

*Mutta, M., P. Lintunen, I. Ivaska & P. Peltonen (toim.) 2014. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2014 / n:o 7. 101–117.*

**Tarja Leblay**

Turun yliopisto

## **Uudenlaisen kuunneltaviin puhenäytteisiin perustuvan itsearviointimenetelmän toimivuus suhteessa perinteiseen itsearviointiin**

This empirical study focuses on French oral language skills and self-assessment. The main purpose of the study was to test the functionality of a new type of speech sample-based self-assessment. The study was designed to explore how well the learners (n=209) could use different methods of self-assessment (traditional paper and pencil self-assessment and a new speech sample-based self-assessment). The data were analyzed by using both classical and IRT-based item analyses. In addition, the study included a qualitative part which provided additional information about the students' reflections and feelings about their own actions in the self-assessment process. In the new method of self-assessment the learners compared their performances on speaking tasks with speech samples collected from other learners. The results of the study showed that the self-assessment based on listening to speech samples and comparing one's own performance with them functioned very well.

**Keywords:** self-assessment, self-assessment based on speech samples, speaking, reflectivity

## 1 Johdanto

1990-luvulla yleistynyt kielitaidon itsearviointi sijoittuu niin kutsuttujen vaihtoehtoisten arviointitapojen joukkoon. Itsearvioinnille on tyypillistä **oppijakeskeisyys**, ja se liittyy tärkeänä osana elinikäiseen oppimiseen. Oppijakeskeisessä opetusnäkemyksessä opettaja luovuttaa osan tavanomaiseen opetustilanteeseen liittyvästä vallasta oppijan käsiin. Itsearvioinnin tarkoitus ei kuitenkaan ole opettajan työn turhaksi tekeminen, vaan ihannetilanteessa oppimistilanne rakentuu itsearviointia käytettäessä oppijan ja opettajan välisen jatkuvan dialogin pohjalta.

Itsearvioinnin käyttöönottoon sisältyy ennakko-odotus, jonka mukaan oppija on halukas opiskelemaan tavalla, joka vaatii aktiivista sitoutumista, uskoa omiin kykyihin ja halua muuttaa tuttuja opiskelutottumuksia. Itsearviointiprosessi etenee parhaimmillaan esittely- ja harjoittelujakson kautta siten, että oppija vähitellen kasvaa ottamaan vastuun omasta oppimisestaan (**oppimaan oppiminen**). Prosessin lopullisena päämääränä on se, että oppijasta tulee autonominen ja reflektiivinen oman oppimisensa suunnittelija, toteuttaja ja tarkkailija.

Itsearviointitaito mainitaan yhtenä oppimistavoitteena suomalaisen peruskoulun ja lukion opetussuunnitelmien perusteissa (POPS 2004; LOPS 2003). Suullisen kielitaidon opettaminen ja arviointi sekä riittävän hyvän suullisen kielitaidon hallinta ovat yhä edelleen erityisen ajankohtaisia asioita Suomessa, kun kansalliseen ylioppilastutkintoon on lähivuosina suunnitteilla suullista kielitaitoa testaava koe. Itsearviointitaitojen tärkeyttä korostetaan myös tämän hetken ja tulevaisuuden työelämän muutoksiin liittyväenä elinkaaren osana (EK 2011: 32). Itseohjautuvuutta tarvitaan kielitaidon itsearvioinnin lisäksi yleisemminkin tulevaisuuden työelämässä, jossa ennusteiden mukaan vaaditaan entistä luovempaa otetta työhön. Työtehtävissä tarvittavia metodeja ja toimintamalleja ei voi välttämättä aina saada ikään kuin valmiina, vaan tehtävien ”nuotittomuus” edellyttää tekijältään halua oppia uutta, etsiä vaihtoehtoisia toimintatapoja ja kykyä tarkastella tietoa eri näkökulmista (EK 2011: 5–8).

## 2 Aiemmasta kielitaidon itsearviointia koskevasta tutkimuksesta

Itsearvioinnin käyttämistä puoltavat monet tutkimukset, joiden tulosten mukaan itsearvioinnilla on kielen oppijan motivaatiota kasvattava vaikutus (ks. esim. Blanche & Merino, 1989; von Elek 1985). Suurin osa tutkimuksista tähdentää myös, että oppijalle on hyötyä metakognitiivisten taitojen kehittymisestä sekä opintojen aikana että luultavasti myös myöhemmin elämässä (esim. Oscarson 1989; EVK 2004; Atjonen 2007; Perrenoud

2011). Valtaosassa tutkimuksista on saatu todisteita itsearvioinnin ja ulkoisen arvioinnin välisestä selvästä, joskus vahvastakin yhteydestä (esim. Evers 1981; LeBlanc & Painchaud 1985; Blue 1988). Itsearvioinnin käytön ja sen tulosten luotettavuuden kannalta on kuitenkin syytä muistaa, että oppijoiden itsearviointitaidot näyttävät paranevan vain jatkuvan harjoituksen myötä (ks. esim. Mok, Lung, Cheng, Cheung & Ng 2006; Dragemark Oscarson 2009).

Itsearvioinnin luotettavuutta tarkastelleista tutkimuksista enemmistö osoittaa, että oppijat onnistuvat tietyin ehdoin itsearvioimaan osaamistaan validilla ja luotettavalla tavalla. Itsearvioinnin onnistumiseen voivat vaikuttaa muun muassa itsearviointiin liittyvä harjoitus ja kokemus (ks. esim. Tudor 1996; Oscarson 1997, 1980; Taras 2001), arvioitava taito (ks. esim. Rasch 1979; Ross 1998; Lee-Ellis 2009), oppimiskulttuuri, koulutus- ja kulttuuritausta (ks. esim. Holec 1990; Thomson 1996; Matsuno 2009) sekä oppijan persoonallisuuteen ja minäkuvaan liittyvät seikat (ks. esim. MacIntyre, Noels & Clément 1997; Ockey 2011; Sinharay, Powers, Feng, Saldivia, Giunta, Simpson & Weng 2009).

Oscarsonin (1998: 137–138) mukaan kielten oppimiseen liittyvää oppijakeskeistä tutkimusta on tehty erityisesti itseopiskelusta ja oppijan autonomiasta, mutta suoraan arviointiin liittyvä tämänkaltainen tutkimus on ollut vähäisempää. Tutkija toteaa, että tälle alalle tulisi kehittää uusia metodeja ja materiaaleja. Tämän tutkimuksen pyrkimyksenä onkin tuottaa tietoa konkreettisten arviointivälineiden käytöstä ja toimivuudesta suullisen kielitaidon itsearvioinnissa. Tutkimuksen lähtökohtana on aiempiin tutkimustuloksiin pohjautuen oletamus siitä, että itsearviointia voidaan pitää luotettavana arviointitapana erityisesti silloin, kun oppijat kokevat käyttämänsä itsearviointimenetelmän ymmärrettäväksi ja suhteellisen helpokäyttöiseksi.

### 3 Tutkimuksesta

Edellä olleen lyhyehkön johdanto-osuuden ja alan joitakin aiempia tutkimuksia esitelleen osuuden jälkeen tässä artikkelissa käsitellään tutkimuksen yleisiä reunaehtoja (luku 3), tutkimuksen käytännön toteutusta (luku 4), perinteisten tilannekohtaisten kirjallisten itsearviointiväittämien (luku 5) ja uudentyyppiseen itsearviointiin liittyvien kuunneltavien puhenäytteiden toimivuutta (luku 6). Artikkelin päätösluvussa käsitellään uuden itsearviointimenetelmän toimivuutta, luotettavuutta ja käytettävyyttä (luku 7). Luku 8 sisältää pohdintaa tutkimuksen rajoituksista ja mahdollisista jatkotutkimusaiheista.

Tämän artikkelin aineiston pohjana toimii soveltavan kielentutkimuksen alaan kuuluva väitöskirjatutkimus (Leblay 2013), jossa tarkasteltavana kielenä on ranska vierana kielenä. Tutkimuksen osallistujia ovat ranskan kielen aikuisopiskelijat (n=209), joista valtaosa on yliopisto-opiskelijoita, mutta tutkimukseen osallistui myös kaksi lukio-

luokkaa. Tutkimukseen sai osallistua kuka tahansa ranskan kieltä taitava, ja siksi osallistuneiden opiskelijoiden ryhmä on hyvin heterogeeninen. Joukossa on niin ranskan kielen pääaine- kuin sivuaineopiskelijoita sekä kielikeskusopiskelijoita, joiden opiskelemat aineet ovat muita kuin kieliaineita.

Edellä mainittu tutkimus on empiirinen tutkimus, jonka aineistoa tarkastellaan sekä kvantitatiivisesta että kvalitatiivisesta näkökulmasta. Kvantitatiivisesti aineistoa analysoitiin sekä perinteisen osioanalyysin että moderniin testiteoriaan perustuvan Rasch-analyysin avulla. Tutkimuksen kvalitatiivisten osien tarkoituksena on antaa lisäselvitystä niistä ilmiöistä, joita tutkimusaineiston käsittely nosti esille. Tutkimuksella on yhteyksiä Eurooppalaiseen viitekehukseen (EVK, arviointiasteikot), DIALANGiin (kirjalliset itsearviointiväittämät) sekä Yleisiin kielitutkintoihin (puhumisen tehtävät).

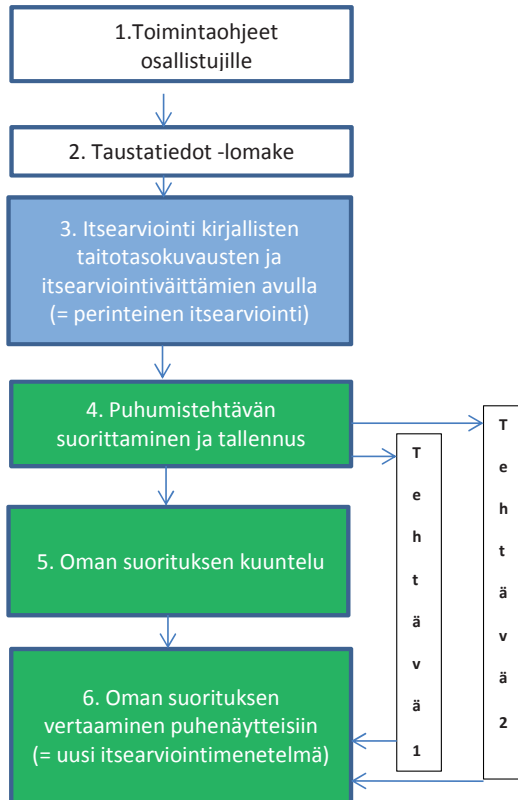
Tutkimuksessa käytetyt kaksi puhumisen tehtävää lainattiin Yleisten kielitutkin-tojen ranskan kielen testistä. Tehtävä 1 oli kuvasta kertomistehtävä, jonka tarkoituksena oli saada osallistujalta mahdollisimman pitkä vapaamuotoinen suullinen tuotos. Tehtävä 2 oli tilanteessa reagointitehtävä, jossa suomalainen ranskaa puhuva henkilö esitteli työpaikalleen saapuneen tärkeän ranskalaisen vieraan niin ikään ranskaa puhuvalle kollegalleen.

Tutkimuksessa kokeiltu uusi itsearviointimenetelmä perustuu siihen, että oppijat kuuntelevat nauhoitettuja puhenäytteitä edellä selostetuista puhumisen tehtävistä ja vertaavat omaa puhe-suoritustaan näytteisiin. Artikkelissa tarkastellaan myös perinteisen kirjallisiin itsearviointiväittämiin perustuvan itsearvioinnin toimivuutta suhteessa uuteen itsearviointimenetelmään, jossa oppivat kuuntelevat puhenäytteitä ja vertaavat omaa suullista ranskan kielen taitoaan niihin. Uuden menetelmän kokeilussa korostuvat menetelmän laadulliset ja toiminnalliset aspektit. Artikkelin taustana toimivan aineiston laadukkuus ja toimivuus selvitettiin väitöskirjatutkimuksessa. Tarkastelun kohteena väitöskirjassa olivat siten muun muassa eri itsearviointimenetelmien toimivuuden tarkastelun lisäksi monet muutkin seikat kuten oppijoiden käsitykset itsearvioinnista. Tässä artikkelissa, jonka aineisto siis pohjautuu edellä mainittuun väitöstutkimukseen, keskitytään selvittämään kahden eri itsearviointimenetelmän, perinteisen kirjallisiin itsearviointiväittämiin ja uuden kuunneltaviin puhenäytteisiin perustuvan menetelmän, toimivuuteen liittyviä seikkoja.

## 4 Tutkimuksen käytännön toteutus

Tutkimusaineisto kerättiin neljästä suomalaisesta yliopistosta sekä kahdesta lukioluokasta. Aineiston keruu tapahtui kielistudioistunnoissa, joiden kesto oli yhteensä noin kaksi tuntia. Kuvio 1 havainnollistaa tutkimusprosessin kielistudioistuntojen eri vaiheita.

Perinteisen itsearviointimenetelmän vaiheet on merkitty kuvioon sinisellä värillä, kun taas uuden menetelmän eri vaiheet on merkitty vihreällä.



KUVIO 1. Uuden itsearviointimenetelmän eri vaiheet.

Kuvion 1 kohta 3 liittyy kirjallisten taitotasokuvausten ja itsearviointiväittämien käyttöön. Uudenlaista puhenäytteisiin perustuvaa itsearviointia havainnollistavat kuvion kohdat 4–6. Taustatietolomakkeen täytön jälkeen oppijat valitsivat yleistaitotason omalle ranskan kielen suulliselle taidolleen kuudesta Eurooppalaisen viitekehysten taitotasosta (A1–C2). Tämän jälkeen oppijat vastasivat rastittamalla joko ”kyllä” tai ”ei” 42 tilannekohtaiseen suullisen kielitaidon itsearviointiväittämään. Vastauksen perusteena toimi se, vastasiko väittämän sisältö osallistujan käsitystä hänen omasta ranskan kielen suullisesta kielitaidostaan (kuvio 1, kohta 3). Uudenlainen itsearviointi alkoi, kun oppijat suorittivat puhumistehtävän 1, minkä jälkeen puhetuotos tallennettiin äänitteelle (kuvio 1, kohta 4). Välittömästi tallennuksen jälkeen oppijat kuuntelivat oman suorituksensa tehtävästä 1 (kuvio 1, kohta 5). Sitten siirryttiin kuuntelemaan muiden oppijoi-

den nauhoitettuja puhenäytteitä samasta tehtävästä (kuvio 1, kohta 6). Tarkoituksena oli verrata omaa suullista puhesuoritusta jokaiseen kuunneltavaan puhenäytteeseen (n=26, tehtävä 1). Kysymys, johon tällä kertaa vastattiin ”kyllä” tai ”ei”, oli ”Suoriuditko paremmin tehtävästä kuin nauhalta kuulemasi henkilö?” Tehtävää 1 koskeva prosessi toistettiin sen jälkeen tehtävässä 2, jossa kuunneltavia puhenäytteitä oli vähemmän kuin tehtävässä 1 (n=23).

## 5 Tilannekohtaisten kirjallisten itsearviointiväittämien toimivuudesta

Tässä luvussa esitellään perinteisen itsearvioinnin toimivuuteen liittyviä havaintoja ja tuloksia (kirjalliset itsearviointiväittämät, kuvio 1, kohta 3). Uuden puhenäytteiden kuuntelemiseen ja oman suorituksen vertaamiseen liittyvän itsearvioinnin toimivuutta käsitellään seuraavassa luvussa.

Vaikka tutkimuksen tarkoituksena oli kokeilla käytännössä uudenlaista kuunneltaviin puhenäytteisiin perustuvaa itsearviointia, tilannekohtaisten kirjallisten itsearviointiväittämien tarkastelu oli tärkeää muun muassa siksi, että voitiin saada kokonaiskuva perinteisten kirjallisten itsearviointiväittämien toimivuudesta ja laadusta sekä arvioida perinteisen ja uuden itsearviointimenetelmän keskinäistä suhdetta. Lisäksi kirjallisten väittämien pohjalta saatiin kerättyä käyttäjäpalautetta siitä, mikä oppijoiden mielestä toimi ja mikä ei. Tällä tavalla saatiin lisätietoa tekijöistä, jotka voivat vaikuttaa oppijan itsearvioinnin onnistumiseen. Käyttäjäpalautteen keräämisellä pyrittiin saamaan tietoa oppijoiden näkemyksistä itsearvioinnista. Jotkut tutkijat korostavat sitä, että itsearviointia voidaan pitää arviointimuotona, joka on – ainakin periaatteessa – ylivertainen sisältönsä kattavuuden kannalta verrattuna esimerkiksi johonkin kielitestiin (Huhta 2003: 8).

Erlaiset itsearviointimenetelmät täydentävät toisiaan ja useamman kuin yhden menetelmän käyttö lisää tutkimuksen luotettavuutta. Tästä syystä tutkimuksessa käytettiin tilastollisia tunnuslukuja sekä kirjallisia itsearviointiväittämiä että kuunneltavia puhenäytteitä analysoitaessa. Käytettyjä tunnuslukuja olivat perinteisen osioanalyysin tuottama osio–testi–korrelaatio sekä Rasch-analyysin tunnusluvut *Infit mean square* (vähemmässä määrin myös *Infit mean Zstd*). Näiden tilastollisten tunnuslukujen lisäksi kirjallisten itsearviointiväittämien ja puhenäytteiden toimivuutta tarkasteltiin osallistujien kommentteja hyödyntäen sekä tarkastellen oppijoiden näkemyksiä väittämien vaikeustasosta (Winstepin Rasch-analyysi, josta ilmenee, käyttäytyvätkö tietyille taitotasolle asetetut osiot odotuksenmukaisesti mittainstrumentin osana). Tässä artikkelissa osiolla tarkoitetaan joko yksittäistä kirjallista itsearviointiväittämää tai kuunneltavaa puhenäytettä.

Perinteisessä osioanalyysissä hyödynnettiin osio–testi-korrelaatioita, jotka saatiin sekä SPSS:n reliabiliteettianalyysistä että Winstepin osio–testi-korrelaatioanalyysistä. Jos osio–testi-korrelaation arvo on yli 0,30, osion voidaan todeta arvon suuruudesta riippuen toimineen joko kohtalaisesti tai hyvin. Kaikki arvon 0,30 alle jäävät arvot puolestaan osoittavat, että osion toimivuudessa on mahdollisesti ongelmia – se on ehkä epäselvästi muotoiltu, tai se mittaa jotakin muuta ominaisuutta kuin muut samassa mittarissa olevat osiot. Rasch-analyysin tunnuslukujen lähtökohtana puolestaan oli se, että väittämien Infit mean square -arvon tulee sijaita 0,70 ja 1,30 välillä ja Infit Zstd -arvon rajojen +2 ja –2 sisällä, jotta niitä voidaan pitää hyvin toimivina mittarin osioina (Linacre 2014). Infit mean square -arvon odotusarvo on 1,0, ja sen on tarkoitus osoittaa, missä määrin jokin osio sopii Rasch-analyysin malliin.

## 5.1 Esimerkkejä hyvin toimineista tilannekohtaisista itsearviointiväittämistä

Yleisesti ottaen voidaan todeta, että 42 tilannekohtaista itsearviointiväittämää sisältävä mittainstrumentti (tässä itsearviointilomake) toimi perinteisen osioanalyysin mukaan kokonaisuutena hyvin. Esimerkiksi taitotasoilta B1, B2 tai C1 ei löytynyt yhtään väittämää, joka olisi toiminut huonosti. Näiden taitotasojen väittämien korrelaatioarvot olivatkin pääosin korkeita.

Myös Rasch-analyysin mukaan tilannekohtaiset itsearviointiväittämät toimivat kokonaisuudessaan hyvin. 42 väittämästä vain neljässä Infit mean square tai seitsemässä Infit Zstd -arvo oli hyväksyttävien rajojen ulkopuolella. Taitotason C2 kaikki väittämät toimivat hyvin Rasch-analyysin pohjalta.

Seuraavaksi esitellään joitakin väittämiä, jotka toimivat hyvin sekä perinteisen että Rasch-analyysin näkökulmasta. Tarkastelun tarkoituksena on yrittää selvittää, mitä sellaisia elementtejä väittämät sisältävät, jotka voisivat selittää niiden hyvän toimivuuden. Tarkastelun kohteena ovat väittämät 42 ja 6.

### Väittämä 42

(taso B2; korrelaatiokerroin 0,68; Infit mean square 0,91; Infit Zstd -0,08; vastausjakauma 106–103)

Osaan kuvailla ja määritellä konkreettisia asioita, joille en juuri sillä hetkellä satu muistamaan sopivaa sanaa (esimerkiksi pysäköintisakko, luottokortti tai vakuutuskirja).

Kun tarkastellaan edellä olevaa sattumanvaraisesti valittua tutkimuksessa hyvin toiminutta väittämää, voidaan aluksi todeta, että väittämä toimi tilastollisesti hyvin sekä perinteisen osioanalyysin (0,68) että Rasch-analyysin näkökulmasta (Infit ms 0,91). Lisäksi voidaan huomata, että väittämä sisältää viittauksia jokapäiväiseen konkreettiseen

kontekstiin. Esimerkkiväittämä on myös suhteellisen lyhyt eikä se sisällä konjunktiolla toisiinsa liitettyjä virkkeitä. Väittämässä on lisäksi selvennetty sulkeissa esimerkein, mitä tarkoitetaan konkreettisilla asioilla. Tutkimukseen osallistujat pitivät väittämää 42 sen tasoisena väittämänä kuin mikä sen määritely taitotaso on (B2).

Myös seuraava C2-tasoinen väittämä, väittämä 6, toimi tilastollisissa analyyseissä hyvin. Kyseisen väittämän toimivuutta on kuitenkin syytä tarkastella kriittisesti, sillä tutkimukseen osallistuneiden oppijoiden kommenttien pohjalta voidaan epäillä sen sisältävän joitakin elementtejä, jotka tekevät siitä vaikeasti ymmärrettävän.

#### **Väittämä 6**

(taso C2; korrelaatiokerroin 0,51; Infit mean square 0,87; vastausjakauma 25–184)

Osaan palata takaisin vaikeisiin kohtiin ja muotoilla uudestaan sanottavani niin, että toiset tuskin huomaavat sitä.

Opiskelijoiden kommentteista käy selkeästi ilmi se, että he eivät ole osanneet tulkita väittämän sisältöä<sup>1</sup>:

*Vaikeisiin asioihin kielellisesti, vai muuten? Miksi palaisin asiaan? Miten niin toisten huomautta? (3).*

*Väittämä 6:ta en oikein ymmärtänyt...mihin vaikeisiin kohtiin? (2).*

*En ymmärtänyt tilannetta. (4).*

*Kohta 6 oli vaikeasti ymmärrettävä. Mitä tarkoittaa palata vaikeisiin kohtiin? (4).*

Väittämä 6 vaikuttaakin olevan väittämä, jonka tarkka merkitys ja sisältö voidaan ymmärtää monella tavalla. Nojautuen oppijoiden antamiin kommentteihin voidaan oikeutetusti kysyä, mitä tarkoitetaan väittämässä mainituilla vaikeilla kohdilla, joihin palataan. Koska väittämän muotoilu on selvästikin tulkinnanvarainen, jokainen tutkimukseen osallistuja on tulkinnut väittämän sisältöä omalla tavallaan. Erityisen huolestuttavaa väittämän ymmärrettävyyden ja toimivuuden kannalta on se, että edes kielitaitonsa taitotasoasteikon ylimmille tasoille (tasot B2 ja C1) arvioineet oppijat eivät ole välttämättä ymmärtäneet puheena olevan väittämän sisältöä. Tässä vaiheessa on syytä muistuttaa, että tutkimuksessa, josta tämän artikkelin aineisto on lainattu, käytettiin englannin kielestä suomennettuja itsearviointiväittämiä<sup>2</sup>. Onkin syytä muistaa, että jos väittämiä käy-

1 Numero oppijan kommenttien perässä tarkoittaa sitä yleistaitotasoa, jolle oppija oman ranskan kielen suullisen kielitaitonsa asetti.

2 Väittämät on lainattu DIALANG-hankkeen suullisen kielitaidon itsearviointiosuudesta.



tetään eri kielissä ja jos niitä käännetään eri kielten välillä, kielikohtaiseen ymmärrettävyyteen on syytä kiinnittää huomiota.

## 5.2 Esimerkkejä huonosti toimineista itsearviointiväittämistä

### Väittämä 36

(taso B2; korrelaatio 0,52; Infit mean square 1,33, Infit ZStd 2,60; vastausjakauma 61–148)

Osaan puhua syntyperäiselle kielenpuhujalle siten, etten ole tahtomattani huvittava, ärsyttävä tai vaadi heitä käyttäytymään toisin kuin toisen syntyperäisen kanssa.

Väittämän 36 korrelaatio on perinteisen osioanalyysin näkökulmasta katsoen kunnossa (0,52), mutta molemmat modernia testiteoriaa edustavat Rasch-arvot ovat kohonneita: Infit mean square 1,33 ja Infit Zstd 2,60. Sen sijaan väittämän vaikeustaso on oppijoiden mielestä ollut täysin kohdallaan eli taitotasolla B2.

Yhteensä 19 oppijaa kommentoi väittämää 36. Kommenteista ilmeni, että oppijat eivät olleet ymmärtäneet väittämän sisältöä eivätkä siten olleet lainkaan varmoja omasta vastausvalinnastaan. Kommenteista löytyy mielenkiintoisia tulkintoja ja asenteita ranskan kielen natiivipuhujien käyttäytymisestä sellaista ei-natiiviva henkilöä kohtaan, joka puhuu heidän äidinkieltään:

*Eivätkö kaikki "ei-ranskankieliset" ole ranskaa äidinkielenään puhuville "tahtomattaan huvittavia"? (4).*

*Kysymys 36 mietitytti myös, sillä mielestäni syntyperäiset puhujat suhtautuvat aina ulkomaiseen puhujaan eri tavalla kuin natiiviin, vaikka kieli olisikin moitteetonta. (4).*

Eräs tutkimukseen osallistuneista oppijoista lähestyi väittämää keskustelukumppanin käyttäytymiseen liittyvästä näkökulmasta:

*Mistä voin itse tietää, puhuuko joku minulle samoin vai eri tavalla kuin syntyperäiselle kielenpuhujalle? (4).*

B1-tasolle oman kielitaitonsa arvioinut oppija kommentoi väittämän 36 vaikeaselkoisuutta seuraavasti:

*Kohta 36 tuntui hankalalta, koska en oikein ymmärtänyt, mitä siinä ajetaan takaa. Erityisesti kohta "tai vaadi heitä käyttäytymään toisin" on vaikeatajuinen. Enhän voi "vaatia" ketään tekemään mitään. (3).*

Vastaanotettujen kommenttien pohjalta voidaan todeta, että tämäntyyppiset väittämät, joissa viitataan vastapuolen asenteeseen tai käyttäytymiseen, tuntuvat saavan paljon

palautetta erityisesti ylimmillä taitotasoilla olevilta henkilöiltä, joilla on jo kokemusta natiivien kanssa keskustelemisesta ja jotka pystyvät ehkä helpommin aistimaan ranskan kielen syntyperäisten puhujien kielelliseen käyttäytymiseen ja kommunikaation sujumiseen liittyviä asenteita.

Hieman yllättäviä saattavat olla toteamukset, joiden mukaan monet oppijat näyttävät pitävän väistämättömänä sitä, että kielen natiivipuhujat joutuisivat aina muuttamaan käyttäytymistään sellaisen ulkomaalaisen henkilön seurassa, joka puhuu heidän äidinkieltään. Nämä näkemykset kuvastavat uskomusta, jonka mukaan vieraan kielen oppijan suullisen kielitaidon taso ei voisi ylittää samalle tasolle kuin kyseisen kielen natiivipuhujan kielitaito on. Osa oppijoista siis vaikuttaa ajattelevan, että vieraan kielen oppija ei pystyisi koskaan kielellisesti olemaan täysin yhdenvertainen natiivipuhujan kanssa suullisessa kommunikaatiotilanteessa. Olipa syy Rasch-analyysin kohonneisiin arvoihin ja oppijoiden lukuisiin kommentteihin sitten mikä tahansa, on todettava, että väittämä 36 tekee oudohkoja oletuksia natiivin ja ei-natiivin kielenpuhujan välisestä kommunikaatiosta ja sen reunaehdoista.

Toisena esimerkkinä huonosti toimineesta itsearviointiväittämästä on väittämä 21, joka on päinvastainen esimerkki kuin edellä esitelty väittämä 36. Kyseisen väittämän korrelaatiokerroin on heikko, kun taas samanaikaisesti kaksi muuta Rasch-analyysin arvoa ovat hyviä.

#### **Väittämä 21**

(taso C2; korrelaatiokerroin 0,21; Infit mean square 0,93; Infit Zstd .0; vastausjakauma 3–206)

Puhun ranskan kieltä melkein yhtä hyvin kuin äidinkieltäni.

Edellisen väittämän kohdalla oppijat tuntuivat ajattelevan, että vierasta kieltä ei voisi hallita samalla tasolla kuin omaa äidinkieltä. Oppijoiden kommentteja tarkasteltaessa tulee mieleen myös, että erityisesti (vieraita) kieliä opiskelevat henkilöt saattavat tulla kriittiseksi omaa kielitaitoaan kohtaan yleisemmälläkin tasolla. Tämä huomio tuntuu joissakin tapauksissa koskevan myös oppijoiden äidinkielen hallintaa, josta esimerkkinä seuraavan oppijan kommentti:

*Oikeastaan olisin voinut vastata joko kyllä tai ei, koska loppujen lopuksi äidinkielikään ei ihan aina riitä, jotta selviytyisi kaikista tämän kaavakkeen esittämistä kielenkäyttötilanteista (esim. kohdassa 16: en tunne vakuusterminologiaa pystyäkseni puhumaan aiheesta asian-tuntevasti). (3).*

Edellä olevasta kommentissa voi niin ikään selvästi havaita kriittisen vivahteen, kun oppija viittaa itsearviointiväittämän 21 ja ehkäpä koko itsearviointilomakkeen sisältöön.

### 5.3 Yhteenvetoa tilannekohtaisten itsearviointiväittämien toimivuudesta

Tilannekohtaisten itsearviointiväittämien laadun ja toimivuuden tarkastelu paljasti monta mielenkiintoista seikkaa. Ensimmäinen havainto oli se, että perinteiseen osio-analyysiin ja moderniin testiteoriaan perustuvat osioanalyysit nostivat esille osittain eri asioita. Täten tutkimuksessa tuli käytännössä testattua ja todeksi havaittua se, että nämä kaksi analyysiä todellakin tuovat esille erilaisen näkökannan tutkittavaan ilmiöön (*different "ways of seeing"*; Gyagenda & Engelhard 2009: 242). Samantyyppisiä huomioita esittivät Lynch ja Namara (1998: 176–177) maahanmuuttajien suullista kielitaitoa koskevassa tutkimuksessaan, jonka aineiston analysoinnissa käytettiin kahta erilaista analyysimetodia (GENOVA- ja Facets-analyysejä). Facets tarkastelee aineistoa mikroskooppisen tarkasti paljastaen aineistosta mahdolliset pienetkin "viat". Tutkijat toteavat kuitenkin, että mainitut analyysit mahdollistavat tasapainoisen ja perusteellisen tulkinnan tarkasteltavasta aineistosta, kun niitä käytetään yhdessä.

Tutkimukseen osallistuneiden oppijoiden laadulliset kommentit väittämien toimivuudesta ja ymmärrettävyydestä toivat lisävalaistusta asian tarkasteluun silloin, kun väittämät osoittautuivat toimivuudeltaan ongelmallisiksi tilastollisessa mielessä. Toisaalta kommentteista oli suurta apua myös silloin, kun molemmat tilastoanalyysit näyttivät kaiken olevan kunnossa, mutta samaan aikaan tarkasteltavana oleva väittäjä keräsi kielteisiä kommentteja oppijoilta. Kuten edellä todettiin, eri menetelmät tarkastelevat väittämiä eri näkökulmista ja antavat siten toisiaan täydentävää hyödyllistä tietoa väittämien toimivuudesta sekä käytetyn itsearviointimittarin luotettavuudesta.

## 6 Kuunneltavien puhenäytteiden toimivuudesta

Uuden menetelmän kohdalla kaikille tutkimukseen osallistuneille oppijoille pakollisia osia olivat molempien puhumistehtävien suorittaminen, puhenäytteiden kuunteleminen samoista tehtävistä ja lopuksi oman suorituksen vertaaminen jokaiseen kuunneltuun näytteeseen (26 näytettä / tehtävä 1; 23 näytettä / tehtävä 2). Vastaaminen vertaamisen pohjalta tapahtui rastittamalla vastauslomakkeen "kyllä"- tai "ei"-vaihtoehdon mukaan, mikä oppijan mielestä oli hänen selviytymisensä tehtävästä suhteessa puhenäytteen puhujaan. Kysymys, johon vastattiin, oli seuraava: "Suoriuduitko paremmin tehtävästä kuin nauhalta kuulemasi henkilö?"

Pakollisen osan lisäksi oppijoilla oli mahdollisuus halutessaan antaa kahdenlaisia lisäkommentteja kuunnelluista puhenäytteistä. Oppijat saivat halutessaan kirjoittaa ylös kommenttinsa siitä, miksi he valitsivat "kyllä" tai "ei" -vaihtoehdon näytteen kohdalla. Lisäksi lomakkeessa oli vaikea päättää -sarake, johon oppijat saattoivat merkitä

rastin niiden näytteiden kohdalla, jotka aiheuttivat heille näytteiden kuuntelun aikana vaikeuksia päättää, suoriutuivatko he paremmin tehtävästä kuin puhenäytteen henkilö vai ei. Uudessa kuunneltaviin puhenäytteisiin perustuvassa itsearvioinnissa käytetyt puhenäytteet analysoitiin tilastollisesti samojen tunnuslukujen avulla kuin edellä perinteisessä itsearvioinnissa käytettyjen tilannekohtaisten itsearviointiväittämien kohdalla. Kuten kirjallisissa itsarviointiväittämissä, tarkastelun tarkoituksena oli yrittää selvittää, sisältävätkö kyseiset puhenäytteet mahdollisesti sellaisia elementtejä, jotka selittäisivät niiden hyvän toimivuuden tai niiden toimimattomuuden.

## 6.1 Esimerkki hyvin toimineesta puhenäytteestä

Alla on karkeasti litteroitu esimerkki puhenäytteen numero 11 analyysistä (tehtävä 1). Tarkoituksena ei ole tehdä syvällistä kielellistä analyysiä puhenäytteen puhujan suullisesta kielitaidosta, vaan keskittyä tarkemmin näytteiden tilastollisiin ominaisuuksiin ja mahdollisiin oppijoiden kommentteihin. Kyseessä on näyte, joka toimii hyvin sekä perinteisen että Rasch-analyysin näkökulmasta. Näytteessä esiintyvät pääasialliset literointimerkinnot ovat seuraavat: tauot on merkitty \_ -viivoin ja selvästi ranskan kielen ääntämissäännöistä poikkeava ääntäminen on merkitty hakasulkeisiin.

### Näyte 11

(A2; korrelaatio 0,75; Infit mean square 0,82; Infit Zstd -1,70; vastausjakauma 92–117)

-et \_ \_ et ici \_ c'est François\_ François est \_le fils de \_ Cécile \_ il a une amie mais \_ il habite \_ avec sa mere\_ ils habitent en \_ Marseille \_ \_euh\_ François est \_ étudiant [t ääntyy lopussa] \_ il etude photographique \_ \_ \_euh François \_ aime football \_et \_ sport \_ euh \_il \_ jouer [r ääntyy lopussa] \_ souvent [t ääntyy lopussa] football avec \_ ses amis (44 sek.)

Tulkinnan helpottamiseksi seuraavassa esitetään yleisluonnehdinta puhetuoksesta. Kyseessä on tutkijan oma pelkistetty ja luettelomainen näkemys tuotetun puheen ominaisuuksista:

Naisääni, joutuu miettimään aika tavalla tuottaessaan puhetta, mistä johtuen tuotos on hitaahko ja sisältää paljon taukoja; kielioppi ei vielä täysin hallinnassa (esim. artikkelien käyttö puutteellista tai sattumanvaraista), ääntäminen ei vielä kohdekielenomaista, mutta tuotos on ymmärrettävä. Puhuja onnistuu kuitenkin kertomaan tarinan yksinkertaisena luettelona.

Tutkimuksen puhenäytteitä itsearviointiin käytettäessä oppijoiden huomio kiinnittyi useimmiten sujuvuuteen tai sen puutteeseen liittyviin seikkoihin. Näytettä 11 kommentoitiinkin runsaasti juuri siitä näkökulmasta, että näyte ei ole sujuva. Näytteen puhuja edustanee tyyppillistä kielitaidon alatasojen puhujaa, jolla on paljon taukoja puhees-

saan, mistä johtuen oppijoille saattoi jäädä ei-sujuvan puhujan vaikutelma näytteestä. Vaikeustasonsa puolesta näyte sijoittuu oppijoiden mielestä aivan A2-tason ylärajalle, missä se on hyvin lähellä B1-tasoisia näytteitä. Näyte vaikuttaa vastaavan hyvin Viitekehysten A2-tasolle määrittelemää puhumisen osaamista.

## 6.2 Esimerkki huonosti toimineesta puhenäytteestä

Kyseessä on näyte, joka toimi hyvin perinteisen analyysin näkökulmasta, mutta sen molemmat Rasch-analyysin arvot ovat heikkoja. Oppijat ovat kommentoineet näytettä jopa 38 kertaa ja merkinneet vaikeaa päättää -sarakkeeseen rastin suhteellisen usein (16 kpl). Näytteen pääasialliset litterointimerkinnot ovat samat kuin edellisessä näytteessä: tauot on merkitty \_ -viivoin.

### Näyte 24

(taso B2; korrelaatio 0,57; Infit Mean square 1,38/ZStd 3,40; vastausjakauma 104–105)

Et voilà Francois\_ euh\_ il est étudiant à l'université\_ dans la faculté de méd- médecine\_ \_  
et\_ \_ il aime\_ \_vraiment beaucoup des\_ des sports\_ \_ basket ball \_ euh\_ surtout\_ \_ il est  
âgé de\_ \_âgé de\_ \_ de\_ vingt-quatre ans \_ \_ \_ bon\_ \_ euh\_ \_ \_ \_ il va \_ \_ \_ il va faire \_ une  
carrière terrible je crois (39 sek.)

Kuten edellisen näytteen kohdalla, tutkija on laatinut pelkistetyn yleiskuvauksen näytteestä 24:

Näytteen miespuhujaa ääntää ranskan kieltä kohdekielenomaisesti ja käyttää ranskan kielelle tyypillisiä lauserakenteita. Tuotos on kuitenkin lyhyt (39 sekuntia), mutta sen ajan, kun henkilö puhuu, hän tuottaa puhetta sujuvasti ja keksii aiheesta suhteellisen paljon sanottavaa. Puhuja lopettaa tuotoksensa puhekielenomaiseen humoristiseen ilmaisuun (*il va faire une carrière terrible*<sup>3</sup>), jota monet tutkimukseen osallistuneet aloittelevat ranskan oppijat pitivät outona. Sen sijaan monien pidemmälle kieliopinnoissaan edenneiden oppijoiden mielestä lopetus oli erityisen hauska.

Viisi arvioijaa, jotka arvioivat kuunneltavat näytteet taitotasolle ennen tutkimusaineiston keruun aloittamista, olivat jakautuneet kahteen ryhmään tämän näytteen arvioinnissa siten, että kolme arvioijaa on sijoittanut tämän näytteen taitotasolle B2, kun taas kaksi arvioijaa oli sijoittanut sen alemmalle tasolle (B1). Näytteen sujuvuuteen ja puhekielisyyteen viitaten on todennäköistä, että myös oppijat saattoivat jakaantua mielipiteissään sen suhteen, millaisena he pitivät tätä näytettä. Todennäköistä onkin, että näyte saattaa edustaa pikemmin taitotasoa B1 kuin tasoa B2 tai ainakin siitä voidaan

3 Suora käännös: 'Hän tulee tekemään hirvittävän uran [jalkapalloilijana].'

perustellusti olla kahta eri mieltä riippuen siitä, mitä kriteeriä puhetuotoksen arvioinnissa painotetaan.

### 6.3 Yhteenvedoa puhenäytteiden toimivuudesta

Kun tarkastellaan puhenäytteiden toimimista, voidaan listata joitakin ominaisuuksia, jotka näyttävät vaikuttavan siihen, että jokin puhenäyte toimi tilastollisesti hyvin. Osioiden pysyminen niille määritellyllä taitotasolla eli se, että oppijat kokevat osioiden kuuluvan vaikeustasoltaan niille määritellyllä tasolla, näyttää ennakoivan sitä, että osio toimii hyvin myös tilastollisesti. Lisäksi ne puhenäytteet, jotka sisältävät tietylle taitotasolle tyypillisiä ominaisuuksia, näyttävät edustavan hyvin toimivia näytteitä.

Tilastollisissa analyyseissä huomosti toimiville puhenäytteille tyypillistä puolestaan näytti olevan se, että kahden eri taitotason rajalle (joko asteikon yläpäässä tasoille C1–C2 tai alapäässä tasoille A1–A2) sijoittuvat näytteet käyttäytyvät ”epävakaasti”. Niillä tuntuu olevan selvä vaikutus tietyn osion korrelaatioon suhteessa koko käytetyn instrumentin kokonaispistemäärään.

Yksi tärkeimmistä havainnoista, joka puhenäytteiden avulla suoritetusta itsearvioinnista voidaan tehdä, on se, että oppijoilla oli eniten vaikeuksia verrata omaa kielitaitoaan niihin näytteisiin, jotka edustivat likimain samantasoista kielitaitoa kuin heidän oma suullinen kielitaitonsa oli. Tämä tulos saatiin esille tarkastelemalla näytteille annettujen vaikea päättää -rastien lukumääriä. Kyse ei siis näyttäisi olevan siitä, että jonkin kielen vasta-alkajilla olisi vaikeuksia verrata omaa tuotostaan kuultuihin puhenäytteisiin, vaan siitä, että oppijoiden oli vaikea verrata oman puhetuotoksensa paremmuutta tai huonommuutta suhteessa niihin puhenäytteisiin, jotka kuulostivat ”tutuilta” ja olivat samantasoisia kuin heidän oma ranskan kielen suullinen taitonsa. Toisaalta tässä kohden on syytä painottaa, että tutkimuksessa omaa kielitaitoa arvioitaessa oppijoiden oli mahdollista valita vain ”kyllä” tai ”ei” -vaihtoehtojen väliltä. Jos vastauslomakkeessa olisi ollut samaa tasoa osoittava vastausvaihtoehto, olisivat tutkimuksen tulokset voineet olla erilaisia kuin mitä ne nyt ovat.

Kuten perinteisen itsearvioinnin tilannekohtaisten itsearviointiväittämien kohdalla jo todettiin, eri testiteoriat nostavat tarkasteluun osioiden eri puolia. Tästä johtuen eri analyytit antavat viitteitä mahdollisista ongelmallisista osioista, jotka eivät kuitenkaan välttämättä ole samoja osioita. Jonkinlaisena nyrkkisääntönä voidaan varmasti pitää sitä, että jos molempien teorioiden ilmaiset tilastolliset tunnusluvut osoittavat jonkin osion mahdollisen ongelmallisuuden, osiossa on mitä todennäköisimmin jonkin tasoinen ongelma. Nämä toteamukset pitävät paikkansa yhtä hyvin tilannekohtaisten itsearviointiväittämien kuin kuunneltavien puhenäytteidenkin osalta.

## 7 Uuden itsearviointimenetelmän toimivuudesta, luotettavuudesta ja käytettävyydestä

Uusi puhenäytteiden kuuntelemiseen ja oman suorituksen vertaamiseen perustuva itsearviointimenetelmä korreloi tutkimuksessa varsin hyvin ulkoisen arvioinnin (koulutettujen arvioijien suorittama arviointi, jota ei käsitellä tässä artikkelissa) sekä kirjallisten itsearviointiväittämien avulla tehtävän itsearvioinnin kanssa. Uusi itsearviointimenetelmä osoittautui niin validiksi ja reliaabeliksi, että sitä voidaan ajatella käytettävän suullisen kielitaidon itsearviointi- ja opetustilanteissa.

Uusi puhenäytteiden kuuntelemiseen pohjautuva itsearviointimenetelmä näyttää soveltuvan hyvin suullisen kielitaidon itsearviointiin konkreettisuutensa ansiosta. Oppijat pitänevät puhenäytteiden kuuntelemista ja oman suorituksen vertaamista niihin autenttisena, motivoivana ja mielekkäänä tehtävänä. Puhenäytteiden kuuntelemiseen pohjautuva itsearviointimenetelmä sopii todennäköisesti varsinkin niille oppijoille, jotka eivät ole halukkaita lukemaan kirjallisia itsearviointiväittämiä. Uuden menetelmän käyttö poistaa myös ymmärtämis- tai tulkintaongelmat, joita saattaa ilmetä kirjallisten itsearviointiväittämien käytön yhteydessä. Lisäksi uusi menetelmä kehittää oppijan auditivista erottelukykyä ja tietoisuutta suullisen kielitaidon ominaispiirteistä sekä totuttaa oppijaa suhteuttamaan omaa kielitaitonsa tasoa eritasoisten kielen puhujien taitotasoon. Kun oppija kuulee muiden puhenäytteitä tutusta tehtävästä, hän oppii myös huomaamaan, mitkä ovat hänen oman kielitaitonsa puutteet. Tämän huomion taas toivoisi johtavan siihen, että oppija motivoituisi oppimaan lisää. Uutta itsearviointimenetelmää voidaan suositella käytettäväksi perinteisen kirjallisten itsearviointiväittämien avulla tehtävän itsearvioinnin rinnalla, koska sen käyttö tekee arviointikäytänteistä monipuolisempia ja oikeudenmukaisempia.

## 8 Tutkimuksen ansioista, rajoituksista ja jatkotutkimusaiheista

Tutkimuksen ansioina voidaan pitää sen arviointi-instrumenttien laadukkuuden varmistamista ja perinpohjaisia tilastollisia analyysejä. Rajoituksia sen sijaan ovat ainakin tutkimusaineiston hankintatapa (erityisesti ”kyllä” ja ”ei”-vaihtoehtojen dikotoomisuus) ja taitotasosteikossa ylimmille tasoille sijoittuvien puhenäytteiden vähäinen määrä. Jatkotutkimusaiheita ovat esimerkiksi samankaltaisen tutkimuksen tekeminen käyttämällä monipuolisempia vastausvaihtoehtoja, puhenäytteiden tarkempi kielellinen ja sisällöllinen analyysi sekä samankaltaisen tutkimuksen toteuttaminen jossain muussa kuin ranskan kielessä.

## Kirjallisuus

- Atjonen, P. 2007. *Hyvä, paha arviointi*. Helsinki: Tammi.
- Blanche, P. & B. J. Merino 1989. Self-assessment of foreign language skills: implications for teachers and researchers. *Language Learning*, 39 (3), 313–340.
- Blue, G. M. 1988. Self-assessment: the limits of learner independence. Teoksessa A. Brookes & P. Grundy (toim.) *Individualisation and autonomy in language learning*. ELT documents 131. London: Modern English Publications in association with the British Council, 100–118.
- Dragemark Oscarson, A. 2009. *Self-assessment of writing in learning English as a foreign language. A study at the upper secondary school level*. Väitöskirja. Göteborg: Acta universitatis Gothoburgensis.
- EK = Elinkeinoelämän keskusliitto 2011. *Oivallus loppuraportti* [online]. [Luettu 11.6.2014]. Saatavissa: [http://ek.multiedition.fi/oivallus/fi/liitetiedostot/Oivallus\\_loppuraportti\\_web.pdf](http://ek.multiedition.fi/oivallus/fi/liitetiedostot/Oivallus_loppuraportti_web.pdf).
- von Elek, T. 1985. A test of Swedish as a second language. Teoksessa Y. P. Lee, A. C. Y. Y. Fok, R. Lord & G. Low (toim.) *New directions in language testing: papers presented at the international symposium on language testing*. Oxford: Pergamon Press, 47–57.
- Evers, R. 1981. Self-evaluation of proficiency levels in French as a second language. Esitelmä konferenssissa *Sixth International Congress of Applied Linguistics*, Lund, Sweden.
- EVK = Eurooppalainen viitekehys. *Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys*. 2003. Helsinki: WSOY.
- Gyagenda, I. S. & G. Engelhard, Jr. 2009. Using classical and modern measurement theories to explore rater, domain, and gender influences on student writing ability. *Journal of Applied Measurement*, 10 (3), 225–246.
- Holec, H. 1990. Apprendre à l'apprenant à s'évaluer: quelques pistes à suivre. *Etudes de Linguistique Appliquée*, 80, 39–47.
- Huhta, A. 2003. Itsearviointi kielitaidon arviointimuotona. Teoksessa A. Airola (toim.) *Kokemuksia suullisen kielitaidon arviointihankkeesta*. Joensuu: Pohjois-Karjalan ammattikorkeakoulu, 20–37.
- LeBlanc, R. & G. Painchaud 1985. Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19 (4), 673–687.
- Leblay, T. 2013. "Voi ei, nää on tosi hyviä verrattuna muhun!": uudenlaisen arviointimenetelmän toimivuus ranskan suullisen kielitaidon itsearvioinnissa. *Jyväskylän studies in humanities* 211. Jyväskylä: Jyväskylän yliopisto. Saatavissa: <https://jyx.jyu.fi/dspace/handle/123456789/42162>.
- Lee-Ellis, S. 2009. The development and validation of a Korean C-Test using Rasch analysis. *Language Testing*, 26 (2), 245–274.
- Linacre, J. M. 2014. *A user's guide to FACETS. Rasch-model computer programs. Program manual 3.71.4* [online]. [Luettu 11.6.2014]. Saatavissa: <http://www.winsteps.com/a/facets-manual.pdf>.
- LOPS = Lukion opetussuunnitelman perusteet. 2003. Opetushallitus. Helsinki: Opetushallitus.
- Lynch, B. K. & T. F. McNamara 1998. Using G-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15 (2), 158–180.
- MacIntyre, P. D., K. A. Noels & R. Clément 1997. Biases in self-ratings of second language proficiency: the role of language anxiety. *Language Learning*, 47 (2), 265–287.
- Matsuno, S. 2009. Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26 (1), 75–100.



- Mok, M. M. C., C. L. Lung, D. P. P. W. Cheng, R. H. P. Cheung & M. L. Ng 2006. Self-assessment in higher education: experience in using a metacognitive approach in five case studies. *Assessment & Evaluation in Higher Education*, 31 (4), 415–433.
- Ockey, G. 2011. Self-consciousness and assertiveness as explanatory variables of L2 oral ability: a latent variable approach. *Language Learning*, 61 (3), 968–989.
- Oscarson, M. 1980. *Approaches to self-assessment in foreign language learning*. Oxford: Pergamon Press.
- Oscarson, M. 1989. Self-assessment of language proficiency: rationale and applications. *Language Testing*, 6 (1), 1–13.
- Oscarson, M. 1997. Self-assessment of foreign and second language proficiency. Teoksessa C. Clapham & D. Corson (toim.) *Language testing and assessment 7. The encyclopedia of language and education*. Dordrecht: Kluwer Academic, 175–187.
- Oscarson, M. 1998. Om självbedömning av språkfärdighet – empiri och reflektioner. Teoksessa B. Ljung & A. Pettersson (toim.) *Perspektiv på bedömning av kunskap*. Stockholm: Lärarhögskolan i Stockholm, Institutionen för pedagogik, 133–156.
- Perrenoud, P. 2011. *Quand l'école prétend préparer à la vie: développer des compétences ou enseigner d'autres savoirs?* Issy-les-Moulineaux: ESF.
- POPS = *Perusopetuksen opetussuunnitelman perusteet*. 2004. Opetushallitus. Helsinki: Opetushallitus.
- Rasch, G. 1979. *Probabilistic models for some intelligence and achievement tests*. Copenhagen: Danish Institute for Educational Research.
- Ross, S. 1998. Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing*, 15 (1), 1–20.
- Sinharay, S., D. E. Powers, Y. Feng, L. Saldivia, A. Giunta, A. Simpson & V. Weng 2009. Appropriateness of the TOEIC Bridge test for students in three countries of South America. *Language Testing*, 26 (4), 589–619.
- Taras, M. 2001. The use of tutor feedback and student self-assessment in summative assessment tasks: towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26 (6), 606–614.
- Thomson, C. K. 1996. Self-assessment in self-directed learning: issues of learner diversity. Teoksessa R. Pemberton, E. S. L. Li, W. W. F. Or & H. D. Pierson (toim.) *Taking control: autonomy in language learning*. Hong Kong: Hong Kong University Press, 77–91.
- Tudor, I. 1996. *Learner-centredness as language education*. Cambridge: Cambridge University Press.