutkintoa, sillä tämä ei

ollut vaatimuksena tutkittavia rekrytoitaessa. Suoraan YKI-tutkintolautakunnalta tuloksia ei myöskään tavoittelusta huolimatta saatu. Tutkielman tilastollisessa osuudessa vedottiin pienehköstä tutkimusjoukosta huolimatta keskeiseen raja-arvolauseeseen, jonka on osoitettu pätevän myös huipukkaassa aineistossa (esim. Ahad ym., 2011). Tuloksia kuitenkin tarkasteltiin myös ei-parametrisen Spearmanin korrelaatiokertoimen valossa ja havaittiin, etteivät ei-parametrisen tarkastelun tulokset eroa Pearsonin korrelaatiokertoimen tuottamista tuloksista.

Tutkittavien taustat olivat vaihtelevia, ja niiden osalta aineistoa ei ollut mahdollista kontrolloida. On todettu, että kulttuuritausta voi vaikuttaa vastaustyyliin kyllä/ei-testissä (Huibregtse ym., 2002; Milton, 2007; Read, 2000; Lemhöfer & Broersma, 2012), esimerkiksi tietystä kulttuuritaustasta tulevat henkilöt voivat olla taipuvaisempia arvaamaan. Suomen kaltaisen pienen kielen kohdalla ilmiötä voi olla vaikea tutkia, sillä eri taustoista tulleiden suomen kielen oppijoiden osajoukot voivat olla varsin pieniä. Tämän tutkielman aineistoa kerätessä kasaantui kuitenkin paljon tämän tutkielman aineiston ulkopuolelle jäänyttä dataa, jota voitaisiin mahdollisesti hyödyntää tämän seikan tarkastelussa tulevissa tutkimuksissa. Joidenkin taustaominaisuuksien yhteyksistä Lexizeen raportoidaan Salmelan ja kumppaneiden tutkimuksessa (2021).

Tässä tutkielmassa YKI-tutkinnon arvosanoja verrattiin Lexizen pisteytyskaavan mukaisiin pisteisiin. Voidaan pohtia, onko pisteytyskaavan valinnalla ollut vaikutusta korrelaatioihin. Esimerkiksi Harsch ja Hartig (2016) kirjoittavat, että kyllä/ei-testiä tarkasteltaessa raakapisteet tuottavat paremman tuloksen laajoissa lukemisen ja kuullun ymmärtämisen testeissä suoriutumisen ennustamisessa kuin yksikään Beeckmansin (2001) tai Huibregsten ja kumppaneiden (2002) tarkastelemista ja vertailemista vaihtoehtoisista tavoista pisteyttää kyllä/ei-testejä. Lexizen pisteytyskaava on intuitiivisesti järkevä, mutta voidaan miettiä, onko se paras mahdollinen. Beeckmans ja tutkimusryhmä (2001) perustelevat artikkelissaan Lexizen kaavan kaltaisten yksinkertaisten kaavojen etuja ja heikkouksia. Monimutkaisemmat tilastomalleihin perustuvat kaavat ottavat teoreettisesti pätevämmin huomioon arvaamisen vaikutukset. Useissa tutkimuksissa (Lemhöfer ja Broersma, 2012; Nakata ym., 2020; Mochida & Harrington 2006) yksinkertaisempi kaava on kuitenkin tuonut parhaan korrelaation muiden mittareiden kanssa. Harsch ja Hartig (2016) taas tuovat artikkelissaan esille, että väärin menneiden vastausten määrää voidaan sellaisenaan pitää arvausfaktorina. Harsch ja Hartig (2016) havaitsivatkin, että arvausfaktorin pistemäärä kasvoi silloin, kun kyseessä oli

harvinaisempi sanasto, eli ihmiset todennäköisemmin arvasivat enemmän. Lisäksi korkeammalla kielitaitotasolla olleet osallistujat arvasivat vähemmän.

Tutkimuksen toteutukseen liittyen tutkimuksen ohjeistuksessa havaittiin pieni harhaanjohtava lause, jossa kerrottiin tutkittavalle, että hän voisi pitää taukoja tutkimuksen suorittamisen aikana, vaikka ajastetussa tutkimuksessa tämä ei kesken itse testin ollut mahdollista. Tauon saattoi pitää esimerkiksi taustatietolomakkeen täyttämisen ja testin suorittamisen välissä, mutta itse suorituksen aikana taukojen pitämisellä on voinut olla vaikutusta tallentuneisiin reaktioaikoihin. Tutkimuksessa sovelletussa aineistossa ennen maksimiaikojen pois karsimista ei kuitenkaan ollut huomattavaa määrää maksimiaikaa vaatineita vastauksia (127 kpl = 1.9 %). Näin ollen voidaan ajatella, etteivät tutkittavat ole pitäneet taukoja soritustensa keskellä harhaanjohtavasta ohjeistuksesta huolimatta. Myös suhteettoman nopeita vastauksia (alle 200 ms) oli varsin vähän (75 kpl = 1.1 %). Voidaan siis olettaa, että suurinta osaa vastauksista on pohdittu ennen vastaamista.

Eettisyyden suhteen huomiona voidaan esittää tutkimuslomakkeen ja ohjeistuksen ymmärrettävyys ja tutkimukseen osallistumisen yhteydessä esitetty mahdollisuus osallistua arvontaan. Arvontapalkkio oli kohtuullinen eikä sillä uskota olevan vaikutuksia tutkimusjoukon valikoitumiseen. Niin suostumuslomakkeessa kuin tutkimuksen ohjeistuksessa ja toteutuksessa pyrittiin selkeyteen ja yksiselitteisyyteen. Tästä huolimatta pieni osa tutkimukseen osallistuneista ilmaisi hämmennystä tutkimuksen vaiheita kohtaan. Tämä näkyi verkkoalustoille jätetyissä yksittäisissä kommentteissa, joissa koettiin tutkimuksen kulku hankalaksi hahmottaa. Osasyynä voidaan pitää mahdollista kielimuuria, sillä ymmärrettävästi ohjeistusta ei voitu antaa erikseen kaikkien osallistujien äidinkielellä, eikä suomenkielinen ohjeistus olisi ollut asianmukainen aloitteleville suomenoppijoille. Voidaan siis pohtia, olisiko lomaketta voinut jotenkin selkiyttää entisestään.

Sisäänottokriteereistä on mainittava, että ne henkilöt, joiden pistemäärä Lexize-testissä oli alle 0, rajattiin pois aineistosta. Toisin sanoen ne henkilöt, jotka suoriutuivat todella huonosti Lexize-testissä eivät päässeet mukaan aineistoon, olivatpa heidän YKI-tuloksensa kuinka korkeat tahansa. Tällä sisäänottokriteerillä pyrittiin rajaamaan pois testisuoritukset, joissa henkilö on täysin arvannut kaikki vastaukset, jättänyt testin pyörimään vastaamatta ollenkaan tai muuten keskeyttänyt testin tekemisen ennen loppua. Sisäänottokriteereistä jäi myös puuttumaan yksi olennainen seikka, sillä tämä poistettiin vahingossa taustatietolomakkeesta

lomaketta muokatessa – diagnosoituja kielellisiä vaikeuksia ei kontrolloitu. Ei siis voida olla täysin varmoja siitä, millaista joukkoa saadut tulokset edustavat.

Voidaan myös jälleen miettiä YKI-tutkinnon vertautuvuutta huomattavasti yksinkertaisemman Lexize-testin taustalla. Sadeghi (2013) lainaa Bakerin (1989) ajatuksia ja esittää, että testien rakenteiden tulisi olla vertailtavissa, jotta niiden perusteella voitaisiin tehdä testin validiteettia koskevia johtopäätöksiä. Yhtenäistä testeissä on se, että niiden kohdepopulaatio on sama (suomen kielen L2-puhujat) ja molemmat testit mittaavat kielitaitoa. Levitettäessä Lexizeä kuitenkin lienee parempi nimittää kielitaidon mittarin sijaan sanastotestiksi, joka voi välillisesti antaa tietoa myös kokonaiskielitaidosta.

#### **4.5 Yhteenveto ja jatkotutkimus**

Tämän tutkielman tulokset tukevat Lexize-testin validiteettia. Tutkielmassa havaittiin keskiuuret korrelaatiot YKI-tutkinnon ja Lexize-testin välillä, mitä voidaan pitää hyvin positiivisena löydöksenä, kun otetaan huomioon Lexize-testin yksinkertaisuus ja helppokäyttöisyys. Osaltaan tutkielman löydökset tukevat koko testiformaatin validiteettia. Toisaalta kuten aiemmissa kappaleissa on mainittu, tulee korrelaatioihin perustuvia tulkintoja vahvistaa yhä esimerkiksi tulevaisuudessa kerättävällä käyttäjäkokemuksella ja tutkia testin toimivuutta erilaisilla ryhmillä. Vaikuttaisi kuitenkin siltä, että Lexize on jo tällaisenaan melko valmis paikkaamaan suomalaisen kielipsykologisen tutkimuksen kentällä vallitsevaa aukkoa objektiivisesta ja nopeasta kielitaitoarviosta. Lisäksi Lexizeä voidaan soveltaa muissakin pikaista testausta vaativissa tilanteissa, joissa seuraukset tutkittavalle eivät ole liian painokkaat. Esimerkkejä konteksteissa, joissa ammattialaiset voisivat käyttää Lexizeä hyväkseen ovat maahanmuuttoon liittyvä tutkimus ja soveltuvilta osin myös kliininen puheterapiatyö. Myös reaktioajan osalta tutkielma antaa alustavia viitteitä siitä, että Lexize on toimiva sellaisenaan, eikä reaktioaikojen lisääminen testiin tuo sille lisäarvoa kuin korkeintaan testiolosuhteiden kontrolloijana. Tämän seikan harkitseminen jää tulevan käyttäjäkokemuksiin pohjautuvan tutkimuksen käsiin. Kokonaisuudessaan tämä tutkimus tuo mielenkiintoisen vastakkaisen näkökulman reaktioaikojen hyödyntämiseen sanaston testaamisessa ja osoittaa, että aihe on lisätutkimuksen arvoinen.



## Lähteet

- Ahad, N.A., Yaacob, C.R., Othman, A.R., Ling, N.S. & Hoon, T. S. (2011). Central Limit Theorem in a Skewed Leptokurtic Distribution. *Jurnal Sains dan Matematik*, 3(1), 64-71.
- Ahola, S. & Hirvelä, T. (2016). Mikä merkitys osallistujan taustatekijöillä on menestymiseen kielitestissä? *Kieli, koulutus ja yhteiskunta*, 7(4). Saatavilla: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-syyskuu-2016/mika-merkitys-osallistujan-taustatekijöillä-on-menestymiseen-kielitestissa>
- Ahola, S. & Leblay, T. (2014). Kaksi vuosikymmentä Yleisiä kielitutkintoja. *Kieli, koulutus ja yhteiskunta*, 5(6). Saatavilla osoitteessa: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-joulukuu-2014/kaksi-vuosikymmentä-yleisiä-kielitutkintoja>
- Ahola, S. (2016). Puhetta arvioinnista – Yleisten kielitutkintojen arvioijien käsityksiä arvioinnista. *Kielitaidon arviointitutkimus 2000-luvun Suomessa, AFinLA-e: soveltavan kielitieteen tutkimuksia*, 9, 89-109.
- Ahola, S., Hirvelä, T. & Neittaanmäki, R. (2016). Puheen ymmärtämisen tehtävien taitotasolle asettaminen Yleisissä kielitutkinnoissa. *Kielitaidon arviointitutkimus 2000-luvun Suomessa, AFinLA-e. Soveltavankielitieteen tutkimuksia*, 9, 68-88.
- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading/Language Research*, 2, 231–256.
- Baker, D. (1989). *Language testing: A critical and practical guide*. London: Edward Arnold.
- Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M. & van de Velde, H. (2001). Examining the Yes/No vocabulary test: some methodological issues in theory and practice. *Language Testing*, 18(3), 235-274. DOI: 10.1191/026553201680188960
- Brown, J.D., & Rodgers, T. (2002). *Doing second language research*. Oxford: Oxford University Press.
- Brysbart, M. (2013). Lextale FR A Fast, Free, and Efficient Test to Measure Language Proficiency in French. *Psychologica Belgica*, 53(1), 23–37. DOI: <http://doi.org/10.5334/pb-53-1-23>
- Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research*, 6(2), 145–173.
- Chang, I.L. & Chang, C.B. (3-5. marraskuuta, 2017). *LEXTALE\_CH: A quick, character-based proficiency test for Mandarin Chinese*. Proceedings of the Annual Boston University Conference on Language Development (BUCLD), 42, 114– 130.
- Chignella, M., Tonga, T., Mizobuchib, S., Delangea, T., Hoa, W. & Walmsley, W. ( 2015). Combining Multiple Measures into a Single Figure of Merit. *Procedia Computer Science*. 69, 36-43. DOI: 10.1016/j.procs.2015.10.004
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2. painos.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Coltheart, M., & Rastle, K. (1994). Serial Processing in Reading Aloud: Evidence for Dual-Route Models of Reading. *Journal of Experimental Psychology. Human Perception and Performance*, 20(6), 1197–1211. <https://doi.org/10.1037/0096-1523.20.6.1197>
- Crossley, S. A., Salsbury, T. & McNamara, D. (2015). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied linguistics*, 36(5), 570-590. DOI:10.1093/applin/amt056
- Dabbagh, A. (2016). The Predictive Role of Vocabulary Knowledge in Listening Comprehension: Depth or Breadth? *International Journal of English Language & Translation Studies*, 4(3), 1-13.
- De Jong, N.H, Steinel, M.P, Florjin, A.F, Schoonen, R. & Hulstjin, J.H. (2012). Facets of speaking proficiency. *Studies in second language acquisition*, 34(1), 5-34. DOI: <https://doi.org/10.1017/S0272263111000489>
- D'Este, C. (2012). New views of validity in language testing. *EL.LE: Educazione Linguistica. Language Education*, 1(1), 61-76. DOI: <http://doi.org/10.14277/2280-6792/5p>
- Eyckmans, J., Van de Velde, H., van Hout, R., & Boers, F. (2007). Learners' response behaviour in Yes/No Vocabulary Tests. Teoksessa H. Daller, M. Milton & J. Treffers-Daller (Toim.), *Modelling and assessing vocabulary knowledge* (s. 59–76). Cambridge: Cambridge University Press.
- Fairclough, M. (2011). Testing the lexical recognition task with Spanish/English bilinguals in the United States. *Language Testing*, 28(2), 273-297. DOI: <https://doi.org/10.1177/0265532210393151>
- Feng, T. (2014). Assessing the Depth and Breadth of Vocabulary Knowledge with Listening Comprehension. *Journal of Language Teaching and Learning in Thailand*, 48, 29-56.
- Ferré P. & Brysbart M. (2017). Can Lextale-Esp discriminate between groups of highly proficient Catalan-Spanish bilinguals with different language dominances? *Behavioral Research Methods*, 49(2), 717-723. DOI: 10.3758/s13428-016-0728-y
- Fulcher, G. & Davidson, F. (2009). Test architecture, test retrofit. *Language Testing*, 26(1), 123–144. DOI: <https://doi.org/10.1177/0265532208097339>

- Gungör, F., & Yayli, D. (2016). The interplay between text-based vocabulary size and reading comprehension of Turkish EFL learners. *Educational Sciences: Theory & Practice*, 16(4), 1171–1188. DOI: <https://doi.org/10.12738/estp.2016.4.0078>
- Han, D. (2017). *Evaluating the relationship between vocabulary knowledge and listening comprehension in English as a foreign language*. [Julkaisematon tohtorintutkimus]. Hong Kong Polytechnic University.
- Harrington, M. & Carey, M. (2009). The on-line Yes/No test as a placement tool. *System*, 37(4), 614-626. DOI: <https://doi.org/10.1016/j.system.2009.09.006>
- Harrington, M. (2006). The lexical decision task as a measure of L2 lexical proficiency. *EUROSLA Yearbook: Annual Conference of the European Second Language Association*, 6, 147-168.
- Harsch, C., & Hartig, J. (2016). Comparing C-tests and Yes/No vocabulary size tests as predictors of receptive language skills. *Language Testing*, 33(4), 555–575. <https://doi.org/10.1177/0265532215594642>
- Huibregtse I., Admiraal W., Meara P. (2002). Scores on a Yes–No vocabulary test: Correction for guessing and response style. *Language Testing*, 19, 227–245.
- Härmälä, M., Neittaanmäki, R. & Tarnanen, M. (2010). Aikuisten kielitaito ja Yleiset kielitutkinnot. *AFinLA-e: Soveltavan kielitieteen tutkimuksia*, 2, 59–71.
- IBM Corp. (2021). IBM SPSS Statistics Windowsille, Versio 28.0.0.0. Armonk, NY: IBM Corp.
- Ibrahim, E.H.E, Sarudin, I. & Muhamad, A.J. (2016). The Relationship between Vocabulary Size and Reading Comprehension of ESL Learners. *English Language Teaching*, 9(2), 116-123.
- Issa, S. H., Awadh, F. H. A., & Ahmed, H. R. (2022). The role of proficiency level in the speed of lexical activation. *Cogent Arts & Humanities*, 9(1). <https://doi.org/10.1080/23311983.2021.1999613>
- Izura, C., Cuetos, F. & Brysbaert, M. (2014) Lextale-Esp: A test to assess the Spanish vocabulary size rapidly and efficiently. *Psiologica*, 35(1), 49-66.
- Jeon, E. H. & Yamashita, J. (2014). L2 reading comprehension and its correlates: A Meta-analysis. *Language Learning*, 64, 160–212.
- Jordan, E. (2012). Cognates in vocabulary size testing – a distorting influence? *Language Testing in Asia*, 2(3), 5-17. doi:10.1186/2229-0443-2-3-5
- Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, 44(1), 287–304. <https://doi.org/10.3758/s13428-011-0118-4>
- Kwak, S.G & Kim, J.H. (2017). Central limit theorem: the cornerstone of modern statistics. *Korean J Anesthesiol*, 70(2), 144-156. DOI: 10.4097/kjae.2017.70.2.144.
- Laufer, B. & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, 1(1), 7-28. DOI: <https://doi.org/10.1075/eurosla.1.05lau>
- Laufer, B. & Ravenhorst-Kalovski, G.C. (2010). Lexical threshold revisited: Lexical test coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15-30.
- Leki, I. & Carson, J.G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. *TESOL Quarterly*, 28(1), 81-101. DOI: <https://doi.org/10.2307/3587199>
- Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44(2), 325-343. DOI: 10.3758/s13428-011-0146-0
- Li, M., & Kirby, J. R. (2015). The Effects of Vocabulary Breadth and Depth on English Reading. *Applied Linguistics*, 36(5), 611–634. DOI: <https://doi.org/10.1093/applin/amu007>
- Lin, L. H. F., & Morrison, B. (2010). The impact of the medium of instruction in Hong Kong secondary schools on tertiary students' vocabulary. *Journal of English for Academic Purposes*, 9, 255–266.
- Liu, Y. (2020). Relating lexical access and second language speaking performance. *Languages (Basel)*, 5(2), 1–15. DOI: <https://doi.org/10.3390/languages5020013>
- McLean, S., Kramer, B. & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language testing research*, 19(6), 741-760. DOI:10.1177/1362168814567889
- Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary tests. *Language testing*, 4(2), 142-154. DOI: <https://doi.org/10.1177/026553228700400202>
- Meara, P. (1990). Some notes on the Eurocentres vocabulary tests. Teoksessa J. Tammola (Toim.), *Foreign language comprehension and production*, pp. 103–113. Turku: AFinLa Yearbook.
- Mehrpour, S. & Rahimi, M. (2010). The impact of general and specific vocabulary knowledge on reading and listening comprehension: A case of Iranian EFL learners. *System*, 38(2), 292-300. DOI: <https://doi.org/10.1016/j.system.2010.01.004>
- Microsoft Corporation. (2022). Microsoft Excel. Haettu osoitteesta <https://office.microsoft.com/excel>.
- Milton J. (2007). Lexical profiles, learning styles and the construct validity of lexical size tests. Teoksessa Daller H., Milton J., Treffers-Daller J. (Toim). *Modelling and assessing vocabulary knowledge* (s. 47–58). Cambridge: Cambridge University Press.
- Milton, J., Wade, J. & Hopkins, N. (2010). Aural Word Recognition and Oral Competence in English as a Foreign Language. Teoksessa R. Chacón-Beltrán, C. Abello-Contesse & M. Torreblanca-López

- (Toim.), *Insights into Non-native Vocabulary Teaching and Learning* (pp. 83-98). Bristol, Blue Ridge Summit: Multilingual Matters. DOI: <https://doi-org.ezproxy.utu.fi/10.21832/9781847692900-007>
- Miralpeix, I. & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International review of applied linguistics in language teaching*, 56(1), 1-24. DOI: <https://doi.org/10.1515/iral-2017-0016>
- Miralpeix, I., & Meara, P. (2013). Knowledge of the written word. Teoksessa J. Milton & T. Fitzpatrick (Eds.), *Dimensions of vocabulary knowledge* (s.30–44). Bloomsbury Publishing Plc.
- Mochida, A. & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, 23(1), 73-98. DOI: 10.1191/0265532206lt321oa
- Nakata, T., Tamura, Y., & Scott, A. (2020). Examining the validity of the LexTALE test for Japanese college students. *Journal of Asia TEFL*, 17(2), 335-348. DOI: <http://dx.doi.org/10.18823/asiatefl.2020.17.2.2.335>
- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63(1), 59–82. DOI: <https://doi.org/10.3138/cmlr.63.1.59>
- Opetushallitus. (2011). Yleisten kielitutkintojen perusteet. Finlex.fi. [https://finlex.fi/data/normit/39607/Oph\\_24\\_011\\_2011\\_s](https://finlex.fi/data/normit/39607/Oph_24_011_2011_s)
- Pellicer-Sánchez, A. & Schmitt, N. (2012). Scoring Yes-No vocabulary tests: Reaction time vs. nonword approaches. *Language testing*, 29(4), 489-509.
- Perea, M., Rosa, E. & Gómez, C. (2002). Is the go/no-go lexical decision task an alternative to the yes/no lexical decision task? *Memory & Cognition*, 30, 34-45. DOI: <https://doi.org/10.3758/BF0319526344>
- Qian, D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Testing*, 52(3), 513-536.
- Qian, D.D. & Lin, H.F.L. (2020). The Relationship Between Vocabulary Knowledge and Language Proficiency. Teoksessa Webb, S. (toim.), *The Routledge Handbook of Vocabulary Studies* (s.66-80). Routledge. DOI: <https://doi.org/10.4324/9780429291586>
- Read J. (1998). Measuring the Vocabulary Knowledge of Second Language Learners. *RELC Journal*, 19(2), 12-25. DOI:10.1177/003368828801900202
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511732942>
- Sadeghi, K. (2013). Doubts on the validity of correlation as a validation tool in second language testing research: the case of cloze testing. *Language Testing in Asia*, 3(1), 15. DOI: <https://doi.org/10.1186/2229-0443-3-15>
- Sahiruddin, S. (23-24. maaliskuuta 2019). *Lexical accuracy and reaction time in second language reading proficiency in Indonesia*. First International Conference on Advances in Education, Humanities, and Language, ICEL 2019, Malang, Indonesia. DOI: 10.4108/eai.23-3-2019.2284951
- Salmela, R., Lehtonen, M., Garusi, S. & Bertram, R. (2021). Lexize: A test to quickly assess vocabulary knowledge in Finnish. *Scandinavian Journal of Psychology*, 62(6), 806-819. DOI: 10.1111/sjop.12768
- Schmitt, N., Nation, P. & Kremmel, B. (2019). Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation. *Language Teaching*, 53(1), 1-12. doi:10.1017/S0261444819000326
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-STS330>
- Stæhr, L. S. (2008). Vocabulary Size and the Skills of Listening, Reading and Writing. *Language Learning Journal*, 36, 139- 152. DOI: <http://dx.doi.org/10.1080/09571730802389975>
- Stæhr, L.S. (2009). Vocabulary knowledge and advanced listening comprehension in English and a foreign language. *Studies in Second Language Acquisition*, 31(4), 577-607. DOI: 10.1017/S0272263109990039
- Tanabe, M. (2016). Measuring second language vocabulary knowledge using a temporal method. *Reading in a Foreign Language*, 28(1), 118-142. Saatavilla: <https://www.proquest.com/scholarly-journals/measuring-second-language-vocabulary-knowledge/docview/1798353336/se-2>
- Tani, H. (2008). Materiaalia aikuisten maahanmuuttajien suomen kielen taidon kartoitukseen ja kehityksen seurantaan. Kielen, Finnish National Agency for Education. Haettu 18 marraskuuta 2022 osoitteesta <https://www.oph.fi/fi/tilastot-ja-julkaisut/julkaisut/kielo>
- Uchihara, T., & Clenton, J. (2020). Investigating the role of vocabulary size in second language speaking ability. *Language Teaching Research*, 24(4), 540–556. DOI: <https://doi.org/10.1177/1362168818799371>
- Van Zeeland, H. & Schmitt, N. (2013). Lexical Coverage in L1 and L2 Listening Comprehension: The Same or Different from Reading Comprehension? *Applied Linguistics*, 34(4), 457–479. DOI: <https://doi.org/10.1093/applin/ams074>
- Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, 22(2), 217–234. <https://doi.org/10.1017/S0142716401002041>

- Wang, Y. & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, 65, 139–150. DOI: <https://doi.org/10.1016/j.system.2016.12.013>
- Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL. Instituut Voor Toegepaste Linguistik*, 168(1), 33–69. DOI: <https://doi.org/10.1075/itl.168.1.02web>
- Yleisten kielitutkintojen asetus 2004/1163, 7 §. Annettu Helsingissä 16.12.2004. Saatavilla sähköisesti osoitteessa <https://www.finlex.fi/fi/laki/ajantasa/2004/20041163>.
- Zhang, S. & Zhang, X. (2020). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 26(4), 696–725. DOI: <https://doi.org/10.1177/1362168820913998>
- Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using single-detection measures, and some correlated of word and nonword recognition. *Intelligence*, 1, 5–31.

# Liitteet

## Liite 1

Rekrytointi-ilmoitus

Hello!

Would you be interested in 1) completing a quick and free vocabulary test to estimate your Finnish language skills, 2) help researchers develop better tools for Finnish language skill assessment and 3) get a chance to win a 40 euro gift card? If yes, participate in our study!

It's easy and fast, and you can do it all at home on your laptop or smartphone: just click on the link below, fill out the short background questionnaire and complete the test! All in all it should take only around 15 minutes of your time, and at the end you will get information on how well you know Finnish vocabulary! We would be really very grateful for your participation.

<https://psyk.abo.fi/LexizeWeb/##/>

You can participate if you are over 18 years old and are learning Finnish. You can be a beginner or an advanced learner or anything in in-between. :)

Best,

Research team at the University of Turku & Åbo Akademi

## Liite 2

Lista rekrytoinnissa hyödynnetyistä verkkosivuista

|   |
|---|
| r/Learn Finnish – Discord-palvelin  |
| Opi suomea –Facebook-sivu   |
| Let’s Study Finnish –Facebook-sivu  |
| Finland International –Facebook-sivu  |
| Ask a Finnish Teacher –Facebook-sivu  |
| Learn Finnish –Facebook-sivu  |
| learn finnish –Facebook-sivu (eri kuin edeltävä)                                |
| Finnish for Foreigners –Facebook-sivu   |
| Duolingo Finnish learners –Facebook-sivu  |
| British people living in Finland and anyone who wants to join us –Facebook-sivu |
| Finnish on Duolingo! -Facebook-sivu   |
| Suomenopettajien Facebook-ryhmä   |
| Turun suomenopettajien Facebook-ryhmä   |
| Random Finnish Lesson – blogin kommenttikenttä                                  |

\* On mahdollista, että linkkiä on jaettu myös muualla, sillä se on ollut vapaasti kaikkien osallistujien jaettavissa.

## Liite 3

## Osallistujien äidinkielet

|            |                 |
|------------|-----------------|
| Bosnia     | 1 osallistuja   |
| Kiina      | 5 osallistujaa  |
| Englanti   | 18 osallistujaa |
| Viro       | 2 osallistujaa  |
| Filippiino | 3 osallistujaa  |
| Ranska     | 1 osallistuja   |
| Galicia    | 1 osallistuja   |
| Saksa      | 4 osallistujaa  |
| Kreikka    | 1 osallistuja   |
| Unkari     | 1 osallistuja   |
| Italia     | 1 osallistuja   |
| Khmer      | 1 osallistuja   |
| Latvia     | 1 osallistuja   |
| Persia     | 1 osallistuja   |
| Puola      | 2 osallistujaa  |
| Portugali  | 2 osallistujaa  |
| Venäjä     | 6 osallistujaa  |
| Serbia     | 1 osallistuja   |
| Slovakki   | 2 osallistujaa  |
| Espanja    | 4 osallistujaa  |
| Vietnam    | 6 osallistujaa  |
| Muu        | 1 osallistujaa  |

#### Liite 4

##### Suostumuslomakkeen teksti

Hello, Welcome to Lexize! This test aims to measure your vocabulary knowledge in Finnish, and it includes a vocabulary test and a questionnaire. The test is part of a research project on language development of the departments of Psychology and Speech-Language Pathology of Åbo Akademi University and University of Turku. To gain insight into how different background factors affect vocabulary size, we ask you to answer a few questions beforehand. After completing the test, you will be informed about your performance. You can send the test results to the researchers by pressing the Send button. The data is processed anonymously.

##### POTENTIAL HARMS, RISKS, DISCOMFORTS

The risks involved in participating in this study are minimal. Your eyes may become tired during the course of the experiment. If this becomes uncomfortable, you are encouraged to take short breaks as needed.

##### CONFIDENTIALITY

All personal information we attain will remain highly confidential and strictly secured. Within this study, your data will be de-identified, i.e., you will be assigned a numeric participant ID and identified only through this ID. No publication or discussion regarding this study will contain any personal and/or identifying information. Your data may be shared with other researchers outside of University of Turku and/or with general public, as part of a journal publication: in this case, it will not contain any information that can be used to identify you. As requested by general federal and university agencies and policies, we will retain your data for at least 5 years. Once out of active use, the electronic data will be archived on a secure server in our research lab for a maximum of 10 years.

##### PARTICIPANT RIGHTS

Participation in this study is completely voluntary. You may choose to withdraw at any time during the study without any consequences. If you withdraw during the study, you will not need to give any explanation for the decision. It is also possible to not click the Send Result-button after you have completed the experiment. In both cases, your data will not be stored. In case you have sent the data, you will be able to withdraw your data up to 6 months after the experiment is completed. If you intend to do so, please contact Raymond Bertram at the Department of Psychology and Speech-Language Pathology at the University of Turku ([firstname.lastname@utu.fi](mailto:firstname.lastname@utu.fi))

##### CONSENT

I have read and understood the information letter about this vocabulary study. I understand that I may withdraw from the study at any time without consequences. I agree that the data collected within the context of this study may be publicly shared as long as no person-identifying information will be made public. By clicking OK I consent to the above statements.

##### LOTTERY

Among the participants who send us their results, there will be a lottery, that gives a chance to win a GoGift super gift card of 40€. The lottery will take place 3.5.2022 and the 10 prize winners will be informed via email within the first 2 weeks of May.



## Liite 5

Taustatietolomakkeen kysymykset

Kindly answer the questions below.

What is your email address? (Note that this information is collected solely for identifying participants across devices and to send you a Gift Card if you win in the lottery. We will not send you spam or distribute your email to third parties.)

What is your gender?

*Male/Female/Other*

What is your native language?

*Pudotusvalikko kielistä aakkosjärjestyksessä - mukana myös vaihtoehto muu*

How would you rate your native language skills (if your native language is other than Finnish)? The rating scale is from 1 (very weak) to 10 (outstanding).

At which age did you start learning Finnish?

How would you rate your Finnish skills? The rating scale is from 1 (very weak) to 10 (outstanding).

Education level (please report the highest level completed)

*Grade school/Vocational school/ Higher secondary/ University of applied sciences/ University (Bachelor or Master) /Doctoral degree*

How old are you?

How long have you lived in Finland (in years)?

Have you done the general Finnish language proficiency test for foreigners (YKI)?

*Yes/No*

What was the result for listening comprehension?

*A1/A2/B1/B2/C1/C2*

What was the result for reading comprehension?

*A1/A2/B1/B2/C1/C2*

What was the result for speaking?

*A1/A2/B1/B2/C1/C2*

What was the result for writing?

*A1/A2/B1/B2/C1/C2*

In which year did you take the YKI-test?

*Pudotusvalikko vuosista*

Are you using a touchscreen device?

*Yes/No*