

Tarja Leblay

”Voi ei, nää on tosi hyviä  
verrattuna muhun!”

Uudenlaisen arviointimenetelmän  
toimivuus ranskan suullisen  
kielitaidon itsearvioinnissa



JYVÄSKYLÄ STUDIES IN HUMANITIES 211

Tarja Leblay

”Voi ei, nää on tosi hyviä  
verrattuna muhun!”

Uudenlaisen arviointimenetelmän toimivuus  
ranskan suullisen kielitaidon itsearvioinnissa

Esitetään Jyväskylän yliopiston humanistisen tiedekunnan suostumuksella  
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212  
syyskuun 21. päivänä 2013 kello 12.

Academic dissertation to be publicly discussed, by permission of  
the Faculty of Humanities of the University of Jyväskylä,  
in Auditorium S212, on September 21, 2013 at 12 o'clock noon.



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2013

”Voi ei, nää on tosi hyviä  
verrattuna muhun!”

Uudenlaisen arviointimenetelmän toimivuus  
ranskan suullisen kielitaidon itsearvioinnissa

JYVÄSKYLÄ STUDIES IN HUMANITIES 211

Tarja Leblay

”Voi ei, nää on tosi hyviä  
verrattuna muhun!”

Uudenlaisen arviointimenetelmän toimivuus  
ranskan suullisen kielitaidon itsearvioinnissa



UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2013

Editors

Ari Huhta, Centre for Applied Language Studies, University of Jyväskylä

Pekka Olsbo, Sini Tuikka

Publishing Unit, University Library of Jyväskylä

Jyväskylä Studies in Humanities

Editorial Board

Editor in Chief Heikki Hanka, Department of Art and Culture Studies, University of Jyväskylä

Petri Karonen, Department of History and Ethnology, University of Jyväskylä

Paula Kalaja, Department of Languages, University of Jyväskylä

Petri Toiviainen, Department of Music, University of Jyväskylä

Tarja Nikula, Centre for Applied Language Studies, University of Jyväskylä

Raimo Salokangas, Department of Communication, University of Jyväskylä

URN:ISBN:978-951-39-5384-3

ISBN 978-951-39-5384-3(PDF)

ISBN 978-951-39-5383-6 (nid.)

ISSN 1459-4323 (nid.), 1459-4331 (PDF)

Copyright © 2013, by University of Jyväskylä

Jyväskylä University Printing House, Jyväskylä 2013

## ABSTRACT

Leblay, Tarja

Oh no, they are so good compared to me! The functionality of a new self-assessment method of French oral language skills

Jyväskylä: University of Jyväskylä, 2013, 321 p.

(Jyväskylä Studies in Humanities;

ISSN 1459-4323 (nid.) 1459-4331 (PDF); 211)

ISBN 978-951-39-5383-6 (nid.)

ISBN 978-951-39-5384-3 (PDF)

Diss.

This empirical study focuses on the 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 could use different methods of self-assessment as well as the extent to which different self-assessments corresponded to each other and to external evaluation. Through qualitative analysis the study aimed at investigating among other things the perceptions of the learners concerning the functionality of the written self-assessment statements. In addition, the study examined the characteristics of those speech samples which the students considered to be either particularly easy or difficult. In the new method of self-assessment learners compared their performances on speaking tasks with speech samples collected from other learners. The results of the speech sample-based self-assessment were then compared to the results of the traditional paper-based self-assessment and to those of the external evaluation. The data were analysed by using both classical and IRT-based item analyses. In addition, the study included a qualitative part which provided additional information about students' reflections and feelings about their own actions in the self-assessment process. The study involved 209 Finnish students studying French, most of whom were university students. The research material consisted of two French speaking tasks, which the students did in the language laboratory, as well as four different self-assessment instruments (general self-assessment on paper, situation-specific self-assessment on paper, self-assessment based on speech samples in tasks 1 and 2). The first two above-mentioned instruments, general self-assessment and situation-specific self-assessment (on paper), were developed for the DIALANG<sup>1</sup> online diagnostic language assessment tool / system. Research also relates to the Council of the Europe's Common European Framework of Reference<sup>2</sup> and the Finnish National Certificates of Language Proficiency<sup>3</sup>. 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. This result is quite promising especially due to the fact that the majority of the participating students did not have prior experience with self-assessment. However, it should be noted that in this study the students managed relatively well also in the self-assessment on paper based on situation-specific self-assessment statements.

Keywords: Self-assessment, assessment, self-assessment statements, paper and pencil self-assessment, self-assessment on paper, self-assessment based on speech samples, self-assessment methods, reliability of self-assessment, self-assessment and students' perceptions, French as a foreign language, speaking, oral proficiency, reflectivity

---

<sup>1</sup> <http://www.lancs.ac.uk/researchenterprise/dialang/about>

<sup>2</sup> Common European Framework of Reference for Languages

<sup>3</sup> [www.oph.fi](http://www.oph.fi)

**Author's address** Tarja Leblay  
Centre for Applied Language Studies  
University of Jyväskylä  
P.O. Box 35  
FI-40014 University of Jyväskylä  
Finland  
tarja.leblay@oph.fi

**Supervisor** Professor Ari Huhta  
Centre for Applied Language Studies  
University of Jyväskylä

**Reviewers** Docent, Ph.D., Raili Hildén  
Department of Teacher Education  
University of Helsinki

Ph.D. Taina Juurakko-Paavola  
Language Centre  
HAMK University of Applied Sciences

**Opponent** Professor Heini-Marja Järvinen  
Department of Teacher Education  
University of Helsinki

## AVANT-PROPOS

Quelques lignes en français pour situer linguistiquement les lignes qui suivent. Ce travail a d'abord été rédigé, presque complètement, en français, bien avant de l'être en finnois. La raison de ce changement de langues vient d'une prise de position revendiquée sur le statut d'une langue minoritaire comme le finnois (environ six millions de locuteurs). Fallait-il m'exprimer en français au risque de négliger ce que peut apporter ma propre langue maternelle à la recherche? Trop peu de travaux de recherche sont écrits en finnois, au détriment de langues comme l'anglais, le français ou l'allemand. De même, la recherche produite en langue française s'efforce d'exister face à celle entreprise en langue anglaise, de même, il m'a semblé opportun de reprendre la rédaction de ma thèse en finnois, afin de lui donner un droit de cité dans l'espace discursif de la recherche. Ce travail en finnois n'est pas, pour autant, une traduction du français, mais bien une complète réécriture qui m'a permis de reprendre et de repenser en deux langues certains concepts développés. Je dois à cette réécriture un très net éclaircissement de mon propos.

## ESIPUHE

Tämä väitöskirjatutkimus on pitkän ajan kuluessa valmistunut todiste siitä, että omiin haasteisiin vastaaminen on tärkeää. Tutkimukseni valmistuminen kesti yli kymmenen vuotta kaikkine elämään liittyvine muutoksineen ja kiireineen. Väitöskirjatutkimukseni on suomenkielinen. Alunperin tarkoitukseni oli kirjoittaa tämä opinnäytetyö ranskan kielellä, mutta vuosien kuluessa mielessäni kypsyi ajatus suomen kielellä kirjoittamisesta. Päätöstäni vauhditti tieto, jonka mukaan Suomessa tehtävistä väitöskirjatutkimuksista vain pieni murto-osa kirjoitetaan suomen kielellä. Olkoon tämä tutkimus siis minun kannanottoni suomen kielen käytön puolesta myös tieteen kielenä.

Tutkimukseni liittyy laajempaan kontekstiin, johon tutustuin tarkemmin aloittaessani työt Jyväskylän yliopiston koordinoimassa DIALANG-hankkeessa 2000-luvun alussa. Tämän eurooppalaisen yhteistyöprojektin itsearviointityöryhmän jäsenenä pääsin perehtymään juurta jaksan kielitaidon arvioinnin moniin kiinnostaviin kysymyksiin. DIALANG-hankkeen edetessä syntyi myös ajatus siitä, että olisi kiintoisaa kokeilla, miten oppijan omaa puhumista voisi arvioida konkreettisten puheasuoritusten avulla.

Tutkimuksessa käsitellään monia asioita, joista on tullut minulle tärkeitä vuosien kuluessa. Tutkimukseni liittyy itsearviointiin. Itsearviointiin liittyvä tutkimusaihe selittyy sillä, että pidän motivoivana ja jopa välttämättömänä sitä, että ihmiset ottavat elämänsä niin sanotusti omiin käsiinsä ja "tekevät sille jotain". Toisaalta itsearviointiin kuuluu myös vertaisarviointi, jota myös sivutaan tässä tutkimuksessa. Siinä minua viehättää se, että se on askel toisen ihmisen kanssa tehtävän yhteistyön ja vuorovaikutuksen suuntaan. Myös suullinen kielitaito on lähellä sydäntäni, koska siihenkin sisältyy pyrkimys lähestyä toista ihmistä ja/tai panna itsensä likoon nopeasti tilanteissa, joita useinkaan ei ole



voinut suunnitella ennakolta. Tärkein kiintymykseni kohde on kuitenkin ranskan kieli, joka minulle merkitsee maailman kauneinta kieltä. Vankka pohja tälle pitkäaikaiselle rakkaudelle valettiin jo lukiossa, ja sen jälkeen ranskan kieli on ollut osa minua ja elämääni.

Haluan kiittää yhteisesti kaikkia niitä henkilöitä, jotka ovat jollakin tavalla tai jossain tutkimuksen vaiheessa osallistuneet tutkimukseni käytännön toteuttamiseen. Erikseen kiitän kollegoitani, jotka toimivat aineiston arviointitehtävissä ja joita ilman tätä tutkimusta ei olisi voitu tehdä. Kiitän siten nöyrimmin seuraavia henkilöitä: Jari Aarnio, Karine Larcher, Christophe Leblay, Simo Lönnblad, Tarja Nyman ja Outi Veivo. Kiitän myös niitä ranskan kielen opettajia eri oppilaitoksissa, jotka antoivat oppilaidensa osallistua tutkimukseen oppituntiansa aikana. Oman tärkeän osansa kiitoksista ansaitsevat tutkimukseen osallistuneet ranskan kielen opiskelijat, jotka osoittivat ihailtavaa sitoutumista, ennakkoluulottomuutta, rohkeutta ja uteliaisuutta osallistumalla tutkimukseen. Tämän lisäksi haluan kiittää eri oppilaitosten kielistudioista vastuussa olleita henkilöitä myötämielisestä suhtautumisesta tutkimukseni aineiston keruuta kohtaan sekä Jyväskylän, Tampereen ja Helsingin yliopistojen hallintohenkilöstöä, joiden ystävällisen suhtautumisen ansiosta sain yhteyden moniin ranskan kielen opiskelijoihin ja tilat käyttööni tutkimusaineiston keräämistä varten.

Erikseen haluan kiittää tutkimukseni ohjaajia. Professori Sauli Takala vastasi tutkimukseni alkuvaiheen ohjauksesta tukemalla laaja-alaisen asiantuntemuksensa avulla työtäni monin eri tavoin. Professori Takalan siirryttyä eläkkeelle tutkimukseni ohjaajaksi tuli professori Ari Huhta, jonka kokemus kielitaidon arvioinnin ja erityisesti itsearviointin alalta on mahdollistanut tutkimukseni aiheen osittaisen eteenpäin kehittelyn.

Oman erityisen mainintansa ansaitsee jo edesmennyt Soveltavan kielen tutkimuksen keskuksen johtaja emeritusprofessori Kari Sajavaara, joka ystävällisen inhimillisellä ja isällisellä otteellaan ohjasi ja neuvoi eteenpäin meitä tutkimusentekijöitä.

Kiitän myös nyt jo edesmennyttä professori Felianka Kaftandjievaa kaikista niistä keskusteluista, joiden pohjalta syntyi idea tämän tutkimuksen tutkimusasetelmasta ja uudenlaisen itsearviointimenetelmän kokeilusta.

Työni esitarkastajia FT Raili Hildéniä ja FT Taina Juurakko-Paavolaa kiitän lämpimästi heiltä saamistani asiantuntevista ja hyödyllisistä kommenteista.

2000-luvun alussa Jyväskylän yliopiston Soveltavan kielentutkimuksen keskuksessa (SOLKI) työskennelleet kollegat ovat myös eri tavoin osallistuneet tämän tutkimuksen edistämiseen. Muun muassa YTM, suunnittelija Ari Maijanen kokosi tutkimukseen osallistuneiden oppijoiden puhe-suoritukset sellaiseen muotoon, että arvioijat pystyivät niitä arvioimaan. SOLKI:ssa työskentely mahdollisti sen, että pääsin tekemään kielitaidon arviointiin liittyvää työtä sekä Yleisissä kielitutkinnoissa että DIALANG-hankkeessa, ja siten perehtymään kielitaidon arviointiin käytännön työssä. SOLKIa kiitän myös matkavastuksista, joita sain matkustaessani eri kaupunkeihin tutkimusaineistoa keräämään.

Haluan kiittää myös työnantajaani Opetushallitusta siitä, että olen saanut virkavapaata tämän työn edistämiseen. Työn taloudellisesta tukemisesta kiitän Ellen ja Artturi Nyyssösen säätiötä sekä Jyväskylän yliopistoa.

Suurin kiitos kuuluu kuitenkin perheeni kaksi- ja nelijalkaisille jäsenille. Osoitan kiitokseni lämpimän rapsutuksen muodossa elämäni kissoille, nyt jo edesmenneelle Senille sekä Mokille ja Mimille. Kissamme asettuivat kukin vuorollaan pöydälle levitettyjen papereideni ja kirjojeni päälle kerälle kehräämään. Kissojemme rauhoittava kehräys loi otollisen työskentelyilmapiirin niihin hetkiin, jolloin olin syventynyt tutkimukseeni. Myös elämäni koira Miel ansaitsee emännältään kunnon rutistuksen. Aina onnellisen oloinen Miel vei minut metsään lenkille tuulettamaan aivojani ja tartutti myös minuun samalla hyvän tuulensa.

Tyttäremme Lila ja hänen elämäntarinansa muistuttavat minua joka päivä siitä, että tyttöjen ja naisten kouluttautuminen on ensisijaisen tärkeää. Pienen tytön äidiksi kasvaminen muistutti minua tutkimuksenteon ohessa siitä, mitä tarkoittaa asioiden välisen tasapainon löytäminen elämässä.

Lopuksi kiitän aviomiestäni Christophea, tukijoistani tärkeintä ja rakkainta. Christophe on aina uskonut minuun; niin myös tämän pitkäkestoisen tutkimuksen valmistumisen suhteen. Hänen kannustuksensa innoittamana olen jo kauan sitten lähtenyt kulkemaan omaa polkuani, mikä onkin osoittautunut antoisaksi valinnaksi. *Merci pour ta patience, ta complicité et surtout pour ton amour, Christophe.*

## KUVIOT

KUVIO 1	Tutkimusmateriaalin valmistelun, aineiston keruun ja tutkimuksen eri vaiheet.....	87
KUVIO 2	Tehtävä 1: arviointiasteikossa käytetyt kommunikatiiviset kompetenssit.....	104
KUVIO 3	Tutkimukseen osallistuneiden oppijoiden ikäjakauma.....	114
KUVIO 4	Tutkimukseen osallistuneiden oppijoiden ranskan kielen opintojen aloittamisikä.....	116
KUVIO 5	Paperilla tehty itsearviointi: väittämien vaikeus oppijoiden vastausten perusteella.....	149
KUVIO 6	Tehtävä 1: puhenäytteiden vaikeus oppijoiden vastausten perusteella.....	169
KUVIO 7	Tehtävä 2: puhenäytteiden vaikeus oppijoiden vastausten perusteella.....	181
KUVIO 8	Tutkimuksessa käytetyn kuunneltaviin puhenäytteisiin perustuvan itsearviointimenetelmän käyttömahdollisuuksia kielten opetuksessa.....	263

## TAULUKOT

TAULUKKO 1	Tehtävän 1 ominaisuuksien luokittelua Fulcherin (2003) ja Weirin (1990) mallien pohjalta.....	90
TAULUKKO 2	Tehtävän 2 ominaisuuksien luokittelua Fulcherin (2003) ja Weirin (1990) mallien pohjalta.....	92
TAULUKKO 3	Puhumisen itsearviointilomakkeen väittämien luokittelu kielenkäyttötilanteen ja taitotason mukaan.....	97
TAULUKKO 4	Tutkimuksen osallistujien oppilaitokset, osallistujamäärät ja prosenttiosuudet / oppilaitos.....	112
TAULUKKO 5	Tutkimuksen osallistujien kontaktit ranskan kieleen tutkimushetkellä.....	117
TAULUKKO 6	Tutkimuksen osallistujien ranskan kielen käytön frekvenssi tutkimushetkellä.....	118
TAULUKKO 7	Kirjallisiin itsearviointiväittämiin perustuva itsearviointi: perinteisen osioanalyysin mukaiset hyvin ja huonosti toimineet väittämät.....	130
TAULUKKO 8	Kirjallisiin itsearviointiväittämiin perustuva itsearviointi: modernin osioanalyysin (Rasch-analyysi) mukaiset hyvin ja huonosti toimineet väittämät.....	131
TAULUKKO 9	Mahdollisesti ongelmalliset väittämät perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta (paperilla tehty itsearviointi).....	146

TAULUKKO 10	Kuunneltuihin puhenäytteisiin perustuva itsearviointi: perinteisen osioanalyysin mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 1). ....	158
TAULUKKO 11	Kuunneltuihin puhenäytteisiin perustuva itsearviointi: modernin osioanalyysin (Rasch-analyysi) mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 1).....	159
TAULUKKO 12	Mahdollisesti ongelmalliset puhenäytteet perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta (tehtävä 1).....	167
TAULUKKO 13	Kuunneltuihin puhenäytteisiin perustuva itsearviointi: perinteisen osioanalyysin mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 2). ....	172
TAULUKKO 14	Kuunneltuihin puhenäytteisiin perustuva itsearviointi: modernin osioanalyysin (Rasch-analyysi) mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 2).....	173
TAULUKKO 15	Mahdollisesti ongelmalliset puhenäytteet perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta (tehtävä 2).....	178
TAULUKKO 16	Perinteinen osioanalyysi: kolmen itsearviointimittarin reliabiliteetti ja osio-testi-korrelaatiot (.30 ylittävien prosenttiosuus, kappalemäärät ja prosenttiosuudet suuruusluokittain). ....	183
TAULUKKO 17	Rasch-analyysi: kolmen itsearviointimittarin reliabiliteetti ja Infit mean square -arvot (.70–1.30 välille sijoittuvien osioiden prosenttiosuus, kappalemäärät ja prosenttiosuudet suuruusluokittain).....	185
TAULUKKO 18	Kolmen itsearviointimittarin ongelmalliset osiot perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta. ....	187
TAULUKKO 19	Arvioijan korrelaatio suhteessa muihin arvioijiin ja arvioijan poistamisen vaikutus reliabiliteettiin. ....	191
TAULUKKO 20	Ulkoisten arvioijien väliset korrelaatiot tehtävissä 1 ja 2. ....	192
TAULUKKO 21	Yhdistettyjen ulkoisten arviointien (tehtävät 1 ja 2) ja itsearviointien väliset korrelaatiokertoimet. ....	195
TAULUKKO 22	Ulkoisten arviointien ja itsearviointien väliset korrelaatiokertoimet (tehtävä 1). ....	196
TAULUKKO 23	Ulkoisten arviointien ja itsearviointien väliset korrelaatiokertoimet (tehtävä 2). ....	197
TAULUKKO 24	Oppijoiden ranskan kielen puhumistaito itsearvioituna holistisesti 6-portaisella EVK-asteikolla. ....	200
TAULUKKO 25	Tehtävien 1 ja 2 kuunneltuja puhenäytteitä koskevat kommenttikategoriat. ....	209
TAULUKKO 26	Puhenäytteiden litterointimerkintöjä. ....	211

TAULUKKO 27	"Vaikea päättää" -rastien prosentuaalinen määrä analyysiin valituista opiskelijoiden vastauslomakkeista (40 % kaikista lomakkeista), tehtävä 1.....	212
TAULUKKO 28	"Vaikea päättää" -rastien prosentuaalinen määrä analyysiin valituista opiskelijoiden vastauslomakkeista (40 % kaikista lomakkeista), tehtävä 2.....	218
TAULUKKO 29	Tutkimuksessa hyvin tai huonosti toimineiden itsearviointiväittämien ominaisuuksia. ....	232
TAULUKKO 30	Bachmanin ja Palmerin (1996) testin luotettavuutta ( <i>test usefulness</i> ) tarkasteleva malli. ....	241
TAULUKKO 31	Aineiston keruun yksityiskohtien erittelyä. ....	253
TABLE 32	Correspondence of different methods of self-assessment: three highest correlation coefficients.....	269
TABLE 33	Correspondences of the self-assessments and external assessments: the three highest correlation coefficients.....	269

## SISÄLLYS

ABSTRACT

AVANT-PROPOS

ESIPUHE

KUVIOT JA TAULUKOT

1	JOHDANTO.....	15
1.1	Tutkimuksen tausta.....	15
1.2	Tutkimuksen toteutuksesta ja tarkoituksesta.....	16
1.3	Tutkimuksen rakenne.....	18
2	PUHUMINEN.....	20
2.1	Puhumista koskevia käsityksiä ja määrittelyjä.....	21
2.2	Kommunikatiivinen kompetenssi.....	22
2.3	Kielitaidon arvioinnin aikakaudet.....	28
2.4	Suullisen kielitaidon arviointi Suomessa.....	34
2.4.1	Yleiset kielitutkinnot.....	36
2.4.2	Valtionhallinnon kielitutkinnot.....	37
2.4.3	Ylioppilastutkinto.....	38
3	KIELITAIDON ITSEARVIOINTI.....	40
3.1	Kielitaidon itsearvioinnin taustaa.....	40
3.1.1	Oppija uudessa roolissa.....	40
3.1.2	Oppijan autonomia elinikäisessä oppimisessa.....	42
3.1.3	Itsearviointi muiden arviointitekniikoiden joukossa.....	45
3.1.4	Itsearvioinnin reflektiivisyys ja dialogisuus.....	48
3.1.5	Itsearvioinnista DIALANG-järjestelmässä.....	50
3.2	Kielitaidon itsearvioinnin käyttötarkoituksia.....	52
3.3	Itsearviointimateriaaleja ja -metodeja.....	54
3.4	Luotettavuudesta kielitaitoa koskevissa itsearviointitutkimuksissa.....	56
3.4.1	Itsearviointimenetelmät.....	59
3.4.2	Tietokoneen avulla tehtävä itsearviointi.....	59
3.4.3	Kokemus itsearvioinnista.....	61
3.4.4	Arvioitava taito.....	62
3.4.5	Oppijan taitotaso.....	66
3.4.6	Oppijan ikä ja sukupuoli.....	67
3.4.7	Oppimiskulttuuri ja koulutustausta.....	68
3.4.8	Kulttuuritausta.....	69
3.4.9	Oppijan persoonallisuus ja minäkuva.....	70
3.5	Itsearviointi ja asenteet.....	71
3.6	Itsearviointi opettamisen ja opetussuunnitelmien näkökulmasta.....	76
3.7	Vertaisarviointi.....	78
3.8	Yhteenvetoa itsearvioinnista.....	81

4	TUTKIMUSKYSYMYKSET, -AINEISTO JA -MENETELMÄT .....	83
4.1	Tutkimuskysymykset.....	83
4.2	Tutkimusmateriaalin valmistelu .....	86
4.2.1	Tutkimuksessa käytettyjen puhumistehtävien esittely .....	88
4.2.2	Yki-korpuksen tehtävien valinta (tehtävät 1 ja 2).....	92
4.2.3	Puhesuoritusten valinta Yki-korpuksesta .....	93
4.2.4	Kuunneltavien näytteiden valinta (tehtävät 1 ja 2) .....	94
4.2.5	Näytteiden kokoaminen korpukseksi .....	95
4.2.6	Lomakkeiden laatiminen .....	96
4.2.7	Arviointiasteikoista .....	99
4.2.8	Tehtävän 1 arviointiasteikon esittely.....	104
4.2.9	Tehtävän 2 arviointiasteikon esittely .....	107
4.2.10	Yki-korpuksen käsittely .....	109
4.2.11	Aineiston ja menettelytapojen toimivuuskokeilu.....	111
4.2.12	Tutkimukseen ilmoittautuminen ja kielistudiot .....	111
4.3	Tutkimusaineiston keruu .....	112
4.3.1	Osallistujien rekrytointi ja tietoa osallistujista .....	112
4.3.2	Työn tarkoituksen esitleminen ja toimintaohjeiden antaminen.....	118
4.4	Tutkimusaineiston käsittely .....	121
4.4.1	Maamerkkikasetin laatiminen .....	121
4.4.2	Ohjeistuksen antaminen arvioijille .....	121
4.4.3	Osallistujien puhesuoritusten arviointi.....	123
4.4.4	Aineiston koodaus .....	123
4.4.5	Palautteen antaminen osallistujille .....	124
5	AINEISTON ANALYYSIT JA TULOKSET .....	125
5.1	Itsearviointimittareiden luotettavuus .....	126
5.1.1	Paperilla tehty itsearviointi .....	126
5.1.2	Tehtävä 1: kuunneltaviin puhenäytteisiin perustuva itsearviointi.....	155
5.1.3	Tehtävä 2: kuunneltaviin puhenäytteisiin perustuva itsearviointi .....	170
5.1.4	Yhteenvetoa kolmen itsearviointimittarin luotettavuudesta.....	182
5.2	Ulkoisten arvioijien luotettavuus .....	190
5.2.1	Kolmen arvioijan luotettavuus, tehtävät 1 ja 2 .....	190
5.3	Itsearviointien ja ulkoisten arviointien väliset korrelaatiot .....	194
5.4	Tutkimukseen osallistuneiden laadullisia kommentteja .....	198
5.4.1	Itsearviointi paperilla -mittarin kommenttikategoriat ja osallistujien kommentit .....	199
5.4.2	Kuunneltavat puhenäytteet -mittareiden kommentti- kategoriat ja osallistujien kommentit .....	207

6	LUOTETTAVUUSTARKASTELU JA JOHTOPÄÄTÖKSET .....	225
6.1	Tutkimusprosessin tarkastelua .....	225
6.2	Vastauksia tutkimuskysymyksiin .....	226
6.3	Tutkimuksessa kehitettyjen itsearviointivälineiden laadun ja hyödyllisyyden tarkastelua Bachmanin ja Palmerin mallin pohjalta .....	240
6.3.1	Reliaabelius .....	242
6.3.2	Käsitevalidius .....	245
6.3.3	Autenttisuus.....	248
6.3.4	Vuorovaikutteisuus .....	249
6.3.5	Vaikuttavuus .....	250
6.3.6	Käytännöllisyys .....	252
6.4	Tutkimuksen rajoituksia .....	254
6.4.1	Tutkimusaineiston hankintatapa .....	254
6.4.2	Väsymyksen vaikutus oppijoiden suoritukseen .....	256
6.4.3	Oppijoiden puhesuoritusten arvioijat .....	257
6.4.4	Kuunneltavien näytteiden tasot .....	258
6.4.5	Tutkimukseen osallistuneet oppijat.....	259
6.5	Mahdollisia jatkotutkimusaiheita.....	259
6.6	Tutkimuksen tulosten hyödynnettävyys opetuksessa .....	261
	SUMMARY .....	266
	LÄHTEET JA LYHENTEET .....	272
	LIITTEET.....	289



# 1 JOHDANTO

## 1.1 Tutkimuksen tausta

Itsearviointia on pidetty 1990-luvulta lähtien yleisesti tärkeänä *elinikäisen oppimisen* edistämismuotona. Itsearviointi sijoittuu niin kutsuttujen vaihtoehtoisten arviointitapojen joukkoon. Niille on tyypillistä *oppijakeskeisyys*. Oppijakeskeinen opetus- ja oppimisenäkemyksessä vaikuttaa suuresti siihen, mitä oppijalta odotetaan. Oppijakeskeisessä opetusnäkemysessä opettaja luovuttaa osan tavanomaiseen opetustilanteeseen liittyvästä vallasta oppijan käsiin. Muutokseen sisältyy kuitenkin ennako-odotus, jonka mukaan oppija on halukas opiskelemaan tavalla, joka vaatii aktiivista sitoutumista ja uskoa omaan kykyihinsä. Itsearviointiprosessi etenee parhaimmillaan esittely- ja harjoittelujakson kautta siihen, että oppija vähitellen oppii ottamaan vastuun omasta oppimisestaan (*oppimaan oppiminen*) ja kokee uuden asemansa motivoivana. Prosessin lopullisena päämääränä on se, että oppijasta tulee *autonominen* ja *reflektiivinen* oman oppimisensa toteuttaja ja tarkkailija. Itsearvioinnin tarkoitus ei kuitenkaan ole opettajan työn turhaksi tekeminen, vaan ihannetilanteessa oppimistilanne rakentuu itsearviointia käytettäessä oppijan ja opettajan välisen jatkuvan *dialogin* pohjalle. Monet tahot ovat nostaneet itsearviointitaidon tärkeyden esille, kun on puhuttu muun muassa oppimisen ja työelämän muutoksista ihmisen elinkaaren aikana. Itsearvioinnin merkittävydestä kertoo jotain myös se, että itsearviointitaito mainitaan erikseen yhtenä oppimistavoitteena suomalaisen peruskoulun ja lukion opetussuunnitelmien perusteissa (POPS 2004; LOPS 2003). Myös suomalaista elinkeinoelämää edustavat tahot painottavat itsearviointitaidon tärkeyttä (EK 2011, 32).

Itsearvioinnissa toteutuu motivaatiota edistävän oppimisen ja osaamisen arviointi. Itsearviointiin sisältyvä itseohjautuvuus näyttääkin kuuluvan erottamattomana osana tulevaisuuden työelämän osaamisvaatimuksiin. Muita työelämässä tarvittavia taitoja ovat esimerkiksi vuorovaikutus- ja yhteistyötaidot, monipuolinen kielitaito sekä kulttuuriosaamisen ja oman oppimisen suunnitte-

lu ja arviointi. Riittävän hyvän suullisen kielitaidon merkitys painottuu entisestään muun muassa siksi, että ekologinen ajattelu näyttää tulleen jäädäkseen työelämän käytänteisiin ja toimintatapoihin. Kustannuspaineiden kasvaessa monilla aloilla on alettu suosia mahdollisuuksien mukaan esimerkiksi videoneuvotteluita pitkien ulkomaille suuntautuvien kokousmatkojen sijaan. Sujuvan suullisen kielitaidon hallinta onkin välttämätöntä esimerkiksi videoneuvottelujen kaltaisissa tilanteissa, joissa viesti tulee saada perille rajallisessa ajassa mahdollisimman ymmärrettävästi ja yksiselitteisesti. Itseohjautuvuutta tarvitaan kielitaidon itsearviointiin lisäksi yleisemminkin tulevaisuuden työelämässä, jossa ennusteiden mukaan korostuvat työtehtävien ”nuotittomuus”, jatkuva oppiminen, vaihtoehtoisten toimintatapojen etsiminen ja kyky tarkastella tietoa eri näkökulmista (EK 2011, 5–18).

Tässä tutkimuksessa itsearviointi kohdistuu kielitaitoon. Edellä mainitut itsearviointiin liittyvät yleiset lähtöoletukset, ominaisuudet ja tavoitteet pätevät myös kielitaidon itsearvioinnissa.

## 1.2 Tutkimuksen toteutuksesta ja tarkoituksesta

Tämä tutkimus käsittelee vieraalla kielellä puhumista ja oppijoiden suorittamaa oman vieraskielisen puheensa laadun itsearviointia. Tutkimus kohdistuu ranskan kieleen, mutta tutkimuksen tulokset ja työssä kehitetty itsearviointimenetelmä ovat sovellettavissa mihin tahansa kieleen. Tarkemmin sanottuna tutkimus keskittyy ranskan kielen aikuisoppijoiden suullisen kielitaidon itsearviointiin sekä itsearvioinnin ja ulkoisen arvioinnin vastaavuuteen. Empiirisessä tutkimuksessa kokeillaan käytännössä uutta suullisen kielitaidon arviointimenetelmää, jolla pyritään helpottamaan ja konkretisoimaan oppijan itsearviointia kuunneltavien puhe-suoritusnäytteiden avulla. Itsearvioinnin onnistumista verrataan sekä ulkopuolisten arvioijien tekemiin arvioihin oppijoiden puhe-suorituksista että oppijoiden perinteisemmällä tavalla eli paperilla suorittamaan itsearviointiin. Tutkimusaineisto koostuu kahdesta ranskan kielen puhumistehävästä (tehtävät 1 ja 2) sekä neljästä seuraavasta eri itsearviointitavasta:

1. Yleisitsearviointi paperilla
2. Tilannekohtainen itsearviointi paperilla
3. Kuunneltuihin puhenäytteisiin perustuva itsearviointi (tehtävä 1)
4. Kuunneltuihin puhenäytteisiin perustuva itsearviointi (tehtävä 2)

Tutkimuksessa hyödynnetään valmiita standardoituja kielitaidon arvioinnin välineitä. Käytetyt kaksi puhetehtävää ovat Yleisten kielitutkintojen puhumisen osakokeen todellisia tutkintotehtäviä, ja tutkimuksessa käytetyt arviointias-teikot pohjautuvat Euroopan neuvoston Eurooppalaisen viitekehyksen (EVK 2004, 93–139) ja laajan kansainvälisen DIALANG-järjestelmän kielitaitoa kuvaaviin asteikkoihin ja itsearviointiväittämiin. Paperilla olevat itsearviointiväit-

tämät on laadittu 2000-luvun alussa DIALANG-hankkeen kehittämisvaiheessa Jyväskylän yliopiston Soveltavan kielentutkimuksen keskuksessa, jossa tutkimuksen tekijä toimi tuolloin kyseisen hankkeen itsearviointiryhmän jäsenenä.

Aineistonkeruuprosessin aluksi oppijat sijoittivat oman suullisen ranskan kielen taitonsa yhdelle Eurooppalaisen viitekehyksen kuudesta yleistaitotasosta A1 – C2. Tämän jälkeen oppijat vastasivat lomakkeen 42:een yksittäistä kielenkäyttötilannetta kuvaavaan itsearviointiväittämään, joista kunkin kohdalla he valitsivat joko ”kyllä” tai ”ei” sen mukaan, sopiko väittämän sisältö heidän taitoonsa puhua ranskan kieltä vai ei.

Tutkimuksessa käytetty uusi itsearviointimenetelmä perustuu äänitteeltä kuunneltavien puhenäytteiden ja oman puhe-suorituksen vertaamiseen. Oppijat suorittivat ensin kielistudiossa kaksi erilaista puhumistehtävää, jotka tallennettiin. Tämän jälkeen oppijat kuuntelivat oman suorituksensa, minkä jälkeen he arvioivat, oliko heidän suorituksensa puhumistehtävässä parempi (= kyllä) vai huonompi tai yhtä hyvä (= ei) kuin nauhalta yksitellen esitetyt puhenäytteet, joissa erilaiset ja eritasoiset ranskan oppijat suorittivat saman puhumistehtävän.

Aineiston keruun aikana tutkimukseen osallistuneilta oppijoilta kerättiin monenlaista palautetta. He antoivat palautetta paperilla suoritettua itsearvioinnissa käytetyistä itsearviointiväittämisistä, itsearvioinnista yleensäkin sekä siitä, oliko heillä jossain kohdin itsearvioinnin aikana mahdollisesti vaikeuksia vastata paperilla tehdyn itsearvioinnin väittämiin. Jos oppijat vastasivat vaikeuksia ilmenneen, heitä pyydettiin tarkentamaan, millaisia nämä vaikeudet olivat. Lisäksi heitä kehoitettiin merkitsemään väittämän numero, jos he pitivät jotain väittämää tavallista hankalampana.

Paperilla tehdyn yksityiskohtaisen itsearvioinnin väittämiä koskevan palautteen lisäksi oppijoita pyydettiin kommentoimaan, miksi he valitsivat ”kyllä” tai ”ei”-vaihtoehdon kunkin kuunneltavan puhenäytteen kohdalla. Oppijoilla oli tässä yhteydessä mahdollisuus ilmaista myös se, pitivätkö he vertailua vaikeana kuunneltuaan jonkin tietyn puhenäytteen. Tutkimus sisältää siis tietoa myös siitä, mihin kielen piirteisiin oppijat kiinnittivät huomiota kuunnellessaan nauhoitettuja puhenäytteitä ja millaiset seikat tai kielen piirteet mahdollisesti vaikuttivat siihen, että valinta vastausvaihtoehtojen välillä tuntui hankalalta näytteitä kuunneltaessa.

Tutkimuksen ensisijaisena tarkoituksena on tuottaa tietoa uudella tavalla toteutetun itsearvioinnin luotettavuudesta ja käyttökelpoisuudesta vertaamalla sen tuloksia ulkoiseen arviointiin sekä perinteisiin itsearviointitapoihin. Tutkimuksen tarkoitus ei kuitenkaan ole verrata erilaisia itsearviointimenetelmiä eikä esimerkiksi asettaa niitä paremmuusjärjestykseen. Käytettyjen itsearviointimittareiden vertaaminen toisiinsa perustuu ainoastaan siihen, että vertailun avulla pyritään selvittämään mittareiden ja niiden sisältämien osioiden toimivuutta ja luotettavuutta. Tutkimuksen tarkoituksena on kehittää sellainen itsearviointimenetelmä, joka olisi mahdollisesti perustellumpi ja motivoivampi oppijoiden näkökulmasta kuin perinteinen paperilla tehtävä itsearviointi. Kyse on siis pikemminkin uudenlaisen lähestymistavan ottamisesta itsearviointiin. Lisäksi tutkimus antaa tietoa itsearviointiväittämiä koskevan palautteen antamisesta, oppijoiden

käsityksistä itsearvioinnista ja heidän omasta ranskan kielen taidostaan sekä siitä, mitä he pitävät suullisen kielitaidon tärkeimpinä piirteinä.

### 1.3 Tutkimuksen rakenne

Johdanto-osan jälkeinen luku 2 käsittelee puhumista. Tämän tutkimuksen tarkoituksena ei ole perehtyä puhumisen koko kenttään, vaan olen valinnut käsiteltäväksi joitakin puhumiseen liittyviä peruskäsitteitä. Oman kronologisesti etenevän osansa lukuun 2 muodostaa puhumisen arviointia käsittelevä osa. Tässä osassa esitellään lyhyesti puhumisen arviointia myös suomalaisessa kontekstissa kolmen suomalaisen kielitutkintojärjestelmän avulla.

Luku 3 yhdessä luvun 2 kanssa muodostaa tutkimuksen teoreettisen viitekehysten. Luvussa 3 esitellään itsearviointiin liittyviä käsitteitä, esimerkiksi oppijakeskeisyyttä ja reflektiivisyyttä sekä itsearviointimateriaaleja ja -tutkimuksia. Itsearvioinnin luotettavuus on luvun 3 avainsana, ja siten itsearviointitutkimusten tuloksia tarkastellaankin itsearvioinnin luotettavuuteen vaikuttavien tekijöiden valossa. Luvussa käsitellään myös itsearviointiin liittyviä asenteita sekä luodaan katsaus opettamiseen ja suomalaisen koulun opetussuunnitelmiin itsearvioinnin näkökulmasta. Luvun 3 lopussa esitellään lyhyesti itsearviointiin kuuluva vertaisarviointi.

Luvussa 4 esitellään tutkimuksen tutkimuskysymykset, joita on kuusi kappaletta. Tämän jälkeen käydään ensin läpi tähän empiriiseen tutkimukseen liittyvät tutkimusprosessin käytännön työvaiheet. Luvussa esitellään siis yksityiskohtaisesti tutkimusmateriaalin valmisteluun, tutkimusaineiston keruuseen, arviointiin ja käsittelyyn liittyviä asioita.

Lukuun 5 sisältyvät aineiston analyysit ja tulokset. Luvussa tarkastellaan tilastollisten menetelmien avulla tutkimuksessa käytetyn kolmen eri itsearviointimittarin toimivuutta kokonaisuutena, yksittäisten itsearviointiväittämien ja puhenäytteiden toimivuutta sekä ulkoisten arvioijien luotettavuutta. Tilastomenetelmien tuottamien analyysien lisäksi luvussa tarkastellaan laadullisesti oppijoiden kommentteja paperilla tehdyn itsearvioinnin itsearviointiväittämistä sekä kuunneltavista puhenäytteistä. Puhenäytteiden kommentteja analysoidaan itsearviointilomakkeessa olevan ”vaikea päättää” -sarakkeen rastien määrän pohjalta. Laadulliseen tarkasteluun on otettu kolme eniten ja vähiten ”vaikea päättää” -rasteja kerännyttä puhenäytettä kummastakin tehtävästä.

Luvussa 6 pohditaan tutkimusprosessin onnistumista, esitellään vastaukset tutkimuskysymyksiin ja tarkastellaan tutkimuksessa kehitettyjen itsearviointivälineiden laatua ja hyödyllisyyttä Bachmanin ja Palmerin (1996, 17–42) mallin pohjalta. Luvun loppupuolella otetaan esille tutkimukseen liittyviä rajoituksia ja ehdotetaan mahdollisia jatkotutkimusaiheita. Lopuksi tarkastellaan tutkimuksen sijoittumista muiden vastaavien itsearviointitutkimusten joukkoon. Tutkimuksen loppu sisältää myös pohdintaa tutkimuksessa kokeillun uuden itsearviointimenetelmän soveltuvuudesta opetukseen ja syistä, jotka

ovat voineet vaikuttaa menetelmän toimimiseen. Tässä osassa sivutaan myös itsearviointiin liittyvien asenteiden tärkeyttä.

## 2 PUHUMINEN

Tämän tutkimuksen pääasiallisena tarkoituksena on kokeilla uutta kuunneltavien puhenäytteiden pohjalle rakentuvaa itsearviointimenetelmää ja verrata sitä perinteisempään itsearviointimenetelmään ja ulkoiseen arviointiin. Itsearviointia tehtiin tässä tutkimuksessa suullisen kielitaidon näkökulmasta, ja siksi tässä luvussa käsitellään puhumista ja sen arviointia. Toisen ja vieraan kielen opetus, tutkimus ja arviointi ovat osittain erillisiä kokonaisuuksia, mutta toisaalta niiden kaikkien kiinnostuksen kohteena on kielen oppijan jonkin muun kielen kuin hänen äidinkiellensä oppiminen ja opitun taidon arvioiminen. Tämän vuoksi suullista kielitaitoa ja siihen liittyviä näkemyksiä, määrittelyjä ja malleja käsitellään tässä luvussa sekä toisen että vieraan kielen tutkimustulosten valossa ilman, että tekstissä välttämättä erikseen mainitaan kummalta puolelta kulloinkin tarkasteltava asia on peräisin. Luvussa saatetaan käyttää synonyymisesti käsitteitä ”puhuminen” ja ”suullinen kielitaito”, ja niillä tarkoitetaan molemmilla puhumista.

Luvussa 2.1 esitellään aluksi puhumista koskevia länsimaiseen kontekstiin liittyviä käsityksiä ja määrittelyjä eri aikakausilta. Luvussa 2.2 tarkastellaan kommunikatiivisen kompetenssin määrittelyä ja siihen liittyviä malleja. Luku 2.3 sisältää kielitaidon arvioinnin tarkastelua eri näkökulmista eri aikakausille tyypillisten arviointitendenssien valossa. Itsearviointia ei tarkastella tässä yhteydessä, vaan sitä käsitellään pääluvussa kolme. Luvun 2.3 lopussa luodaan vielä lyhyet katsaukset suullisen kielitaidon arviointiin ja puhumisen opettamiseen suomalaisessa kontekstissa.

On huomattava, että tässä tutkimuksessa ei pyritä käsittelemään puhumista tyhjentävästi, koska tämän tutkimuksen pääfokus ei ole puheen tai puhumisen tutkimus. Siten tämän luvun tavoitteena on luoda tiivis katsaus siihen, miten puhumista on määritelty eri aikoina ja miten sen arviointi on muuttunut.

## 2.1 Puhumista koskevia käsityksiä ja määrittelyjä

Suullisen kielitaidon arviointiin ja määrittelyyn ovat vaikuttaneet monet eri tieteet ja suuntaukset. Yksi vahvimista vaikuttajista on psykologinen suuntaus, joka keskittyi 1970-luvulle saakka ihmisen erilaisten taitojen rakenteen selvittämiseen (Lussier & Turner 1995, 4; Spolsky 1995, 10). Muita aloja, joissa kieli ja sen käyttö nähtiin sosiaalisena ilmiönä ja jotka vaikuttivat kieltä koskeneiden käsitysten muuttumiseen, olivat mm. sosiologia, kulttuuriantropologia, etnografia ja filosofia (Huhta 1993, 83). Myös kielitiede on antanut oman tärkeän lisänsä kielitaidon arvioinnille muun muassa Chomskyn (1965), Hallidayn (1970), Austinin (1962) ja Searlen (1969) tutkimusten ansiosta.

1960-luvulla Lado (1961, 239) määritteli vieraan kielen puhumiseen liittyvää taitoa seuraavasti:

*The ability to speak a foreign language is without doubt the most highly prized language skill, and rightly so, because he who can speak a language well can also understand it and can learn to read it with relative ease unless it happens to be a language like Chinese or Japanese, whose writing systems constitute major learning problems of their own.*

Kuten yllä olevasta lainauksesta ilmenee, Lado tuntuu arvostavan paljon suullista kielitaitoa. Samalla tekstissä voi havaita tietynlaisia käsityksiä, joiden yhteyksiä nykyisin ei ehkä enää pidettäisi automaattisina. Tällainen seikka on esimerkiksi se, että henkilö, joka hallitsee jotakin hänelle vierasta kieltä suullisesti hyvin, oppisi suhteellisen helposti myös lukemaan tätä samaa vierasta kieltä. Toki Ladon ajatus sopii mitä luultavimmin suurimpaan osaan vieraiden kielten oppijoista. Aina on kuitenkin ollut ihmisiä, jotka ovat oppineet puhumaan vierasta kieltä hieman epätavallisissa yhteyksissä tai ainakin sellaisissa olosuhteissa, joissa muun kuin suullisen kielitaidon hallinnalla ei ole ollut suurta merkitystä (esimerkiksi vieraassa kulttuurissa väliaikaisesti työssä olevat henkilöt). Lado, joka julkaisi vuonna 1961 uraa uurtavan teoksen *Language Testing*, määrittelee kielitaidon koostuvan neljästä erillisestä taidosta, jotka olivat puhuminen, ymmärtäminen, lukeminen ja kirjoittaminen. Näiden neljän kielitaidon osataidon lisäksi Lado mainitsee viidentenä kääntämistaidon, jota hänen mukaansa tulee testata erillisenä taitona, ei kuten yleistä kielen hallintaa (1961, 26). Lado esittelee kielen kommunikaatiojärjestelmänä, jossa vaikuttavat itse kieli, kulttuuri ja puhujat. Lado korostaa myös kulttuurisen aspektin olemassaoloa kielen opiskelussa.

Yksi puhutun kielen ja sen tehtävien lähestymistapa on tarkastella sitä, millaisia toimintoja testattavien odotetaan tekevän, kun he sanovat jotain (Luoma 2004, 32). Tämän lähestymistavan taustalla on kielifilosofinen näkökulma, johon liittyvät yksittäiset kielelliset suoritukset eli puheaktit (Austin 1962). Amerikkalaisen filosofi Searlen (1969) pidemmälle kehittämä puheaktiteoria (*Speech Act Theory*) sekä myöhemmin 1970-luvulla Gricen maksimit - käsite (Grice 1975) edustavat tätä näkökulmaa. Puheaktiteoria ja Gricen mak-

siimit keskittyvät puhujien väliseen yhteistyöhön ja siihen ajatukseen, että kun sanotaan jotain, samalla tehdään jotain. Englantilainen filosofi Austin alkoi kiinnostua siitä, miten sanoilla tehdään asioita ja millaisten ehtojen on toteuduttava tilanteessa, jotta tietyt sanat lausumalla teko onnistuisi. Searle määrittelee puhumisen seuraavasti: ”*puhua tarkoittaa sitä, että suoritetaan toimintoja sääntöjen mukaan*” (Searle 1969, 59). Puheakteina pidetään toimintoja, jotka liittyvät kielen avulla tehtäviin toiminnan lajeihin, joista on olemassa erilaisia luokituksia. Puheaktiteorian mukaan yleisiä puhumalla suoritettavia toiminnan lajeja voi kuitenkin olla vain rajallinen määrä. Yksi mahdollinen tapa on luokitella toiminnan lajit ilmausten mukaan esimerkiksi seuraaviin luokkiin: esitys, ohjaus, sitoumus, mielenilmaus, julistus (Kieli ja sen kieliopit 1994, 89).

Amerikkalaisen filosofi H. P. Griceen maksiimit puolestaan lähtevät siitä, että vuorovaikutus on mahdollista. Siten Griceen mukaan keskustelun perustana onkin keskustelijoiden välinen yhteistoiminnan periaate. Maksiimit-sanaa Grice käyttää sopivuussäännöistä; hänen mukaansa keskustelijat tuntuvat vuorovaikutuksessa ollessaan olettaen, että keskustelun repliikit noudattavat tiettyjä julki lausumattomia sopivuussääntöjä (Kieli ja sen kieliopit 1994, 90).

Suullisen kielitaidon pohjaa eli puhumista ovat tutkineet ja määritelleet eri tavoin monet tutkijat. Muun muassa Bygate (1987, 3) erottaa selkeästi toisistaan kieleen liittyvän tiedon (*knowledge*) ja sen käyttämiseen liittyvän taidon (*skill*). Hän määrittelee puhumisen olevan näiden kahden yhdistelmä. Bygaten mukaan kielen käyttämiseen liittyviä taitoja on kahdenlaisia: motoriseen toimintaan ja havaitsemiseen liittyviä taitoja (*motor-perceptive skills*) ja vuorovaikutuksellisia taitoja (*interaction skills*). Ensimmäisillä taidoilla tarkoitetaan kielenkäytön kontekstiin liittymättömiä taitoja, joita ovat esimerkiksi havainnoiminen, muistaminen ja kielen äänteiden ja rakenteiden artikuloiminen oikeassa järjestyksessä. Jälkimmäiset taidot taas sisältävät kyvyn käyttää kieltä tiettyjen tarpeiden tyydyttämiseksi (Bygate 1987, 5, 7).

## 2.2 Kommunikatiivinen kompetenssi

Ennen kuin tarkastelen kieliin liittyvän kommunikatiivisen kompetenssin ominaispiirteitä, pohdin lyhyesti kompetenssi-käsitettä yleisellä tasolla. Eräät näkemykset asettavat vastakkain tiedon (*savoirs*) ja kompetenssin (*compétences*). Näiden näkemysten mukaan tieto olisi yhteydessä kulttuuriin ja ihmiskunnan perintöön, kun taas kompetenssi rajoittuisi hyötynäkökulmaa korostaviin suuntauksiin, joiden puolestaan nähdään liittyvän uusliberalistiseen individualismiin. Samojen näkemysten mukaan monet kansainväliset järjestöt, kuten esimerkiksi OECD, suosivat kompetenssilähtöistä ei-tietopohjaista suuntautumista, koska tietopohjaisen suuntautumisen ei katsota olevan välittömästi hyödyllinen taloudelliselta tai sosiaaliselta kannalta (esim. Hirt 2010; Laval ym. 2011).

Kompetenssi-käsitteen määrittelyn ympärillä on joidenkin mielestä vallinnut niin suuri epämääräisyys, että performanssin arviointi tai tietyn tehtävän toteuttaminen on saanut huomattavasti enemmän huomiota kuin vaikeasti



määriteltävä kompetenssi-käsite (Cuq 2003, 212). 1970-luvun puolivälistä lähtien alettiin korostaa sitä, että kielenoppijoiden tulisi oppia kieltä, joista olisi todellista hyötyä jokapäiväisessä elämässä. Tässä vaiheessa korostuivat toiminnot (*functions*), eli mitä oppijoiden tulisi pystyä tekemään, ja käsitteet (*notions*), eli niiden sanojen ja fraasien oppiminen, joita oppijat tarvitsevat pystyäkseen keskustelemaan tietyistä aiheista. Näkemys merkitsi suurta muutosta aiempaan kielioppipainotteiseen kielenopetukseen, ja tästä näkemyksestä seurasiakin kommunikatiivisen kielenopetustavan nopea yleistyminen 1980-luvulla. (Luoma 2004, 33.) Uudet ajatukset eivät olleet peräisin pelkästään kielitieteen, vaan myös sosiologian, kulttuuriantropologian, etnografian ja filosofian piiristä. Ne näkivät kielen ja sen käytön sosiaalisena ilmiönä (Huhta 1993, 83).

Kommunikatiivinen kompetenssi korostaa kielenkäyttäjän roolia ja kielen käyttämistä kommunikointiin (Luoma 2004, 100). Myös Widdowsonin käsityksen mukaan (1983) kaikenlainen yksilön eristetyssä tilanteessa tekemän suoriutumisen (*performance*) ulkopuolelle ulottuva kielenhallinta edustaa kommunikatiivista kykyä tai kompetenssia. Uuden kommunikatiivisen näkemyksen mukaan niin kielitaidossa kuin sen testaamisessakin korostui kielen funktionaalisuus ja kielenkäyttötarkoitusten kuvaamista pidettiin tärkeänä (Huhta 1993, 83). Kommunikatiivinen kompetenssi ja sen ero perinteiseen kielen hallintaan verrattuna ilmenee selkeästi Mothen (1990,133) määrittelystä, jossa asiaa lähestytään arvioinnin näkökulmasta. Hänen mukaansa kommunikatiivisen kompetenssin arvioinnissa on kyse siitä, mitä oppija osaa tehdä sillä, mitä hän osaa ("ce que l'apprenant sait faire avec ce qu'il sait").

Chomsky (1965) (Huhdan 1993, 83 mukaan) määrittelee kielitaidon olevan lingvistinen kompetenssi. Hänen kielitieteelliseen ajatteluunsa 1960-luvulta lähtien suuresti vaikuttanut näkemyksensä kielestä on kaksijakoinen: toisaalta on olemassa *kompetenssi* eli synnynnäinen kielikyky ja toisaalta varsinainen puhe-suoritus eli *performanssi*.

Hymes (1972) otti kommunikatiivisen kompetenssin käsitteen käyttöön 1970-luvulla sen jälkeen, kun kielioppipainotteiset kieleen liittyvät teoriat, jotka analysoivat kieltä riippumattomana sen käyttäjistä, alkoivat tuntua aikansa eläneiltä. Hymes korostaa kielenkäytön sosiaalisen aspektin tärkeyttä. Hänen tärkeä havaintonsa on, että kielenkäyttökyky (*ability to use language*) pitäisi sisällyttää kielitiedon ohella kommunikatiiviseen kompetenssiin (Hymes 1972, 282).

Aiemmin mainitun puheaktiteorian vaikutus voidaan jäljittää mm. Wilkinsiin (1973 ja 1976) ja *Threshold level* -julkaisuun (1975 ja 1990), ja siten puheaktiteoriaa voidaankin pitää tietyllä tapaa kielen kommunikatiivisen lähestymistavan edelläkävijänä (Figueras 2001, 16). Sama koskee Euroopan neuvoston vuonna 1976 julkaisemaa ranskankielistä kielen funktionaalisen lähestymistavan mukaan laadittua *Un niveau-seuil* -julkaisua. Sen johdanto-osassa viitataan englannin kieltä koskevaan, vuotta aiemmin julkaistuun *Threshold Level* -julkaisuun ja todetaan näiden kahden julkaisun olevan erilaisia kolmessa tärkeässä kohdassa: ranskankielisessä teoksessa otetaan laajemmin huomioon puheaktit (*actes de paroles*), sen todetaan olevan pikemminkin kommunikaatiokielioppi (*grammaire de communication*) kuin kielitieteellinen kielioppi, ja sen lähtö-

kohtana on kielenkäyttö perinteisen kielen muotoihin perustuvan tarkastelun sijaan (Coste ym. 1976, v).

Euroopan neuvoston kehittämistyöllä oli suuri vaikutus siihen, että eri maiden kielten opetusta koskevia opinto-ohjelmia alettiin määritellä mm. kielen funktioiden ja tiettyihin tilanteisiin liittyvän kommunikaatiokontekstin ja kommunikaatiotaitojen näkökulmasta. Euroopan neuvoston kehitteillä olleen Viitekehysten taustaideana oli alusta alkaen aikuisten kielen oppimisen ja opettamisen rinnakkainen esille tuominen sekä sen korostaminen, että opetuksen tulee perustua oppijan tarpeisiin ja olla toimintasuuntautunutta. Myöhemmin huomattiin, että aivan samat asiat, joihin kiinnitettiin huomiota aikuiskoulutuksen puolella, sopivat myös koulujen tarpeisiin. Euroopan neuvoston Viitekehysten tausta-ajatusten sanotaankin vaikuttaneen esimerkiksi Suomen koulujen opetussuunnitelmien perusteisiin jo 1960-luvun lopulta alkaen (Huttunen & Jaakkola 2003).

Kommunikatiiviseen kompetenssiin eli kielelliseen viestintätaitoon liittyviä määritelmiä ja malleja on kehitelty Hymesin jälkeen vuosikymmenien aikana lukuisia. Näistä tärkeimpinä voidaan mainita Canalen ja Swainin malli (1980), Canalen edellistä täydentävä malli (1983) sekä Bachmanin (1990) ja Bachmanin ja Palmerin (1996) kehittämät mallit.

Canalen ja Swainin (1980) viestinnälliseen opettamiseen ja kielitaidon testauksen teoriaan liittyvässä mallissa laadittiin sarja ohjeita ja periaatteita siitä, kuinka kommunikatiivisia lähestymistapoja voidaan soveltaa toisen kielen opettamiseen ja arviointiin liittyvien instrumenttien kehittämiseen. Mallissa kieltä lähestytään joko kieliopillisesta tai kielen funktioiden näkökulmasta. Ensimmäiseen sisältyvät esim. fonologiset, morfologiset ja syntaktiset muodot ja toiseen erilaiset toiminnot, joihin kielen tiettyä kieliopillista muotoa käytetään: esimerkiksi kutsuminen tai anteeksipyyttäminen. Malli koostuu kolmesta pääkomponentista, jotka ovat kieliopillinen kompetenssi, sosiolingvistinen kompetenssi ja strateginen kompetenssi. Mallissa määritelty sosiolingvistinen kompetenssi liittyy suullisen kielenkäytön sääntöjen tuntemiseen ja niiden tilannekohtaiseen oikeaan käyttöön sekä diskurssisääntöjen tuntemiseen ja käyttöön. Mallissa ehdotetaan uudeksi kielen komponentiksi strategista kompetenssia (*strategic competence*), joka liittyy siihen, miten kielenkäyttäjät selviytyvät erilaisista kommunikaatiokatkoksista ja miten he toimivat todellisissa kielenkäyttötilanteissa pitääkseen kommunikaatiokanavan avoimena. Figueras (2001, 20) toteaa Canalen ja Swainin mallista, että se ei onnistu määrittelemään kielenkäyttökykyä (*ability to use*) ts. sitä, mikä saa puhujassa aikaan sen, että tämä pystyy tuottamaan ja ymmärtämään kieltä. Canalen vuonna 1983 julkaisema kommunikatiivista kompetenssia kuvaava malli täydentää Canalen ja Swainin mallia lisäämällä siihen diskursiivisen kompetenssin komponentin.

Bachmanin (1990) malli ja Bachmanin ja Palmerin (1996) 1990-luvulla edelleen kehittämä malli veivät eteenpäin kommunikatiivista kompetenssia määritteleviä malleja. Bachmanin (1990) kielitaidon testaukseen tarkoitettu kielikompetenssia määrittelevä malli koostui aluksi kahdesta pääkomponentista, jotka olivat organisatorinen kompetenssi ja pragmaattinen kompetenssi. Organisatoriseen kompetenssiin sijoittuvat kieliopillinen ja tekstuaalinen kompe-

tenssi ja pragmaattiseen kompetenssiin sosiolingvistinen ja illokutiivinen kompetenssi, joista jälkimmäinen esiteltiin uutena kompetenssina. Bachman kehitti edelleen vuoden 1990 mallia ja päätyikin korvaamaan kompetenssi-sanalla ”tieto”. Myös mallin sisältö muuttui tieto-sanalla käyttöön oton lisäksi siten, että pragmaattisen tiedon alakategorioita tuli kolme entisen kahden (illokutiivinen ja sosiolingvistinen) sijaan. Pragmaattisen tiedon alakategoriat olivat propositionaalinen tieto (uusi, viittaa sanastoon), funktionaalinen tieto (aikaisemmin illokutiivinen kompetenssi; viittaa kykyyn ymmärtää ja tuottaa ilmauksia, joilla on tietty tarkoitus) sekä jo alkuperäisessä mallissa mukana ollut sosiolingvistinen tieto. Bachmanin mallin soveltuminen erityisesti kielitaidon testaukseen perustuu siihen, että kielen rakenteen selittämisen lisäksi sen avulla pyritään selvittämään, miten kielen kompetenssi toteutuu kielen käytössä ja mitkä tekijät vaikuttavat henkilön menestymiseen testissä (mallin analyysistä ks. tarkemmin myös Huhta 1993). Vuodelta 1996 peräisin olevassa Bachmanin ja Palmerin mallissa on edelleen mukana aiemmin esitellyt neljä kielen komponenttia sijoittuneena seuraavasti kahden yläkategorian alle:

#### **Kielitieto**

1. Organisatorinen tieto:
  - Tekstuaalinen tieto
  - Kieliopillinen tieto
  
2. Pragmaattinen tieto:
  - Funktionaalinen tieto
  - Sosiolingvistinen tieto

#### **Strateginen kompetenssi (metakognitiiviset komponentit)**

- Arviointi
- Päämäärän asettaminen
- Suunnittelu

Muita kommunikatiivista kompetenssia määritteleviä malleja ovat kehittäneet 1980-luvulla muun muassa Moirand (1982) ja Savignon (1983) sekä 1990-luvulla Celce-Murcia, Dörnyei ja Thurrell (1995) ja Schachter (1990).

Euroopan neuvoston Viitekehyksessä suullinen kielitaito asettuu kommunikatiivisten kielellisten toimintojen alle. Nämä toiminnot jaetaan tuottamis- ja vuorovaikutteisiin toimintoihin, joihin liittyvät kiinteästi tuottamis- ja vuorovaikutukselliset strategiat. Monet viestintätilanteet perustuvat vuorovaikutukseen (esim. keskustelu) tai viestien välittämiseen (esim. kahden tai useamman henkilön välillä, jotka eivät pysty olemaan suoraan keskenään yhteydessä). Samalla muistutetaan myös, että useimmat viestintätilanteet ovat erilaisten toimintatyyppien yhdistelmiä. (EVK 2004, 91.) Suullisen kielen tuottamiseen liittyviä Viitekehyksen taitotasoasteikkoja ovat seuraavat: puhumista koskevat yleiset taitotasot, kokemuksen kuvaaminen, asian esittäminen, julkinen tiedottaminen ja yleisölle puhuminen (EVK 2004, 93–95).

Strategia-sanan määritellään tarkoittavan mahdollisimman tehokkaan toimintatavan valitsemista (EVK 2004, 92). Kyse on sisäisten voimavarojen aktiivisista siten, että henkilön suorituskyky saataisiin vastaamaan kulloisenkin tehtävän vaatimuksia (EVK 2004, 98). Tuottamisstrategioihin kuuluu monenlaisia kielen tuottamistilanteen aikana realisoituvia tehtäviä, jotka ryhmitellään seuraavasti: suunnittelu (esim. kohderyhmän huomioon ottaminen), toimeenpano (esim. kompensointi), arviointi (esim. onnistumisen monitorointi) ja korjaaminen (esim. omaehtoinen korjaaminen) (EVK 2004, 99). Tuottamisstrategioihin liittyviä taitotasosteikkoja on kolmenlaisia: suunnittelu, kompensointi sekä tarkkailu ja korjaaminen (EVK 2004, 100–101).

Eurooppalaisessa Viitekehysessä luetellaan suuri määrä suullisia vuorovaikutustoimintoja, kuten esimerkiksi asiointikeskustelut, vapaamuotoiset keskustelut ja neuvottelut. Suullisten vuorovaikutustoimintojen määrä näkyy myös taitotasosteikkojen lukumäärässä (9 kpl). Esimerkkeinä tyypillisistä vuorovaikutustoimintoihin liittyvistä asteikoista voidaan mainita virallisten ja vapaamuotoisten keskustelujen, neuvottelujen, tavoitteellisen yhteistyön ja tiedonvaihdon arviointiin tarkoitettut asteikot (EVK 2004, 111).

Vuorovaikutusstrategioista teoksessa todetaan, että vuorovaikutukseen kuuluu sekä vastaanottavaa (reseptiivistä) että tuottavaa (produktiivista) kielellistä toimintaa. Reseptiivisten ja produktiivisten strategioiden lisäksi on olemassa yksinomaan vuorovaikutukseen liittyviä strategioita, joiden tarkoituksena on ohjata vuorovaikutusprosessia (esim. suunnitteluun kuuluvana sen arviointi, mitä voidaan olettaa puhekuppanin tietävän). Puheenvuoron vaihtamisen strategioita hyödynnetään vuorovaikutustoiminnon aikana (esim. puheenvuoron ottaminen, avun pyytäminen). Vuorovaikutusstrategioihin liittyviä taitotasosteikkoja on kolme: puheenvuoron vaihtaminen, yhteistyö ja tarkennuksen pyytäminen (EVK 2004, 123–124.)

Viitekehysessä esitellään myös kielellistä viestintätaitoa ja sen komponentteja. Kielellinen viestintätaito koostuu kolmesta pääkomponentista, joita ovat kielellinen, sosiolingvistinen ja pragmaattinen komponentti (EVK 2004, 33). Kielelliseen kompetenssiin kuuluvat monenlaiset kieliin liittyvät tiedot ja taidot (esim. sanastolliset ja syntaktiset), mutta myös ne komponentit, joiden kuulumista kielelliseen kompetenssiin ei ehkä heti osata hahmottaa (esim. tiedon jäsentäminen ja saatavuus eli aktivoituminen, mieleen palauttaminen ja käytettävissä olo; EVK 2004, 33). Viitekehys ryhmittelee kielelliseen kompetenssiin sanastolliset, kieliopilliset, semanttiset ja fonologiset taidot sekä oikeinkirjoitustaidon ja kirjoituksen luontevan ääneenlukutaidon (EVK 2004, 156).

Merkitystä koskevat kysymykset ovat Viitekehysten mukaan viestinnässä keskeisiä. Niitä käsitelläänkin teoksessa monissa eri kohdissa. Kielellistä kompetenssia käsitellään muodon näkökulmasta, jolloin kielellisen viestintätaidon monet osatekijät erotetaan toisistaan. Erottelun käyttäminen edellyttää kuitenkin sitä, että viestintätaidon osatekijöihin luetaan kielen muotorakenteen tuntemus ja kyky käsitellä tätä muotorakennetta. Kiinnostavana esimerkkinä yhdestä tavasta kuvata kielellistä kompetenssia esitetään aiemmin mainittuun Kynnystaitotason ranskankieliseen versioon (*Un niveau-seuil*) sisältyvän ”kommunikatiivisen kieliopin” kuvaus. Tämä kuvaus eroaa rakenteeltaan huomatta-

vasti Euroopan neuvoston tunnetuista funktionaalis-notionaaliseen kielen tarkastelutapaan perustuvista taitotasokuvauksista (Waystage 1990, Threshold Level 1990 ja Vantage Level 1997). Näiden taitotasokuvausten tapa kuvata kieltä etenee merkityksestä muotoon. (EVK 2004, 164–165.)

Sosiolingvistinen kompetenssi koskee tietoja ja taitoja, joita tarvitaan yhteisön kielenkäytössä. Se liittyy sosiokulttuuristen ehtojen vaatimaan kielenkäyttöön ja on kiinteästi yhteydessä sosiaalisiin käytänteisiin, kuten esimerkiksi kohteliaisuussääntöihin ja erilaisten ryhmien välisiä suhteita sääteleviin normeihin. Sosiolingvistiseen kompetenssiin kuuluu aktiivisen tilanteeseen sopivan kielenkäytön hallinnan lisäksi myös kyky tunnistaa kielimuodon vaihtelua, joka liittyy kielenkäyttäjän taustaan (EVK 2004, 34, 168–170).

Pragmatiivinen kompetenssi säätelee kielellisten voimavarojen käyttöä tiettyihin tarkoituksiin ja perustuu kielenkäyttäjien keskinäisen vuorovaikutuksen skenaarioihin. Pragmaattiseen kompetenssiin kuuluu – ehkä odotuksenvastaisesti – diskurssin, merkityskoherenssin ja koheesion hallinta sekä tekstityyppeihin ja erilaisiin teksteihin liittyvien sävyjen (ironia ja parodia) tunnistaminen. Viitekehyksen mukaan tämän kompetenssin kehittymiseen vaikuttanevat suuresti henkilön kokeumukset ja se kulttuurinen ympäristö, johon henkilöllä on yhteyksiä (EVK 2004, 34). Viitekehyksessä pragmaattinen kompetenssi on yläkompetenssi, jonka alle sijoittuvat diskurssikompetenssi, funktionaalinen kompetenssi ja skeemakompetenssi (EVK 2004, 173).

Pohdittaessa erilaisten oppimistavoitteiden asettamista ja Viitekehyksen mahdollista roolia siinä Viitekehyksessä esitellään ajatusmalli, jonka mukaan kielellistä viestintätaitoa voidaan pitää kielenkäyttäjän monikielisenä ja monikulttuurisena kompetenssina, jonka kokonaisuus muodostuu äidinkielen erilaisista kielimuodoista ja yhden tai useamman vieraan kielen erilaisista kielimuodoista (EVK 2004, 189).

Vaikka kommunikatiivisten kompetenssien arviointi on vallannut laajalti alaa, jotkut tutkijat näkevät ongelmia tämän kompetenssin määrittelyssä ja arvioinnissa. Muun muassa Kahn (2012) toteaa, että kommunikatiivisten kompetenssien arviointiin liittyy yleinen ongelma: suurin osa oppijoista ja opettajista on edelleen sitä mieltä, että arvioida voidaan vain jotakin sellaista, mitä on opetettu (Kahn 2012). Kompetenssien arviointi taas ei liity millään tavalla arviointia edeltävään opetukseen, mikä nostaakin esille tietojen ja kompetenssien arvioinnin välisen eron. Jotkut tutkijat ovat sitä mieltä, että arvioinnissa piilee vaara. Esimerkiksi Crahay (2006) toteaa, että jotkut oppilaat näyttävät pystyvän toisia paremmin hyödyntämään käytettävissä olevat resurssinsa. Crahay arvelee luultavasti oikeasuuntaisesti, että hyödyntämiskykyyn vaikuttaa esimerkiksi tottumus kirjoitettuun tekstiin, erilaisiin älyllisiin käsitteisiin ja resursseihin. Crahay varoittaakin, että kompetenssien arviointi voi entisestään lisätä oppijoiden välistä epätasa-arvoa, mikä puolestaan johtaa siihen, että koulutus osallistuisi elitististen käytäntöjen ylläpitämiseen. Sato (2011, 225) puolestaan huomauttaa, että kuuluisimmatkaan mallit kielitaidosta (esim. Bachman & Palmer 1996; Canale & Swain 1980) eivät välttämättä ota riittävästi huomioon kommunikatiivisen kyvyn ei-kielellisiä komponentteja ja sitä, kuinka ne vaikuttavat kielelliseen

suoritukseen. Kommunikatiivista lähestymistapaa kieleen ja sen testaamista on kritisoitu niin ikään siitä syystä, että se ei yleensä ota huomioon sitä, kuinka tehokkaasti oppija prosessoi kieltä. Tutkijoiden mukaan prosessoinnin tehokkuutta testaamalla voitaisiin saada hyödyllistä informaatiota oppijoiden (toisen) kielen hallinnasta (Ellis 2005b; Van Moere 2012, 325).

### 2.3 Kielitaidon arvioinnin aikakaudet

Spolsky jakaa toisen ja vieraan kielen arvioinnin kehittymisen kolmeen pääkauteen. Alla olevaan listaan on merkitty luvun jälkeen ensin ne kausien nimet, joiden pohjalle Spolsky alunperin rakensi arviointia koskevan ajanjaksojen luokittelun. Kauttaviivan (/) jälkeen on merkitty Spolskyn myöhemmin uudelleen nimeämien kausien nimet:

1. Ennen vuotta 1950 ajoittuva esitieteellinen kausi / traditionaalinen kausi
2. 1950–60-luvuille sijoittuva psykometris-strukturaalinen kausi / moderni kausi
3. Vuosien 1960–1980 välille sijoittuva psykolingvistis-sosiolingvistinen kausi / postmoderni kausi.

(Spolsky 1995, 353-354.)

Lussier ja Turner (1995, 4) nimittävät 1980-luvulta alkavaa toisen ja vieraan kielen didaktiikassa vallalla olevaa kautta kommunikatiivis-metakognitiiviseksi kaudeksi.

*Ennen vuotta 1950 ajoittuva esitieteellinen kausi*

Euroopassa testejä alettiin käyttää 1700- ja 1800-luvuilla, kun valittiin valtion virkamiehiä tai muita henkilöitä tärkeisiin työtehtäviin. 1800-luvun loppuun mennessä Länsi-Euroopassa oli yleisesti käytössä julkisia kokeita, joiden tarkoituksena oli koulutuksen kontrollointi ja virkamiesten valinta. (Spolsky 1995, 10.) Englannissa otettiin käyttöön jo vuonna 1913 ensimmäinen Cambridgen yliopiston testi, joka sisälsi melko rajallisen puhumisen osakokeen, mikä oli viitoittamassa tietä vuonna 1939 käyttöön otetulle puhumisen taitoa testaavalle Lower Certificatelle<sup>4</sup>. (Fulcher 2003, 6.)

Esitieteellisellä kaudella tyypillisiä kielitaidon testaustapoja olivat muun muassa esseet ja suulliset kokeet, joiden tuloksen päättivät intuitiivisesti auktorisoidut ja autoritääriset testaajat (Spolsky 1995, 353). On kiinnostavaa havaita, että ranskan kielen *docimologie*-sana, joka tarkoittaa kokeisiin liittyviä arvosanajärjestelmiä tutkivaa tiedettä, otettiin käyttöön ranskalaisen psykologin

<sup>4</sup> Nykyisin First Certificate in English eli FCE

Henri Piéron'n ehdotuksesta jo vuonna 1922. Docimologian tutkimuskohteet ovat kielitaidon arvioinnin näkökulmasta edelleen ajankohtaisia: niitä ovat myönteiset seikat, joihin arvioijat kiinnittävät huomiota, ja heidän arvostuksensa muuttaminen pisteiksi, korjaajien väliset erot, asteikkojen soveltaminen, arvosana-asteikot, arvioijien väliset korrelaatiot ja korjaajien välinen yhdenmukaisuus. Docimologian päätehtävänä olikin pyrkiä siihen, että kaikessa arvioinnissa minimoidaan sattuman vaikutus (Cuq 2003, 75). Kielitaidon arviointi oli nykynäkökulmasta katsottuna aina 1940-luvulle asti hyvin intuitiivista, subjektiivista, epätarkkaa eikä keskenään vertailukelpoista toimintaa, jonka validiutta tai luotettavuutta ei varmistettu minkäänlaisten säännönmukaisten laadunvarmennuskeinojen avulla (Lussier & Turner 1995, 12–13).

Suullisen kielitaidon arvioinnin historia liittyy läheisesti erityisesti Yhdysvaltojen sotilaalliseen ja poliittiseen historiaan. Ensimmäisen maailmansodan aikoihin aina 1920-luvulle asti suullisen kielitaidon testaukseen ei kiinnitetty huomiota. Sen sijaan toisen maailmansodan aikaan huomattiin, että yhdysvaltalaisen sotilashenkilöstön sotilastarkoituksiin tarvittava kielitaito (puhuminen, ymmärtäminen ja lukeminen) oli niin heikko, että siitä arveltiin olevan haittaa yleiselle turvallisuudelle. Tästä syystä aloitettiin suullisen kielitaidon testaus sotilashenkilöstön keskuudesta. Sieltä testaus sitten levisi yliopistopiireihin (Fulcher 2003, 1).

Yhdysvalloissa käytössä ollut ensimmäinen todellinen puhumisen testi oli *The College Board's English Competence Examination*, joka otettiin käyttöön vuonna 1930. Testin tarkoituksena oli varmistaa, että yhdysvaltalaisiin yliopistoihin pyrkivillä ulkomaalaisilla opiskelijoilla oli opintojen kannalta riittävä suullisen kielen taito. Alun perin opintokontekstiin suunniteltua testiä alettiin pian käyttää myös Yhdysvaltoihin hakeutuvien maahanmuuttajien kielitaidon testaamiseen. Yhä lisääntyvän maahanmuuton vuoksi kielitaidon testauksella on myös suorat yhteydet Yhdysvaltain armeijassa käytettyihin älykkyystesteihin (*Army Alpha* tai *Beta Tests*), joiden johtopäätökset ovat nykyisten yksilöiden tasavertaisuutta korostavien näkemysten valossa puistattavaa luettavaa. Älykkyystestit näet jakoivat henkilöt eri älykkyysluokkiin, joissa luokkiin jaon perusteena olivat mm. testattavan henkilön maahanmuuton lähtömaa ja ihon väri (Fulcher 2003, 2–4). Fulcherin (2003, 4) mukaan silloisen Yhdysvaltojen presidentin Calvin Coolidgen maahanmuuttopolitiikkaan sidotun testauksen tarkoituksena oli älykkyydeltään vajavaisten ulkomaalaisten pitäminen poissa Yhdysvalloista (ks. myös Spolsky 1995, 36–37). Tuohon aikaan vallalla olleen yleisen huolen mukaan Yhdysvaltoja oli varjeltava lain keinoin (*Immigration Act*, 1924) siltä, ettei liika maahanmuutto pääsisi vaikuttamaan heikentävästi Amerikan väestön älykkyyteen (Fulcher 2003, 4).

Armeijan tarpeiden lähtökohdista oppilaitoksissa kehitettiin erilaisia kielikoulutusohjelmia (ASTP)<sup>5</sup>, joissa kieliopillinen tieto oli toisarvoista ja joiden tarkoituksena oli tehdä sotilaista todelliseen käyttöön tarkoitetuilta suullisilta kommunikaatiotaidoiltaan joko asiantuntijoita (*expert*) tai taitavalla (*competent*) tasolla olevia (Fulcher 2003, 6; Spolsky 1995, 102). Spolsky lisää vielä, että ar-

<sup>5</sup> Army Specialized Training Program

meijalla oli käytössään kahden edellisen kielenhallintaa osoittavan tason lisäksi kaksi muutakin tasoa, ei-kompetentin (*non-competent*) ja ”toivottoman” (*hopeless*) kielenkäyttäjän tasot. Näitä tasoja kielikoulutusta antavat oppilaitokset eivät kuitenkaan yleisesti käyttäneet (Spolsky 1995, 102.)

Merkille pantavaa on, että jo 1940-luvulla pidettiin tärkeänä sitä, että testiosioista saatujen tulosten pohjalta voitaisiin saada luotettavaa näyttöä siitä, miten testattu henkilö selviytyisi todellisesta puhetilanteesta. Jo tuolloin armeijalle ehdotettiin sellaisia suullisen kielitaidon testejä, joissa kiinnitettäisiin huomiota samankaltaisiin seikkoihin kuin nykyäänkin haastattelutesteissä: testispesifikaation tuli olla käytännöllisiä, suoritusten tuli olla mahdollisia arvioida koneellisesti, kielen sujuvuuden eri puolia tuli testata (esim. kyky muotoilla kysymyksiä ja vastata niihin), testaajan tuli olla tehtävään koulutettu, mieluiten ulkomailla asunut sivistynyt henkilö, testauksen tuli tapahtua rauhallisessa ja miellyttävässä tilassa, testaajan tuli olla ystävällinen mutta samalla ammattimainen (*businesslike*) ja epämuodollinen. Ehdottajana toimineen Kaulfersin mielestä suullista sujuvuutta ei voida testata kirjoittamisen avulla eikä voida ajatella, että kirjallinen taito korreloisi sen kanssa, miten henkilö puhuu spontaanissa puhetilanteessa. Spolsky toteaa, että Kaulfersin ehdotukset eivät toteutuneet vielä toisen maailmansodan aikoihin, mutta ajatukset jäivät elämään ja ne tuntuvat vaikuttavan nykyisinkin käytössä oleviin suullista kielitaitoa mittaaviin haastattelutyyppeihin testeihin. (Spolsky 1995, 104–106.)

#### *1950–60-luvuille sijoittuva psykometris-strukturaalinen kausi*

Psykometris-strukturaalisella kaudella tapahtui monia muutoksia, joista voidaan mainita mm. se, että sekä opetus että sen arviointi kehittyivät entisestään. Arvioinnin ensisijaisena tarkoituksena oli varmistaa, että tavoitteet oli saavutettu, ja toissijaisena tarkoituksena oli selvittää, missä määrin. Monissa maissa yleinen oppivelvollisuus tuli voimaan 1950–1960-luvuilla, mikä sai aikaan sen, että oppilaiden määrä kasvoi suuresti ja oppilaiden tasosijoitteluun ja valintaan tarvittiin mittalaitteita. Asiantuntijat alkoivat vaatia, että mittavälineiden tuli olla tarkkoja, tieteellisiä ja objektiivisiä (Lussier & Turner 1995, 16–17). Tälle kaudelle tyypillistä oli kielitaidon arvioinnissa nykyäänkin toteutuva psykometrikkojen ja lingvistien yhteistyön käynnistyminen. Psykometrikot etsivät objektiivisiä ja toistettavissa olevia testausmetodeja, joiden avulla voitaisiin testata niin yhdenmukaisesti kuin mahdollista, kuinka hyvin yksittäinen oppija hallitsi lingvistien testiin valitsemaa osiota (Spolsky 1995, 353). Käyttöön otettiin tilastolliset analyysimenetelmät, joiden avulla selvitettiin testitulosten ja yksittäisten osioiden toimivuutta uusiin reliabiliteetti- ja valideiteetti-käsitteisiin pohjautuen (Lussier & Turner 1995, 17).

Edellä mainittuihin armeijan käyttämiin ASTP-ohjelmiin liittyviä testejä kehitettiin New Yorkin Queen’s Collegessa. Kolmiosaiset testit sisälsivät kuvastakerrontaa, pidempikestoista (*sustained*) puhetta ja ohjattua keskustelua. Tätä yliopistossa kehitettyä testiä voidaan pitää FSI:n (Foreign Service Institute) kuuluisan puhumisen OPI-testin (*Oral Proficiency Interview*) edeltäjänä (Fulcher 2003, 7–8). FSI:n kehittämää testiä pidettiin 1960-luvulla niin luotettavana, että



sitä alkoivat käyttää lukuisat valtion laitokset, kuten CIA , the Defence Language Institute ja Peace Corps. Nämä laitokset kehittivät sitten yhteisesti taitotasoista uuden standardisoidun version, joka nykyisin tunnetaan nimellä ILR (*Interagency Language Roundtable*; Lowe 1987, lainattu Fulcherin teoksesta 2003, 10). Toisen maailmansodan jälkeen tarvetta kielitaidon testaamiseen pidettiin tärkeänä sotilashenkilöiden testauksen lisäksi myös diplomaattikunnan testauksessa; Yhdysvaltojen hallinnossa haluttiin pidettävän kirjaa tulevien tarpeiden varalta näihin kahteen ryhmään kuuluvien henkilöiden kielitaidosta (Fulcher 2003, 17).

Vuonna 1952 kehitettiin ensimmäinen julkinen, laajalle levinnyt kuusitasoinen arviointiasteikko, jota myöhemmin täydennettiin. Fulcher (2003, 9) toteaa, että kun arviointiasteikot kehittyivät, suullisen kielitaidon testaus parani huomattavasti. Asiantuntijat pohtivat jokseenkin samoja aiheita kuin nykyisinkin eli toisin sanoen sitä, mitkä ovat kieliopillisen kompetenssin ja kommunikaatiivisen taidon (*ability*) roolit ja pystytäänkö testillä osoittamaan ja ennustamaan todellista kommunikaatiota.

#### *1960–1980-luvuille sijoittuva psykolingvistikis-sosiolingvistinen kausi*

1960–1980-luvuille sijoittuvaa psykolingvistikis-sosiolingvististä kautta pidetään kielen opettamisen ja pedagogisen arvioinnin kannalta erittäin hyödyllisenä. On kuitenkin syytä muistaa, että 1960-luku oli murroskautta, jolloin sekä vanha että uusi ajattelutapa elivät vielä rinnakkain. Onkin mahdotonta tehdä tarkkoja rajauksia vuosilukujen tasolla varsinkin, kun muutos tapahtui usein hitaasti ja eri tahtiin eri puolilla läntistä maailmaa. Tänä aikana lukuisat tutkijat esittivät uusia näkökantoja kielen rakenteesta ja kielitaidon arvioinnista. 1950–60-luvuilla vallalla olleet strukturalismi ja behaviorismi määrittivät kielen koostuvan eri osista, jotka tuli pitää erillään myös testauksessa (ks. esim. Lado 1961). Carroll (1961) esimerkiksi esitti, että kielen rakenteen ja sanaston ei tulisi olla opettajien ja arvioijien ensimmäisenä kiinnostuksen kohteena, vaan sen sijaan viestin tuottamisen ja arvioitavan selviytymisen kommunikaatiotilanteissa. Carroll (1964) hylkäsikin behavioristien ärsyke–reaktio-mallin ja ehdotti tilalle psykologian ja psykolingvistiikan kielenoppimista koskevia teorioita. Arvioinnissa siirryttiin strukturalistiseen ajatteluun nojanneen näkemyksen mukaan tarkastelemaan kieltä aikaisempaa kokonaisvaltaisemmin eikä enää vain erillisten elementtien summana. Cooper (1968) ehdotti, että Carrollin vuonna 1961 tekemään arviointimalliin lisättäisiin sosiolingvistinen elementti, sillä hänen mukaansa tehokas kielellä kommunikointi on muutakin kuin kielen tuntemista.

Psykolingvistikis-sosiolingvistikisella kaudella vuonna 1967 järjestettiin Michiganin yliopistossa kielitaidon arvioinnin kannalta merkittävä vieraan kielen arviointiin liittyviä ongelmia pohtiva kongressi, johon osallistui kielitaidon arvioinnin asiantuntijoita eri maista. Kongressijulkaisussa otettiin esille monia arviointiin liittyviä ongelmia tai linjanvetoja. Katsottiin, että arvioinnissa on alusta lähtien pyrittävä tarkkuuteen ja selvyyteen ja että oppijoiden suorituksia ei enää pidä arvioida kielen yksittäisten elementtien arvioinnin pohjalta vaan tutkia kommunikaatiotilanteita. Lussier ja Turner (1995, 61) ehdottavatkin, että

Spolskyn antama nimitys psykolingvistik-sosiolingvistinen kausi muutettaisiin mieluummin *psykolingvistik-integratiiviseksi kaudeksi*.

1970-luvulla alettiin hyväksyä arvioinnin pedagoginen arvo osana opettamista ja oppimista. Toisaalta arvioinnin tuotoksia tarvittiin myös työelämän tarpeisiin. Muun muassa Foreign Service Institutin (FSI) suullisessa haastattelutestissä, joka oli kehitetty alun perin vuosien 1952 ja 1956 aikana (Spolsky 1995, 99) ja jota oli käytetty ympäri maailman jo parinkymmenen vuoden ajan vuonna 1978, otettiin käyttöön kielitaidon arviointiin tarkoitettu viisitason arviointiasteikko (0–5). Tasojen nimet paljastavat, että testiä on käytetty ammatilliseen kielenkäyttökontekstiin liittyvän suullisen ja tekstin- ymmärtämisen taidon arviointiin: (1) alkeistaso, (2) rajallinen työtaso, (3) minimaalinen ammatillinen taso, (4) optimaalinen ammatillinen taso ja (5) äidinkielen tasoon verrattavissa oleva taso (Lussier & Turner 1995, 53–54). FSI:n kehittäjät tiesivät tekevän tärkeää pioneerityötä pohtiessaan suullisen kielitaidon testaukseen liittyviä ongelmia, kuten esimerkiksi sitä, kuinka hankkia sopiva puhenäyte testattavalta ja kuinka tehdä siitä luotettava ja validi arvio (Spolsky 1995, 99).

Lussierin ja Turnerin mukaan (1995) Clark (1975, 1979) teki psykolingvistik-sosiolingvistik-kaudella puhumistaitoa arvioivien testien kolmijaon, joka on vieläkin ajankohtainen. Puhumista mittaavat testit jaettiin seuraaviin ryhmiin:

1. Suorat testit; pohjana kommunikatiivinen kasvokkainen vuorovaikutus (kuten esimerkiksi edellä mainittu FSI:n haastattelutesti),
2. Osittain epäsuorat testit (*semi-direct*) (esim. tarinan kertominen kuvan pohjalta); testattavalta ei odoteta vuorovaikutusta
3. Epäsuorat testit; pohjana kirjoitettu tai nauhoitettu ärsyke

(Lussier & Turner 1995, 55).

Usein puhumista mittaavat testit jaetaan kahteen pääluokkaan, eli toisin sanoen suoriin ja epäsuoriin testeihin. Suoriin puhumisen testeihin sisältyy yleensä kasvokkainen vuorovaikutus, vaikka puhetta voidaan toki testata myös esimerkiksi puhelimen välityksellä. Epäsuorat puhumisen testit taas ovat testejä, joissa testin suorittaja reagoi joko ääninauhalta tai videonauhalla kuultaviin erilaisiin tehtäviin (Galaczi 2010, 30).

*1980-luvulta alkaen nykyaikaan asti ulottuva kommunikatiivis-metakognitiivinen kausi*

1980-luvulta alkaen suullisen kielitaidon testauksessa on tapahtunut paljon, sillä suullisen kielitaidon testausmenetelmät ovat monipuolistuneet käsittämään perinteisen haastattelun lisäksi myös muita tehtävätyyppejä, esimerkiksi paritehtäviä, roolileikkejä tai ryhmäkeskusteluja. Yleinen tietotekninen kehitys viimeisten vuosikymmenien aikana on luonnollisesti myös vaikuttanut kielitaidon arviointiin. Suullisen kielitaidon testaus on yleistynyt ja siten se onkin tärkeä osa useimpia merkittäviä kielitutkintoja. Amerikkalaisen TOEFL-testin suullisen osan kehittyminen on hyvä esimerkki koko alan kehittymisestä. Ai-

koinaan TOEFL-testissä ei ollut lainkaan puhumistaitoa testaavaa osaa, ja sitten puhumisen testin suorittamisesta tuli vapaaehtoista ja se tehtiin TSE-testillä (*Test of Spoken English*). Uusimmassa tietokonepohjaisessa TOEFL iBT -testissä puhumisen testin suorittaminen on pakollinen osa tutkinnon suorittamista (Zareva 2005, 50–53).

Ensimmäinen ihmisten väliseen vuorovaikutukseen perustuva suullisen taidon haastattelutesti, joka edellä jo mainittiin, on *Oral Proficiency Interview* (OPI). 1950-luvulla kehitetyn OPI-testin jälkeen suullisen taidon testaamista kehitettiin 1980-luvulla ottamalla mukaan nauhan avulla toteutettu simuloitu tapa testata suullista kielitaitoa (*Simulated Oral Proficiency Interview* eli SOPI). Sitä on kutsuttu ensimmäisen tietotekniikkasukupolven edustajaksi suullisen kielitaidon arvioinnin alalla (Kenyon & Malone 2010, 2). Tietotekniikka on siis tullut entistä enemmän mukaan myös kielitaidon arviointiin viime vuosikymmeninä. Muita kielen osataitoja pystytään helposti testaamaan tietokoneen avulla, mutta suullisen kielitaidon testaamiseen tarkoitettuja testejä on saatu odottaa melko kauan. Muun muassa tietokonepohjaisessa DIALANG-järjestelmässä ei ole vielä kukaan suullista kielitaitoa arvioivaa osaa. Nykyisin on kuitenkin olemassa jo joitakin tietokonepohjaisia suullista kielitaitoa testaavia testejä. Näistä voidaan mainita toisen tietotekniikkasukupolven edustajana *The Computerized Oral Proficiency Instrument* (COPI), jossa ihmiset arvioivat suoriutuksen edelleen, tai TOEFL iBT -testi sekä alusta loppuun tietokonepohjaiset Versant-testi ja PTE Academic speaking test -niminen testi (Araújo 2010, ii–iii; Kenyon & Malone 2010, 3; Galaczi 2010, Zareva 2005).

Kenyonin ja Malonen (2010, 5) mukaan tietotekniikkaa alettiin käyttää suullisen kielitaidon arviointiin 1990-luvun lopussa osittain siitä syystä, että tietokoneteknologia laajeni tuolloin koulutuksen alalle. Haluttiin selvittää, voitaisiinko tietotekniikan avulla löytää ratkaisuja tiettyihin SOPI:ssa ilmenneisiin rajoitteisiin. COPI-testissä onkin monia parannuksia tutkinnon suorittajan kannalta verrattuna nauhapohjaiseen SOPI-testiin. Tutkinnon suorittajalle tärkeistä helpotuksista COPI-testissä voidaan mainita esimerkiksi se, että suorittaja pystyy tietyn ennalta päätetyn maksimian sisällä päättämään, kuinka paljon aikaa hän käyttää tutkintotehtävään tutustumiseen ja kuinka paljon vastauksen antamiseen. Lisäksi tutkinnon suorittamista helpottaa se, että suorittaja voi valita puheenaiheen.

Tietokoneavusteinen suullisen kielitaidon arviointi on lisääntynyt suuresti viimeisten viiden vuoden aikana, koska suullisen kielitaidon nopeaan ja tehokkaaseen testaamiseen on yleistä tarvetta. Tietokoneen avulla tehty suullisen kielitaidon arviointi toimii kätevänä vaihtoehtoisena ja/tai täydentävänä tapana perinteisille kasvokkaisille tavoille arvioida suullista kielitaitoa. Viime vuosina monet kielitaidon arvioinnin asiantuntijat ovatkin tuoneet markkinoille tietokonepohjaisia puhumisen testejä, esimerkiksi ETS:n TOEFL iBT -testin, Pearsonin akateeminen PTE-testin, tai Cambridge ESOL:n BULATS-testin (Galaczi 2010, 29.)

Vaikka tietokoneavusteinen suullisen kielitaidon arviointi kehittyikin koko ajan, tutkijat varoittavat sen tämänhetkisistä puutteista. Puutteiksi mainitaan

muun muassa se, että tietokoneella ei voida arvioida suulliseen kielitaitoon kuuluvaa vuorovaikutuksellisuutta tai että tehtävätyyppien käytössä mennään kehityksessä taaksepäin, sillä tietokoneella ei voida toteuttaa kaikkia puhumisen arviointiin kiinteästi liittyviä tehtävätyyppejä (Alderson & Bachman 2007, xi; Luoma 2004, 50). Jotkut tutkijat muistuttavat, että ei ole vain yhtä tapaa arvioida suullista kielitaitoa. Olisikin aina pidettävä mielessä, mikä on testin tarkoitus ja sopiiko valittava testi tiettyyn tarkoitukseen ('fitness for purpose'). Tietokoneella toteutettavat testit ovat helposti järjestettäviä ja kaikille suorittajille samanlaisia, mutta toisaalta niistä puuttuu vuorovaikutuksellinen elementti ja ne ovat, ainakin toistaiseksi, suppeampia mahdollisten tehtävätyyppiensä monipuolisuudessa kuin kasvokkain toteutetut puhumisen testit. Kasvokkaisilla suullisen taidon testeillä saadaan tietokonetestejä tarkempaa, laajempaa ja syvempää näyttöä suorittajan kielitaidosta, silloin kun tarvetta tähän on (Galaczi 2010, 47.)

Jotkut kielitaidon arviointia edustavat tahot eivät enää puhu tietokoneavusteisesta (*computer assisted*) vaan täysin automaattisesta puhumisen testisuorituksesta ja arvioinnista (*fully automated*; Van Moere 2010, 84). Näissä 1990-luvulla kehitetyissä testeissä koko prosessi alusta loppuun asti tapahtuu koneellisesti ja ne käyttävät hyväkseen puheentunnistusta ja puheenprosessointiteknologiaa (*speech processing technology*). Vaikka tietokoneella tapahtuvaa täysin automaattista puhumisen testaamista onkin kritisoitu, testien puolestapuhujien mukaan tietokoneen ja koulutetun kielitaidon arvioijan tekemien arviointien välillä on selvä korrelaatio (vähintään .93; Van Moere 2010, 86).

## 2.4 Suullisen kielitaidon arviointi Suomessa

Suullisen kielitaidon merkitystä on alettu korostaa erityisesti 1980-luvulta lähtien kommunikatiivisen kielenopetuksen saatua jalansijaa myös Suomessa. Suullisen kielitaidon opetuksen välttämättömyydestä ja arvioinnista on keskusteltu monissa maissa, kuten myös Suomessa, jo vuosikymmeniä. Jo 1970-luvulla Suomessa tehtyjen tutkimusten mukaan oppilaat yleensä pitävät suullisen kielitaidon oppimista vieraan kielen oppimisen tärkeimpänä tavoitteena (Takala 1977, 10). Takala (1973, 18) toteaa perusopetuksen opetussuunnitelman perusteisiin liittyvässä oppaassa jo 1970-luvun alussa, että huolimatta puhumista mittaavien kokeiden vaikeasta järjestämisestä ja niihin liittyvistä ongelmista on kuitenkin välttämätöntä sisällyttää koeohjelmaan myös suullista kielitaitoa mittaavia osakokeita, koska muuten kielenopetuksen keskeinen tavoite, aktiivisen suullisen kielitaidon saavuttaminen, jää haaveeksi. Takala myös varoittaa opettajia liiallisesta virhekeskeisyydestä ja muistuttaa, että suullisen kielitaidon edistämiseksi koulussa olisi osattava tehdä kompromisseja kielellisesti ideaalin ja toiminnallisesti tehokkaan kielenkäytön välillä (Takala 1973, 18–19). Suullisen kielitaidon tärkeys ymmärrettiin vähitellen suomalaisessa koulujärjestelmässä, ja maininnat suullisen kielitaidon oppimisesta ja sen tavoitteista ilmestyivätkin

sekä peruskoulun että lukion toisen kotimaisen kielen ja vieraiden kielten opetussuunnitelmien perusteisiin 1990-luvulla (POPS 1994; LOPS 1994). Huomattavaa kuitenkin on, että vielä tässä vaiheessa suullista kielitaitoa ei kuitenkaan testattu kouluissa millään yhteismitallisella testillä (ks. jäljempänä).

Lisääntynyt kiinnostus suullista kielitaitoa kohtaan näkyi suomalaisissa kouluissa varsinkin EU:hun liittymisen aikoihin monipuolistuvina kielivalintoina. Yleisesti ottaen oppilaatkin ovat tuntuneet pitävän suullista kielitaitoa ja puhumisen harjoittelua tärkeänä. Jo vuonna 1994 oppilaat toivoivat, että kieltenopetus loisi jo ala-asteelta alkaen paremman pohjan puhumisen harjoittelulle, jotta koulussa opittu kieli antaisi mahdollisuuden selviytyä edes arkipäivän tilanteista vieraalla kielellä (Yli-Renko 1991, 54).

Viitekehysten laatimisessa on painotettu oppimisen ja opettamisen ensisijaisuutta varsinaiseen arviointitoimintaan verrattuna, mistä kertoo dokumentin nimessä olevien sanojen tietoinen esitysjärjestys (oppiminen, opettaminen ja arviointi).

Huhta (1993, 84) muistuttaa, että huolimatta kommunikatiivisen ja funktionaalisen lähestymistavan tulosta opetukseen 1970-luvulla uutta kielitaitonäkemyksiä testaavat ensimmäiset viralliset kielitestit kehitettiin vasta 1980-luvun alussa.

Suomessa tehdyt ja suomalaista kontekstia koskevat suulliseen kielitaitoon liittyvät tutkimukset ovat monissa tapauksissa käsitelleet joko suullisen kielitaidon arviointia tai testausta (ks. esim. Saleva 1997; Hildén 2000). Yli-Rengon 1980-luvulla tekemissä saksan kieltä koskevissa tutkimuksissa suullinen kielitaito ja arkailemattomuus puhua vierasta kieltä nousivat yhdeksi tärkeimmistä kielen oppimisen tavoitteista (Yli-Renko 1985; 1989). Harjanne (2006, 5) puolestaan toi tutkimuksessaan esille näkökannan, jonka mukaan vieraiden kielten opetus, oppiminen ja harjoittelu ovat ensisijaisia verrattuna kielen testaukseen tai arviointiin. Ammattikorkeakoulujen ja yliopistojen yhteinen valtakunnallinen kieltenopetuksen kehittämishanke KORU antaa kokonaiskuvan ruotsin kielen opetukseen ja arviointiin liittyvistä käytänteistä tietyissä korkeakouluissa Suomessa. Ainutlaatuisessa yhteishankkeessa selvitettiin kattavasti vallalla olevia käytänteitä ja kehitettiin arvioinnin tueksi materiaalia (taitotasokuvaukset ja maamerkit). Tutkimuksessa selvitettiin myös itse- ja vertaisarviointiin liittyvien aktiviteettien käyttöä (Elsinen & Juurakko-Paavola 2006). Viime vuosina on kuitenkin todettu, että myöskään KORU-hanke ei sisällä kriteereitä, joita voitaisiin hyödyntää, kun arviointi kohdistuu keskustelutilanteen viestinnällisten piirteiden erottamiseen opiskelijoiden persoonallisuuspiirteistä ja sosiaalisista taidoista (Alanen & Männikkö 2012).

Koulukontekstissa Opetushallitus on huolehtinut vapaaehtoisen lukion suullisen kokeen arviointimateriaalin laatimisesta noin 1990-luvun puolivälistä alkaen. Koska suullisten kokeiden tilaaminen kouluihin sekä opettajan suorittamaan arviointiin osallistuminen on ollut vapaaehtoista, kaikki koulut eivät ole tarjonneet oppilailleen suullisen kokeen suorittamismahdollisuutta. Opetus- ja kulttuuriministeriö, Opetushallitus ja osittain myös Ylioppilastutkintolautakunta ovatkin omalta osaltaan pyrkineet kehittämään suullisen kielitaidon ar-

viointia erityisesti 2000-luvulla. Opetusministeriön yhteyteen vuonna 2006 asetetun lukiokoulutuksen suullisen kielitaidon arviointityöryhmän tehtävänä oli selvittää, kuinka suullisen kielitaidon arviointi olisi yhteismitallisesti ja luotettavasti järjestettävissä, mitä kieliä ja mitä kielten oppimääriä suullisen kielitaidon arviointi koskisi ja mitkä olisivat ehdotuksen aiheuttamat kustannukset (Lukiokoulutuksen suullisen kielitaidon arviointityöryhmän muistio 2006, 26). Työryhmä kartoitti työnsä pohjaksi suullisen kielitaidon arvioinnin senhetkistä tilaa ja arvioinnissa käytettyjä kokeita sekä Suomessa että eräissä muissa Euroopan unionin jäsenmaissa. Työryhmä tarkasteli lukiokoulutuksessa saavutettavan suullisen kielitaidon arvioimiseksi neljää vaihtoehtoa: 1) vapaaehtoista suullisen kielitaidon arviointia Opetushallituksen tuottamien kokeiden pohjalta, 2) suullisen kielitaidon arviointia yleisten kielitutkintojen avulla, 3) suullisen kielitaidon arviointia osana ylioppilastutkinnon kielikokeita tai 4) suullisen kielitaidon koetta osana lukion opetussuunnitelmaa. Työryhmä päätyi yksimielisesti ehdottamaan suullisen kielitaidon arvioimista Opetushallituksen valtakunnallisilla kokeilla lukion A- ja B1-kielten valtakunnallisten syventävien kurssien yhteydessä.

Opetushallitus muutti vuonna 2009 antamallaan määräyksellä (10/011/2009) nuorille tarkoitetun lukiokoulutuksen opetussuunnitelman perusteita (LOPS 2003). Muutoksen yhteydessä tietyt lukion kurssit muutettiin suullisen kielitaidon kursseiksi. Suullisen kielitaidon kokeesta tuli osa suullisen kielitaidon kurssia, ja koe sisältyi osana kurssin arviointiin. Kurssien suoritusarvioinnin perustuu sekä suullisen kielitaidon kokeesta saatuun arvosanaan että muihin kurssin aikaisiin näyttöihin.

Opetushallitus avasi syksyllä 2010 lukioiden käyttöön suullisen kielitaidon arviointia varten koepankin. Koepankki<sup>6</sup> sisältää koetehtävät, kokeiden arviointiohjeet sekä yleiset ohjeet kokeiden järjestämisestä.

#### 2.4.1 Yleiset kielitutkinnot

Vieraiden kielten taidon merkitys korostui aikuiskoulutuksen puolella, kun Suomen valtio 1990-luvun alussa päätti perustaa virallisen lakisääteisen kansallisen monikielisen (9 kieltä) tutkintojärjestelmän Yleiset kielitutkinnot (Yki). Yki-tutkintojen kehittämisen lähtöideana oli aikuisten kieltenoppijoiden tukeminen. Yleisissä kielitutkinnoissa puhumisen osakoe on ollut alusta lähtien mukana joko kielistudiokokeena ja/tai haastatteluna. Puhumistaitoa arvioidaan perus-, keski- ja ylimmällä tasolla kielistudiossa (poikkeuksena ruotsin ja suomen kielen perustaso) tehtävillä, joissa esimerkiksi osallistutaan simuloituihin keskusteluihin, reagoidaan annettuihin tilanteisiin sekä esitetään mielipide tai kerrotaan jostakin annetusta aiheesta. Tämän lisäksi kaikkien kielten ylimmällä tasolla sekä ruotsin ja suomen kielen perustasolla on henkilökohtainen haastattelu. Kaikki kielistudiosuoritukset sekä ruotsin ja suomen perustason haastatte-

<sup>6</sup> ([http://www.oph.fi/koulutus\\_ja\\_tutkinnot/lukiokoulutus](http://www.oph.fi/koulutus_ja_tutkinnot/lukiokoulutus))

lut äänitetään ja ylimmän tason haastattelut videoidaan arviointia varten (Yleisten kielitutkintojen perusteet 2011, 10).

#### 2.4.2 Valtionhallinnon kielitutkinnot

Toinen Suomen valtion lakisääteisistä kielitutkintojärjestelmistä on maamme kansalliskieliä testaava Valtionhallinnon kielitutkinnot (entinen *Valtion kielitutkinnot*). Valtion kielitutkintojärjestelmä pohjautui vuodelta 1922 peräisin olevaan lainsäädäntöön (149/1922; 313/1922). Järjestelmää koskevaa aiempaa lainsäädäntöä tutkimalla selviää, millaisena suullinen kielitaito ja sen arviointi nähtiin ennen kuin järjestelmä siirrettiin Opetushallitukseen ja otettiin käyttöön uudistetussa muodossaan vuonna 2004. Aiemmin tutkintojärjestelmää koskevissa säädöksissä määrättiin, että voidakseen osoittaa suomen tai ruotsin suullista taitoa tutkittavan tuli osoittaa, että hän ymmärsi puhetta tällä kielellä sekä osasi käyttää sitä suullisesti. Näiden nykykäsityksen mukaisten suulliseen kielitaitoon liittyvien määrittelyjen lisäksi säädettiin suomen- tai ruotsinkielisen kirjoituksen ymmärtämisen kuuluvan suulliseen kielitaitoon (442/1987, 7 §). Nykynäkemyksessä suullisesta kielitaidosta ja sen testaustavoista eroaa suuresti aiemmin käytössä olleista. Yhdistetyssä suullisen ja kirjallisen taidon tutkinnossa näet suullista kielitaitoa tutkittiin keskustelemalla tutkinnon suorittajan kanssa hänen omasta ammattialastaan, mikä kuulostaa nykyisinkin perustellulta tavalta arvioida suullista taitoa. Sen sijaan tutkintotilaisuudessa toimeenpannun suullisen keskustelun tuli johtosäännön mukaan osoittaa, ei pelkästään suoritettavan kielen puheen vaan myös senkielisen kirjoituksen ymmärtämistä, mikä nykyisin kuulostaa oudolta suullisen kielitaidon arvioinnin yhteydessä.

Valtion kielitutkinnoissa oli aiemmin neljänlaisia kielitutkintoja:

1. tutkinto, joka koski suomen tai ruotsin kielen suullista ja kirjallista taitoa
2. tutkinto, joka koski suomen tai ruotsin kielen suullista taitoa
3. tutkinto, joka koski kykyä ymmärtää suomen tai ruotsin kieltä
4. tutkinto, joka koski suomen tai ruotsin kielen täydellistä hallitsemista.

Kolmen ensimmäisen suhteellisen helposti hahmoteltavissa olevan tutkinnon lisäksi säädöksissä mainittiin neljäntenä listassa oleva tarkemmin määrittelemätön kielen täydellistä hallitsemista osoittava kielitutkinto. Tämän tutkinnon suorittamisesta kerrottiin tutkintolautakuntien johtosäännössä: tutkinnon suorittajan tuli kirjoittaa omalta ammattialaltaan esitys lautakunnan määräämästä aiheesta sekä keskustella lautakunnan jäsenen kanssa suullisesti omaan ammattialaansa kuuluvasta aiheesta. Lisäkoikeena voitiin käyttää myös käännoästä kansalliskielien välillä (SK 82/1923). Silloinen käsitys kielen erinomaisesta hallinnasta ilmaistiin siis kielen täydellisenä hallitsemisena, ja täydelliseen hallitsemiseen luettiin kuuluvaksi myös suullinen kielenhallinta keskustelun muodossa.

Nykyisessä Valtionhallinnon kielitutkintojärjestelmässä voidaan osoittaa sellaista suomen tai ruotsin kielen taitoa, jota vaaditaan valtion ja muun julkishallinnon työtehtävissä. Järjestelmässä on omat tutkinnot suulliselle kielitaidolle, kirjalliselle kielitaidolle ja ymmärtämiselle. Suullista taitoa arvioidaan

suomen tai ruotsin kielen suullisen viestintäkyvyn perusteella. Suullisen kielitaidon osoittaminen edellyttää puheen ymmärtämisen ja puhumisen osakokeiden hyväksyttävää suorittamista (Valtionhallinnon kielitutkintojen perusteet 2005, 7).

Valtionhallinnon kielitutkintojen hyvänä puolena on niiden joustavuus: ne on suoritettavissa suhteellisen nopeasti ja tietyn taidon tutkintoon kuuluvat osakokeet on mahdollista suorittaa joko yhdessä tai erikseen. Tutkinnon hyväksytyttyä suorittaminen edellyttää kuitenkin tietyn taidon molempien osakokeiden hyväksytyttyä suorittamista.

### 2.4.3 Ylioppilastutkinto

Kuten edellä mainittiin, suullisen kielitaidon tärkeys on tunnustettu opetusala- ja muuallakin yhteiskunnassa jo pitkään, mutta silti suullisen kielitaidon systemaattinen arviointi esimerkiksi osana ylioppilastutkintoa puuttuu. Suullisen kielitaidon kokeen liittämistä ylioppilastutkintoon on keskusteltu monen vuosikymmenen ajan. Hirvonen esitteli väitöskirjassaan jo vuonna 1974 lukion uuden englannin kielen päättötutkinnon, jossa suullista tuottamista olisi arvioitu sisälukukokeen, testaaajan kysymyksiin vastaamisen ja tilanteessa reagoitien avulla (Hirvonen 1974). Myös Takala esitti jo vuonna 1993 konkreettisia vaihtoehtoja suullisen kielitaidon testaamiseksi osana kielten ylioppilastutkintoa (Takala 1993). Opetushallitusta edeltäneellä Kouluhallituksella on niin ikään ollut aktiivinen osa suullisen kielitaidon kokeen suunnittelussa. Tätä tarkoitusta varten Kouluhallituksella oli vuosina 1988–1989 työryhmä, jonka tehtävänä oli muun muassa tehdä konkreettinen ehdotus suomalaiseen ylioppilastutkintoon liitettävästä suullisen tuottamisen kokeesta (Takala 1993, 35). Työryhmän perusteellinen selvitystyö johti suullisen kielitaidon kehittämiseen liittyvään kokeilutoimintaan 1990-luvun alussa, mutta tästä huolimatta suullisen kielitaidon pakollista koetta ei edelleenkään ole saatu liitetyksi valtakunnalliseen tutkintoon.

Suullisen kielitaidon kokeen liittämistä ylioppilastutkintoon on pohdittu monissa työryhmissä ja raporteissa 1980–1990-luvuilla. Suullisen kielitaidon aseman vahvistaminen opetuksessa ja arvioinnissa sisältyy myös vuonna 2007 julkaistun Kielikoulutuspoliittisen projektin loppuraportin keskeisiin suosituksiin. Suosituksessa ehdotetaan, että olisi tarpeen selvittää, voitaisiinko sekä äidinkieliä että vieraista kielistä antaa suullisesta kielitaidosta erillinen arvosana tai kielitaitoprofiili perusasteen ja toisen asteen päättötodistuksessa. Tämän lisäksi tulisi selvittää mahdollisuuksia sisällyttää suullista kielitaitoa arvioivia osia kansallisiin tutkintoihin, esimerkiksi ylioppilastutkintoon ja kokeisiin (esim. peruskoulun valtakunnalliset kokeet; Pöyhönen & Luukka 2007, 469).

Syy, miksi suullista kielitaitoa ei arvioida ylioppilastutkinnossa, on ehkä se, että suullisen kielitaidon arvioinnin järjestäminen valtakunnallisesti isolle määrälle oppilaita sanotaan olevan (liian) hankala tehtävä. Tämä argumentti ei kuitenkaan ole hyvä, sillä monissa muissa maissa ylioppilaskirjoitusten kielten kokeisiin osallistuvien vuosittaiset määrät ovat huomattavasti suuremmat kuin Suomessa, ja silti niissä testataan suullista kielitaitoa ylioppilastutkinnon osana.



Salevan (1997) mukaan yksi syy siihen, että suomalainen ylioppilastutkinto ei vielä vuonna 1997 sisältänyt suullisen kielitaidon koetta, ovat koulujen opettajien asenteet. Hänen mukaansa monet vieraiden kielten opettajat pitävät työmääräänsä jo nykyisellään niin suurena, että uuden suullista kielitaitoa testaavan testin käyttöönotto tuntuu liian haastavalta. Saleva arvelee myös, että kyseessä voi olla myös pelko uutta asiaa kohtaan. Saleva tähdensi jo melkein 15 vuotta sitten, että Suomella ei ole varaa jäädä katsomaan sivusta, kun muu Eurooppa seuraa Euroopan neuvoston suosituksia ja kehittää suullisen kielitaidon arviointia sekä korostaa suullisen kielitaidon tärkeyttä yleensäkin. (Saleva 1997, 146.)

Monet tahot ovat tukeneet suullisen kielitaidon sisällyttämistä ylioppilastutkintoon. Muun muassa Suomen Lukiolaisten Liiton mielestä ylioppilastutkinto ei ole koskaan valmis, ja siten suullisen kielitaidon kokeiden sisällyttäminen tutkintoon olisikin liiton mielestä seuraava askel tutkinnon kehittämisessä 2000-luvulla toteutettujen uudistusten jälkeen (tutkinnon uusi rakenne 2005, ainereaalit 2006 ja äidinkielen koe 2007; Suomen Lukiolaisten Liitto ry 2005).

Ylioppilastutkinnon kehittäminen on mainittu nykyisen hallituksen ohjelmassa kohdassa, jossa käsitellään toista astetta ja lukiota, mutta ohjelmassa ei määritellä tarkemmin, mitä muuta tämä kehittämistyö sisältää kuin tieto- ja viestintätekniikan käyttöönoton asteittaisen valmistelun ylioppilaskirjoituksiin (Hallitusohjelma 2011). Suullisen kielitaidon kehittämisen kannalta on lisäksi syytä muistaa, että jos suullinen kielitaito olisi osa arviointia, se parantaisi myös suullisen kielitaidon asemaa opetuksessa (Tarnanen, Huhta & Pohjala 2007, 406).

Yhteenvetona voidaan todeta, että puhumistaitoa pidetään Suomessa yleisellä tasolla merkittävänä ja siten puhumisesta on tullut tärkeä osa kielenopetusta. Kieltenopettajat ottavat vaihtelevassa määrin puhumisen huomioon arvosanoja antaessaan. Koska suullista kielitaitoa pitävät tärkeänä niin oppijat ja opettajat kuin esimerkiksi elinkeinoelämänkin edustus, olisi tietenkin toivottavaa, että puhumisen taitoa arvioitaisiin säännöllisesti valtakunnallisella kokeella eli toisin sanoen ylioppilastutkinnon yhteydessä.

Suullisen kokeen suunnittelu ylioppilastutkintoon näyttää vihdoinkin etenevän ylioppilastutkinnon sähköistämishankkeen osana. Median (Helsingin Sanomien verkkolehti 11.6.2013) mukaan ylioppilastutkintoon ollaankin saamassa suullisen kielitaidon koe viimeistään vuonna 2019.

## 3 KIELIT AidON ITSEARVIOINTI

Itsearviointia voidaan tehdä monissa eri yhteyksissä, joista kielitaidon arviointi on vain yksi esimerkki. Tämän tutkimuksen kohteena on kielitaidon itsearviointi erityisesti suullisen kielitaidon mittaamisessa. Luvussa 3.1 kuvaillaan aluksi itsearvioinnin taustaa, itsearviointiin liittyviä keskeisiä käsitteitä sekä tutkimukseen läheisesti liittyvän DIALANG-järjestelmän itsearviointiosuutta. Itsearvioinnin taustaa seuraavissa luvuissa tarkastellaan itsearvioinnin käyttötarkoituksia (luku 3.2) ja itsearviointimateriaaleja ja -metodeja (luku 3.3). Luvussa 3.4 käsitellään itsearviointiin liittyvää tutkimusta eri näkökulmista. Luvun 3.5 aiheena on itsearviointi ja sitä koskevat eri toimijoiden asenteet. Luvussa 3.6 luodaan lyhyt katsaus itsearviointiin opettamisen ja opetussuunnitelmien näkökulmasta, ja luvussa 3.7 esitellään lopuksi kielitaidon vertaisarviointia.

### 3.1 Kielitaidon itsearvioinnin taustaa

#### 3.1.1 Oppija uudessa roolissa

Kielitaidon itsearviointi on suhteellisen nuorta verrattuna ulkopuoliseen kielitaidon arviointiin. Käyttäytymistieteiden alalla erilaiset omaa toimintaa tarkastelevat selonteot ovat olleet käytössä jo pitkään, mutta kielitaidon arvioinnissa itsearviointia koskevia pohdintoja alettiin julkaista vasta 1980-luvulla. Tärkeinä taustavaikuttajina itsearvioinnin ilmaantumiselle kielitaidon arvioinnin piiriin toimivat erilaiset Euroopan neuvoston oppijan roolia opetus- ja oppimisproses- sissa korostavat projektit ja hankkeet, joita oli ollut käynnissä jo 1970-luvulta lähtien. Ennen tätä ajankohtaa kielikoulutuksen piirissä tapahtuva itsearviointiin liittyvä tutkimus oli jokseenkin harvinaista. (Oscarson 1997, 176.)

Itsearviointia koskevia tutkimuksia ilmestyi aluksi varsinkin soveltavaan kielitieteeseen ja kasvatustieteelliseen arviointitutkimukseen keskittyvissä julkaisuissa (Ross 1998, 1). Itsearviointi-käsitteen ilmaantuminen sekä julkaisuihin että käytännön tason opetustilanteisiin 1970-luvulta lähtien liittyy laajempaan koulutusfilosofiseen muutokseen: siirryttiin vähitellen opettajakeskeisestä pe-

dagogiikasta oppijan tarpeisiin ja kiinnostuksen kohteisiin keskittyvään pedagogiikkaan. Tällöin virallisissakin yhteyksissä myönnettiin arvioinnin pedagoginen arvo. Arvioinnin avulla saatiin hyödyllistä tietoa oppimisen edistymisestä, diagnosoitiin oppijan vahvuuksia ja heikkouksia sekä saatiin tietoa mahdollisten oppimisessa tarvittavien tukitoimien aloittamisesta. (Lussier & Turner 1995, 41–42.)

Tämä koulutusfilosofinen muutos liittyy yleisemminkin yhteiskunnassa tapahtuneisiin muutoksiin. Tällaisina muutoksina voidaan pitää perinteisten elämänmuotojen (kuten esimerkiksi perheen ja uskonnon) merkityksen ja painoarvon muuttumista yhteiskunnassa. Onkin selvää, että tämänkaltaiset, muun muassa sosiologian tutkimuskohteena olevat muutokset heijastuvat myös koulutukseen puolelle uusina ja erilaisina oppimiskäsityksinä ja opetusikäntöinä. Yleinen suuntaus koulutuksessa on siten jo jonkin aikaa ollut siirtymässä yksilöllisempiin koulutusprosesseihin, jotka mahdollistavat muun muassa itsearvioinnissa tarvittavan omatoimisuuden, itseohjautuvuuden ja tekemällä oppimisen (*learning by doing*). (Hirsjärvi & Remes 1988, 136–153.) Itsearviointiin esiintulo liittyy monien tutkijoiden (esim. Shepard 1994) mukaan siihen, että perinteiset oppimisen arviointitavat eivät ole kiinnittäneet huomiota oppimisen kannalta nykyisin tärkeinä pidettäviin seikkoihin, kuten oppijoiden asenteisiin, motivaatioon, kiinnostuksen kohteisiin ja oppimisstrategioihin.

Euroopan neuvoston 1970-luvulla aloittamien hankkeiden tarkoituksena oli yhtenäisen kielen kommunikatiivisuuteen liittyvän oppimisjärjestelmän luominen aikuisille vieraan kielen oppijoille. Esimerkkinä tällaisista hankkeista voidaan mainita *Modern Languages* -hanke, jonka lähtökohtana oli kommunikatiivisen kompetenssin korostaminen ja jonka tarkoituksena oli mm. kielenoppimisen demokratisoiminen. Tämä projekti teki 1970-luvulla urauurtavaa työtä korostaessaan oppijan itsenäisyyttä ja opettajan ja oppijan yhteistä vastuuta vieraiden kielten opiskelussa (Holec 1988, 1981; Huttunen 1990; North 1992). Euroopan neuvoston monista muista tärkeistä kansainvälisistä kehittämishankkeista voidaan mainita myös vuosina 1989–1996 toteutettu *Language learning for European citizenship* -projekti, jossa korostuivat muun muassa *oppimaan oppiminen* ja oppijan itsenäisyyden edistäminen (*Language learning for European citizenship*, 9).

Euroopan neuvoston mittava työ aikuisten kielenoppijoiden vieraiden kielen oppimisen alalla selittää sen, että itsearviointi on ollut alusta lähtien erottamaton osa 1990-luvulla ilmestynyttä Euroopan neuvoston julkaisemaa Eurooppalaista viitekehystä (EVK). Euroopan neuvosto haluaa tukea sellaisia opiskelu- ja opetustapoja, jotka auttavat oppijoita kehittämään asenteitaan, tietojaan ja taitojaan. Nämä taas edistävät mm. oppijoiden ajattelun ja toiminnan kehittymistä itsenäiseksi ja yhteistyökykyiseksi. Oppijoita halutaan kannustaa oppijan autonomian tielle ja tehdä heistä demokraattisen yhteiskunnan kansalaisia (EVK 2004, 11). Tässä kymmenen vuoden kansainvälisen yhteistyön tuloksena toimitetussa julkaisussa, jolla on virallinen asema sekä EU:ssa että Euroopan neuvostossa kielikoulutuksen alalla, itsearviointi mainitaan useissa eri kohdissa. Kielen kommunikatiivisiin toimintoihin keskittyvä itsearviointiloke-

rikko edustaakin Viitekehysten ensimmäisiä taulukoita. Lisäksi viittauksia itsearviointiin on muun muassa kohdissa ”muiden suorittama arviointi ja itsearviointi”, ”opiskelun suunnittelu” ja ”DIALANG-itsearviointiasteikot” (EVK 2004, 26, 50–51, 261, 306). Itsearvioinnin liittäminen Viitekehykseen olikin luonteva valinta. Viitekehys luotiin alun perin sitä varten, että voitiin helpottaa eri maiden koulutukseen ja opetukseen liittyviä tavoitteita ja vaatimustasoja koskevaa kommunikaatiota sekä tutkintojen keskinäistä tunnustamista vertaamalla näitä yhteistyössä luotuun kalibroituun ja kuvailevaan standardiin (Trim 2001). Eurooppalaisen viitekehysten pääasiallisiin käyttötarkoituksiin kuuluu myös opiskelun itsenäinen suunnittelu, johon itsearviointitaito luonnollisesti sisältyy (EVK 2004, 26). Kuten Little (2005, 325) toteaa, Eurooppalainen viitekehys toimii välineenä, jonka avulla voidaan paitsi linkittää kielitestejä myös suunnitella sekä yleisiä opetussuunnitelmia että henkilökohtaisia opiskeluohjelmia. Kielitaidon arvioinnin alalla eurooppalaisen kehittämistyön näkyvimpinä tuotoksina voidaan pitää sekä Euroopan unionin komission rahoittamaa DIALANG-hanketta että Euroopan neuvoston Eurooppalaista kielisalkkua, joihin molempiin itsearviointi kuuluu olennaisena elementtinä. Näitä molempia käsitellen myöhemmin tässä luvussa.

### 3.1.2 Oppijan autonomia elinikäisessä oppimisessä

Gremmo ja Riley (1995, 152) totesivat jo melkein 20 vuotta sitten, että kyky ajatella ja toimia itsenäisesti on aina ollut ominaisuus, jota valtaosa erilaisten yhteiskuntien ihmisistä on arvostanut. Tutkijat muistuttavat samalla kuitenkin, että todellisuudessa näiden kykyjen harjoittaminen on ollut usein mahdollista vain yhteiskuntien eliittiin kuuluville henkilöille. Oppijan autonomiaa edistävälle oppijakeskeisille kielenopetuksen suuntauksille on tyypillistä se, että ne vaativat oppijoita tekemään oman oppimisensa tavoitteita, sisältöä ja oppimismetodeja koskevia päätöksiä (Little 2005, 321). Itsearvioinnin tärkeys oppijan autonomialle selviää lukemalla joidenkin tutkijoiden kannanottoja, joiden mukaan oppijan autonomiasta tai elinikäisestä oppimisestä ei oikeastaan voida edes puhua puhumatta itsearvioinnista (Barbot 1990; Oscarson 1997). Jotkut tutkijat ovat luetelleet tekijöitä, joilla on vaikutusta oppijoiden itsenäisenä oppijana toimimiseen (*learner autonomy*). Muun muassa Wenden (1991) listasi seuraavat seitsemän tekijää, jotka vaikuttavat siihen, miten oppija kokee itsearvioinnissa edellytettävän oppijan autonomisen roolin. Tekijät ovat seuraavat:

- sosiaalistamisprosessi
- ristiriitaiset odotukset
- roolien monimutkaisuus
- metakognitiivisen tiedon puute
- opittu avuttomuus
- itsearvostus (*self-esteem*)
- minäkuva.

Oppijan autonomian kehittämiseen ei tutkimusten mukaan riitä pelkästään se, että opettaja uskoo oppijakeskeisiin oppimissuuntauksiin, vaan on seurattava tietynlaista prosessia ennen kuin oppija on halukas ottamaan vastuun omasta oppimisestaan. Prosessin käynnistäjänä toimii yleensä opettaja, jolla on tärkeä rooli itsearvioinnin käyntiinpanijana. Viitekehys (2004, 199) määrittelee opettajan uuden roolin siten, että tämän tulisi toimia opetustilanteessa ohjaajana ja auttajana, joka kuuntelee, mitä oppilaat sanovat opiskelustaan ja oppimisestaan ja reagoi siihen sekä koordinoi oppilaiden työskentelyä heidän toimintansa seuraamisen ja siinä opastamisen lisäksi (EVK 2004, 199). Oscarson (1989, 3) toteaa, että juuri itsearvioinnin yhteydessä toteutettu opettajan ja oppijan välinen yhteisvastuu on mainio esimerkki kielenopetuksen kehittämisen demokratisoitumisesta (Oscarson 1989, 3). Yhtenä ensimmäisistä tehtävistä mainitaan, että opettajan tulee selvittää mahdollisimman tarkkaan oppijan koulutustausta ja sen vaikutus hänen vieraiden kielten oppimiseensa. Tämän selvitystyön takana on pyrkimys oppimaan oppimiseen eli toisin sanoen siihen, että oppija on kiinnostunut osallistumaan oman oppimisprosessinsa suunnitteluun, tavoitteiden asetteluun ja arviointiin niin, että se mahdollistaa juuri hänen kohdallaan tehokkaimman oppimisen.

Boud (1981) on todennut jo 1980-luvulla, että itsearvioinnin onnistunut käyttöönotto edellyttää sitä, että uskotaan oppijan autonomiaan opetuksellisena päämääränä. On muistettava, että opettajat eivät suinkaan aina ole varauksettomasti sitä mieltä, että oppijoiden autonomiaan tähtäävä toiminta olisi helposti toteutettavissa. Syitä tähän voi tietenkin olla monia. Mahdollisista syistä voi saada jonkinlaisen käsityksen tarkastelemalla esimerkiksi kielikasvatuksen kehittämisen- ja tutkimushanke OSKUn yhteydessä kerättyjä opettajien mielipiteitä oppilaan itseohjautuvuudesta ja itsearviointitaidoista. Tutkimuksessa näet selvisi, että vaikka tutkimuksessa mukana olleet opettajat pitivät opiskelijoiden autonomiaa tärkeänä, he eivät aina tienneet, kuinka tai missä vaiheessa oppijoille olisi parasta tarjota mahdollisuuksia autonomian kehittämiseen. Oppijoiden ikä näyttää olleen merkittävä tekijä, kun opettajien mielipiteitä järjesteltiin ryhmiin heidän vastaustensa perusteella. Jotkut opettajat uskoivat, että ikä ei vaikuta siihen, kehittykö oppijasta autonominen, kun taas toiset opettajista olivat vastakkaista mieltä: heidän mukaansa vain vanhemmat oppijat voivat kehittyä autonomisiksi. Osa tutkimukseen osallistuneista opettajista piti opiskelija-autonomian tavoitteita hyväksyttävänä, mutta tavoitteisiin sitoutumista liian vaativana itselleen (Salo & Hildén 2011, 24–25.)

Usein itsearvioinnista puhuttaessa korostuu oppijaa hyödyttävä puoli, toisin sanoen se, että oppija oppii reflektion ja muilta (useimmiten opettajalta ja muilta oppijoilta) saamansa palautteen avulla tuntemaan paremmin itseään oppijana. Tällöin kuitenkin unohdetaan, että myös opettajalle voi olla suurta hyötyä oppijan itsearvioinnista. Opettaja oppii tuntemaan oppilaitaan yksilöinä huomattavasti paremmin esimerkiksi juuri itsearvioinnissa käytettävien menetelmien ja siihen liittyvän vuoropuhelun ansiosta (Kohonen n.d., 2). Tudor (1996) korostaa itsearviointikoulutuksen merkityksen lisäksi sitä, että koulutus on vain välivaihe ennen oppijan sitouttamista (*learner involvement*) ja että perimmäisenä päämääränä on oppijan voimaannuttaminen (*learner empowerment*). Toisaalta on syytä huomioida, että näkökannat itsearviointia kohtaan ovat muuttuneet vuosikymmenien kuluessa. Vielä 1990-luvulla esiintyi näkemyksiä, joiden mukaan itsearviointi oli hyödyllistä vain oppijan kannalta. Näissä kommentteissa todettiin, että itsearvioinnissa ei ollut tärkeää oppia arvioimaan omaa osaamistaan ”oikein” eli tavalla, joka tuottaisi samanlaisia tuloksia, kuin jos opettaja arvioisi oppijan osaamista. Kannanotoissa korostettiin sitä, että itsearvioinnin ei olekaan tarkoitus olla jonkun ulkopuolisen tahon toteuttaman arvioinnin kopio. (Carton 1995.)

Viitekehyksessä (EVK 2004, 199–200) pohditaan uudenlaisen opettajanroolin lisäksi sitä, missä määrin kielenoppijoilta, joista pyritään tekemään autonomisia oman oppimisensa vastuuhenkilöitä, odotetaan tai edellytetään kykyä tehdä muun muassa seuraavia asioita (sulkeissa olevat kommentit tutkimuksen tekijän):

- osallistua aktiivisesti oppimisprosessiin; tavoitteiden ja työtapojen sopiminen yhteisesti opettajan ja/tai oppilastovereiden kanssa (yhteisöllinen oppimiskäsitys, motivointi, sitoutuminen, vuorovaikutuksellisuus)
- suostua kompromisseihin (yhteistyö- ja neuvottelutaidot)
- osallistua oppilastovereiden opettamiseen ja vertaisarviointiin (toisilta oppiminen, kielen oppijana kehittyminen vuorovaikutuksessa)
- työskennellä itsenäisesti itseopiskeluun tarkoitetun oppimateriaalin avulla (itseohjautuvuus).

Oppijan autonomiaa ei kuitenkaan tule tarkastella yksisilmäisesti keskittyen autonomia-sanan siihen merkitykseen, joka korostaa oppijan yksin toimimista. Lopullisena päämääränä toki on oppijan autonomisuus, toisin sanoen se, että hän pystyisi myöhemmässä vaiheessa elämässään ottamaan vastuun omasta oppimisestaan (= oppimaan oppiminen). Autonomia ei siis ole automaattista, vaan se vaatii harjoitusta. Siksi harjoitteluvaiheessa ja missä tahansa vaiheessa, kun on kyse uuden asian oppimisesta, korostuvat parhaimmassa tapauksessa vuorovaikutuksellisuus ja toisilta oppiminen, oli kyseessä sitten oppijoiden tai opettajan ja oppijoiden välinen vuorovaikutus. Monien tutkijoiden mukaan yhdessä oppiminen tai tekeminen on luontevaa kaikenikäisille; sekä lapset että aikuiset oppivat toisten ihmisten kanssa, ja tässä yhteistyössä kehittyvät samalla erilaisia yhdessä oppimisen muotoja (esim. Vygotski 1982; EK 2011). 2000-luvulla oppimisen suunnittelun ja tutkimuksen yhteydessä on käytetty käsittei-

tä *sosiaalinen vastuu* ja *sosiaalinen toimijuus*. Niillä ei tarkoiteta pelkästään oppijan omaa aktiivista roolia oppimistapahtumassa vaan myös sitä, että oppijat jakavat aktiivisesti tietojaan omassa oppimisyhteisössään ja toimivat oppijan roolissa koko ajan. Joskus he toimivat yksin ja joskus ryhmässä erityyppistä oppimista toteuttaen ja samalla oppimisyhteisöään ja -ympäristöään muovaten (Université Laval 1996; Lantolf & Thorne 2006). Autonomia-käsite määritelläänkin nykyisin oikeamminkin sosiaaliseksi kuin yksilökeskeiseksi käsitteeksi, ja autonomian sanotaan pohjautuvan vygotskilaisen sosiokulttuurisen teorian sisältämään yhteisvastuullisuuden ja yhteisöllisyyden periaatteisiin (Hildén 2011, 20).

Nykyisin usein korostettu ajatus *elinikäisestä oppimisesta tai kehittämisestä* edellyttää oppijalta omaa aktiivista asiantuntijuutta ja ohjattua itseohjautuvuutta (esim. EK 2011). Kyky itsearvioida on tässäkin keskeistä. Elinikäinen oppiminen edellyttää sitä, että oppija osaa toisaalta asettaa itselleen tavoitteita ja toisaalta arvioida säännöllisesti sitä, onko hän saavuttanut asettamansa tavoitteet. Itsearviointilla sanotaankin olevan kauaskantoisia vaikutuksia oppijan arviointiin sitoutumiseen, ja siten sen arvoa ei tulisi jättää huomiotta (Oscarson 1999, 165). Joidenkin nykytutkimusten mukaan koulun tulisikin valmentaa oppilaitaan aikaisempaa paremmin elämää varten osoittamalla, että nyt koulussa opittavat taidot on pääasiassa tarkoitettu muodostamaan tietynlainen pohja, jolle jatkaa ja syventää osaamistaan eri oppiaineissa pakollisen koulunkäynnin jälkeen (Perrenoud 2011). Suomalaisissa perusopetuksen ja lukio-opetuksen opetussuunnitelman perusteissa mainitaankin elinikäisen oppimisen kasvatustavoite, mitä voidaan pitää oppilaan autonomiaan liittyvänä (Hildén 2011, 10). Oppijan vastuu omasta oppimisestaan on tärkeää, sillä siinä on kysymys valmiina annetun tiedon kyseenalaistamisesta, ja sen perimmäisenä tarkoituksena on tarjota kaikille kansalaisille mahdollisuus sosiaaliseen ja yksilölliseen kehittymiseen jatkuvan tiedon avulla (Springer 1996, 16). Selvää onkin, että nykyisin ikääntyvän väestön määrän kasvaessa elinikäisestä oppimisesta tulee entistäkin tärkeämpää. Muun muassa näistä syistä itsearviointia pidetään välttämättömänä koko elämän kestävässä oppimisessa (esim. Boud 1995).

### 3.1.3 Itsearviointi muiden arviointitekniikoiden joukossa

Kielitaidon itsearviointi liittyy 1980-luvulla yleistyneeseen konstruktivistiseen oppimisenäkemykseen, jonka mukaan oppiminen on sitä tehokkaampaa ja motivoivampaa, mitä enemmän oppija itse osallistuu aktiivisesti koko oppimisprosessiin. Konstruktivisessa oppimisenäkemyksessä oppija on oppimisen keskiössä ja oppijan itsenäisen toiminnan ja vastuun jakamisen tärkeyttä korostetaan koko oppimisprosessin ajan. Näkemys korostaa myös oppijaa yksilönä, joka konstruoi aktiivisesti tietoa. Itsenäistä oppijuutta, yhteisvastuuta, yhteistyötä ja vähitellen oppijan rekonstruoimaa tietoa tarkasteltaessa voidaan todeta, että on siirrytty lopputuotteen tarkastelusta itse prosessin tarkasteluun (esim. Gipps 1995). Blanchen (1988, 75) mukaan jaettu vastuu esimerkiksi osaamisen arvioinnista saa aikaan sen, että oppija ei ole opettajan mielipiteen varassa, ja toisaalta oppija pystyy myös viestittämään paremmin opettajalle omat oppimistarpeensa.

sa. Oppiminen ei voi olla oppijan näkökulmasta tehokasta, ellei hän tiedä, mikä hänen taitonsa ovat, kuinka paljon hän on edistynyt ja mitä hän osaa (tai ei osaa) oppimillaan taidoilla tehdä (Blanche 1988, 75). Itsearviointin käyttäminen mahdollistaa sen, että oppijat saadaan tietoisiksi muun muassa edellä mainituista seikoista.

Jaetun vastuun lisäksi monet tutkijat korostavat itsearvioinnissa sitä, että opettajaa sitoo osallistamisen, aktivoinnin ja vastuun etiikka, minkä pohjalta opettajan voidaan odottaa pohtivan ja käyttävän muitakin arviointitapoja kuin vain perinteistä opettajan tekemää arviointia. Myös oppijan oikeus saada vaikuttaa oman oppimisensa tavoitteenasetteluun sekä tavoitteiden ja tulosten vertailuun on nostettu esille silloin, kun on perusteltu itsearviointin käyttämisen hyötyjä. (Atjonen 2007, 81.) Samanaikaisesti kun on puhuttu vaihtoehtoisista arviointitavoista perinteisten rinnalla, jotkut tutkijat ovat muistuttaneet, että vaihtoehtoinen arviointi on usein opettajille aivan uutta. Monilla heistä on vain vähän tai ei ollenkaan arviointikoulutusta, ja siten heiltä puuttuu olennainen mittaamiseen liittyvä tieto (Stiggins 1991). Onkin jossain määrin paradoksaalista, että samaan aikaan kun itsearviointia kritisoidaan siitä, ettei se täyttäisi arviointilta edellytettäviä yleisen objektiivisuuden vaatimuksia, kielenopettajan uskotaan pystyvän tekemään tarkkoja arvioita oppijan kielitaidosta, vaikka opettajalla ei olisi minkäänlaista kielitaidon arviointiin liittyvää koulutusta. Tosiasia tietenkin on, että ihan yhtä lailla kuin oppija tarvitsee harjoitusta itsearviointiin, myös opettaja tarvitsee (mieluiten jatkuvaa) koulutusta kielitaidon arviointiin.

Itsearviointi koostuu seuraavista kolmesta elementistä:

1. kohdekielisten tilanteiden analyysi  
→ mitä oppijoiden pitää osata tehdä kielellä?
2. tämänhetkinen tilanneanalyysi  
→ mitä oppijat osaavat tällä hetkellä tehdä?
3. strateginen analyysi  
→ miten oppijat saavuttavat parhaiten asettamansa oppimistavoitteet?  
(Tudor 1996, 162).

Vastauksia näihin kysymyksiin sekä monipuolista tietoa oppijan kielen opiskeluun liittyvästä lähtötilanteesta, hänen vahvuuksistaan tai edistyksestään kieliopinnoissa saadaan tavanomaisen summatiivisen arvioinnin lisäksi diagnostisen, formatiivisen, dynaamisen tai muiden niin kutsuttujen vaihtoehtoisten oppijakeskeisten arviointitekniikoiden avulla. Itsearvioinnin yhteydessä mainitaan usein diagnostinen, formatiivinen ja dynaaminen arviointitekniikka, joiden tarkoituksena on olla mukana jatkuvassa opettajan ja oppijan välisessä palaute-prosessissa. Formatiiivista *kehittävää arviointia*, jota käytetään esimerkiksi opetuksen kehittämishankkeiden tai koulutusohjelmien arvioinnissa, ei tavallisesti lueta itsearviointin kanssa yhteisestä lähestymistavasta huolimatta kielitaidon arviointiin; sitä ei yleensä lasketa vaihtoehtoisten arviointimenetelmien joukkoon. Räisäsen mukaan (2005, 110–112) kehittävä arviointi, aivan kuten itsear-



viointikin, edustaa voimaannuttavaa arviointia. Siinä korostuu vahva sitoutuminen arviointiin. Sen ensisijaisia tuloksia ovat reflektioon perustuva oppiminen, kokemus, oppimisen avulla syntyvä uusi ajattelu ja eri tahojen välinen yhteistyö. Kehittävä arviointi on ei-perinteistä arviointia: arviointitietoa tuotetaan hankkeen aloitusvaiheesta lähtien eikä vain hankkeen loppuvaiheessa. Kehittävä arviointi pyrkii nimensä mukaisesti arviointikohteen kehittämiseen, ja sille on keskeistä käytännön toimijoiden osallistuminen arviointiin aivan kuten itsearviointissakin. (Korkeakoski & Silvennoinen 2008, 118, 123.)

Diagnostisen, formatiivisen ja dynaamisen arviointitavan määrittelyn ja toisistaan erottamisen on todettu olevan verrattain hankalaa, sillä näiden termien tarkasta sisällöstä ei vallitse yksimielisyyttä (esim. Huhta 2008). Yhteistä niille kaikille on kuitenkin ainakin se, että niissä itse oppiminen ja sen arviointi etenevät rinnakkain, sekä se, että niiden avulla pyritään saamaan riittävän ajoissa tietoa oppijan oppimisen heikkouksista ja vahvuuksista, jotta oppijan oppimisprosessia voitaisiin ohjata parhaiten oppijaa hyödyttävään suuntaan. Myös oppijat tuntuvat arvostavan diagnostisesta ja formatiivisesta arvioinnista saatavaa hyötyä. Elsisen (2007, 359) laatimassa opiskelijakyselyssä, joka liittyi Joensuun yliopiston kielikeskuksen opetuksen laatutyöhön, selvitettiin kielikeskusopiskelijoiden kokemuksia ja käsityksiä itse- ja vertaisarvioinnista sekä suhtautumista niihin. Kyselyyn vastanneiden opiskelijoiden (N=37) käsityksistä selvisi, että he pitivät kielitaidon arvioinnin pääasäntöisenä tehtävänä juuri diagnostista tai formatiivista tehtävää. Tämä huomio on tärkeä siksi, että ruotsin kielen yliopistollisten kielikeskusopintojen avulla osoitetaan julkishallinnon virkoihin vaadittava niin kutsuttu virkamiesruotsin taito. Siten odotettavaa olisikin ollut, että ainakin osa kyselyyn vastanneista opiskelijoista olisi maininnut julkishallinnon virkoihin vaadittavan kielitaidon osoittamisen olevan ruotsin kielen kielikeskusopintojen pääasiallinen tehtävä. Kukaan kyselyyn vastanneista opiskelijoista ei kuitenkaan pitänyt tätä kielitaidon osoittamisen virallista aspektia kielitaidon arvioinnin tärkeimpänä tehtävänä. (Elsinen 2007, 359.)

Dynaaminen arviointi on staattisen arvioinnin vastakohta. Se ei selvitä, mitä oppijat osaavat tehdä tällä hetkellä, vaan mitä he tulevat osaamaan tulevaisuudessa. Dynaamisessa arvioinnissa katsotaan siis eteenpäin eikä taaksepäin. Jotkut dynaamista arviointia tutkineet ovat sitä mieltä, että arvioinnista puhuttaessa olisi kaiken kaikkiaan syytä olla aikaisempaa enemmän kiinnostunut siitä, mitä oppijat osaavat tehdä silloin, kun heitä autetaan. Tutkijat korostavat, että arviointia ei tulisi lainkaan erottaa opetuksesta erilliseksi elementiksi, vaan että sen pitäisi olla opetuksen olennainen osa (Ellis 2005a; Poehner & Lantolf 2005). Poehner ja Lantolf (2005) tekevät eron formatiivisen ja dynaamisen arvioinnin välille siten, että näistä jälkimmäinen on heidän mukaansa systemaattista, oppijan pitkäaikaista kehittymistä tukevaa sekä opetuksen ja arvioinnin katkeamatonta yhteyttä korostavaa. Näiden arviointitekniikoiden yhteydessä sekä itsearviointilla että jatkuvalla palautteen annolla on tärkeä tehtävä oppijan oppimista ohjaavana tekijänä. Kohonen (1997, 13) mainitsee vielä itsearviointiin liittyvänä arviointitekniikan muotona *autenttisen arvioinnin*. Autenttiselle arvioinnille on tyypillistä muun muassa se, että se on erottamaton

osa opetusta ja siinä korostuvat oppijoiden yksilöllisyys sekä heidän edistymisensä ja vahvuutensa.

### 3.1.4 Itsearviointin reflektiivisyys ja dialogisuus

Oppijan omalla reflektoinnilla on ensisijainen merkitys itsearviointin onnistumisessa. Reflektointi- ja itsearviointitaidot mainitaankin usein yhdessä, kun kyse on näkökulmista, joihin olisi hyvä kohdistaa huomiota omaa opiskelua pohdittaessa. Itsearviointi ja reflektio tunnutaan mielleltävän joskus lähestulkoon samaa tarkoittavaksi – niin läheisesti toisistaan riippuvaisista käsitteistä on kysymys. Tämä kävi ilmi esimerkiksi tutkimuksesta, jossa selvitettiin sitä, mitä lukiolaisille ja opettajaopiskelijoille tuli mieleen sanoista itsearviointi ja reflektio (Alanen & Kajander 2011, 74). Tutkijat varoittavat kuitenkin, että itsearviointi-sanaa ei tulisi käyttää korvaamaan sanaa reflektio, koska se kaventaa ja vinouttaa reflektio-sanana merkitystä. Nämä kaksi sanaa eivät ole toisensa synonyymejä, vaan ”reflektio on onnistuneen itsearviointin edellytys” (Alanen & Kajander 2011, 78).

Reflektiivisyydellä tarkoitetaan Atjosen (2005, 82) mukaan valmiuksia ymmärtää ja pohdiskella omaa toimintaa eri näkökannoilta sekä ohjata ja muuttaa sitä. Hildén (2011, 10) huomauttaa, että autonomiaan kasvattamisessa tarvitaan reflektiivistä toimintaa ja tietoisuutta siitä, mitä ollaan tekemässä. Myös Kohonen (1999) on sitä mieltä, että reflektio ei koske vain oppimisprosessia, vaan se voi koskea esimerkiksi oppimisen sisältöjä, työmuotoja, oppimisympäristöjä tai -konteksteja tai vaikkapa omia tunteita. Kohosen (ei vuotta, 8) mukaan itsearviointi ja reflektointi ovatkin toisiinsa kietoutuneita ominaisuuksia. Baxter Magolda (1992, 106) listaa viisi oppimisympäristöön liittyvää avainaspektia, joista suurin osa mainitaan itsearviointin yhteydessä. Avainaspektit ovat oppijan rooli, vertaisoppijan rooli, opettajan rooli, arviointi ja tiedon luonne. Tutkija toteaa, että oppijan käsitykset mainituista aspekteista ja niiden liittymisestä oppimiseen vaihtelevat sen mukaan, kuinka oppijat hahmottavat tietoa (tiedon hahmottamistavoista ks. tarkemmin luku 3.5).

Reflektointi on mekanismi, jossa oppija ottaa itsensä, omat kokemuksensa ja tietonsa kriittisen analyysin ja tarkastelun kohteeksi. Tutkijat tähdentävät usein, että oppijat on saatava sitoutettua siihen, että he omaksuvat kriittisen reflektointikyvyn. Tämä sitouttaminen sisältää yhtä hyvin ammatillisen arviointikyvyn kuin eettisen tietoisuudenkin, ja sitä pidetään yleisesti tärkeänä osana oppimaan oppimista. (Lucas & Tan 2013, 104.) Reflektion välttämättömyyttä osoittaa se, että sitä pidetään oppimisen ehtona:

”Oppimista ei synny ilman *reflektiota* eli kriittistä yhdessä arviointia, joka koskee sekä itse toimintaa että sen lopputulosta.” (Uusikylä & Atjonen 2005, 149.)

Reflektointi-käsitettä on lähestytty tutkimuksessa myös tutkimalla sitä, millä tasolla oppijat ovat reflektoinnissaan. Huttusen (2000) luokittelussa oppijan reflektointikyvyssä on kolme tasoa, jotka ovat mekaaninen (ei omaa analysointia,

faktatiedon luettelua), pragmaattinen (oman oppimisen pohdinta ja kriittisyys) ja emansipatorinen (opetuksen ja oppimisen näkeminen jatkuvana prosessina). Reflektiossa ei ole kyse siitä, että oppija pohtisi ja analysoisi oppimistaan yksin. Oppimista edistetään yhteistyössä opettajan ja oppilastoverien kanssa, jolloin oppimista voidaan esimerkiksi sanallistaa ajattelemalla ääneen tai kirjoittamalla. Usein tällaisissa harjoituksissa opettaja ja/tai oppilastoverit toimivat auttajina tai palautteen antajina. (Uusikylä & Atjonen 2005, 149.) Reflektointia toteutetaan jonkin verran oppimispäiväkirjojen avulla. Useimmiten opettaja kommentoi oppijan oppimispäiväkirjaa, jolloin saadaan aikaiseksi luonteva vuoropuhelu oppijan ja opettajan välillä. Jotta reflektoinnista voitaisiin saada eniten hyötyä, oppijan tulee joka tapauksessa saada jonkinlaista palautetta reflektoinnistaan. *Palaute* onkin käsite, joka liittyy läheisesti sekä reflektointiin että itsearviointiin. Hattien ja Timperleyn (2007, 102) mukaan palautetta voidaan pitää yhtenä kaikista tärkeimmistä tekijöistä oppijan oppimisen kannalta. Palautteen avulla pyritään tutkijoiden mukaan täyttämään oppimisprosessiin kuuluvat "aukot" ("mikä on tilanteeni?", "minne olen menossa?") sekä tekemään mahdolliset vaihtoehdot näkyviksi, jotta oppija voi edetä oppimisessaan ("minne menen seuraavaksi?").

Opetuksen *dialogisuus* puolestaan ilmenee siten, että oppijalle annetaan opetus- ja oppimistilanteessa tilaa ja valmiuksia toimia tasavertaisena keskustelijana ja aktiivisena toimijana opettajan rinnalla. Dialogisuuden lähtökohdat ovat vastakohtaisia perinteisen opettajajohtoisen opetustavan kanssa. Dialogisuuteen kuuluu siis se, että oppija on keskustelijana samalla tasolla kuin hänen opettajansa, ei alemmalla. Itsearviointi määritelläänkin usein oppijan ja opettajan väliseksi ponnistukseksi kohti yhteistyötä (Nunan 1988, 2). Arvioinnin tulisi parhaimmillaan olla opettajan ja oppijan välinen vastavuoroinen dialogi, jossa ei pakoteta toista hyväksymään omaa kantaa, vaan lopputulokseen päädytään yhdessä neuvotellen (Atjonen 2007, 84). Dialogisuuden tavoitteena opetus- ja arviointitilanteissa on siis se, että opetukseen ja / tai arviointiin otetaan mukaan kaikki ne, joita asia koskee (Takala 1994, 5).

Littlen (1991, 5) mukaan kykymme oppia itseopiskelua kehittyä luultavasti sen kokemuksen pohjalta, joka syntyy oppimistilanteessa vuorovaikutuksessa toisten kanssa. Voidaankin todeta, että itsearviointiin läheisesti liittyvä vuoropuhelu opettajan ja oppijan välillä sekä mahdollinen vertaisarvioinnin käyttö auttavat molemmat oppijaa siinä, että hänestä tulee autonominen ja omasta oppimisestaan vastuunsa kantava aktiivinen toimija. Itsearvioinnin sanotaankin olevan tärkeä keino herättää oppilaita passiivisuudesta, joka on koulutyöskentelylle usein ominaista (Atjonen 2007, 82). Joidenkin tutkijoiden (esim. Carton 1995) mukaan itsearvioinnista puhuttaessa keskityttiin ainakin aikaisemmin liikaa keskustelemaan osaamisen siirtämisestä. Tämän näkemyksen mukaan opettajan osaaminen ikään kuin siirrettäisiin oppijalle, joka sitten toteuttaisi perinteisesti opettajalle kuuluvia tehtäviä. Carton huomauttaa, että itsearviointia ei voida tarkastella siitä näkökulmasta, että opettajan osaamista siirrettäisiin oppijoille, sillä oppijoilta ei luonnollisestikaan voida edellyttää samaa osaamista kuin pedagogisen koulutuksen saaneilta opettajilta. Kuten edellä on jo tullut

esille, itsearviointi käsittää oppijan määrittelemän itsearvioinnin kohteen, joka perustuu etukäteen päätetyille oppimistavoitteille, toimintatavoille, käytettäville arviointikriteereille ja lopuksi oman oppimisen onnistumisen arvioinnille (Cuq 2003, 30).

Itsearviointi korostuu niin kutsutuissa moderneissa oppimisympäristöissä, joissa opettajan rooli ei enää perinteisesti ole siirtää tietoa oppijalle, vaan joissa oppija konstruoi oman tietonsa jo hallitsemansa tietoaineksen pohjalle. Nopea teknologinen kehitys vaikutti oppijan autonomiaa koskevien ajatusten leviämiseen (Gremmo & Riley 1995, 153). 1970- ja 1980-luvuilla perustetut niin kutsutut oppimisen resurssikeskukset, joihin oli kerätty keskitetysti tietoa oppijoiden käyttöön, edustivat aikanaan uusia oppimisympäristöjä. Resurssikeskusten asiakaskunta oli hyvin kirjavaa: oli aikuisia, jotka tavoittelivat ammattitutkinnon suorittamista ja nuoria, jotka pyrkivät parantamaan yleissivistävän koulutuksen piirissä saamiaan arvosanoja (Springer 1996, 18). Vaikka oppimisen resurssikeskusten ideologia pohjautuukin oppijan autonomialle, on todettu, että oppijat tarvitsevat tässäkin kontekstissa ohjausta, jotta heidän oppimisensa olisi tehokasta (Gremmo & Riley 1995, 157). 2000-luvulla resurssikeskusten käyttö on kuitenkin vähentynyt muun muassa tietotekniikan yleisen kehittymisen, yleistymisen ja saatavuuden vuoksi. Nykyisin oppimisympäristöön kuuluviksi lasketaan opettajan ja oppijoiden lisäksi myös erilaiset opetusmateriaalit ja -välineet. Meisalon ym. (2003, 77) mukaan oppimisympäristö-käsitettä käytetään erityisesti silloin, kun oppijan aktiivinen rooli nostetaan tarkastelun keskipisteeseen.

### 3.1.5 Itsearvioinnista DIALANG-järjestelmässä

Tämän tutkimuksen aineisto ja tutkimusasetelma pohjautuvat suurelta osin DIALANG-järjestelmän kehittämiseen, ja tästä syystä keskityn tässä luvussa järjestelmän itsearviointia koskevan osuuden kuvaamiseen.

Itsearviointi on kuulunut alusta alkaen DIALANG-järjestelmään, sillä järjestelmän tarkoituksena on antaa oppijoille palautetta heidän kielitaitonsa vahvuuksista ja heikkouksista, opastaa heitä oppimaan lisää sekä tulemaan tietoisiksi kielen oppimisen ja kielitaidon eri aspekteista (Huhta & Figueras 2004, 65). DIALANG-hankkeella oli 1990-luvun alussa mahdollisuus kehittää itsearviointityökalu, jossa hyödynnettiin Northin (1995) tutkimuksessa jo aiemmin pilotoituja, validoituja ja kalibroituja englanninkielisiä kielitaitoa kuvaavia väittämiä tai kuvaimia. DIALANG-järjestelmän itsearviointiosuuteen valittiin joukko Viitekehysten väittämiä, jotka muokattiin itsearviointiin soveltuviksi tarvittaessa yksinkertaistamalla kuvaimia siten, että ne olisivat ymmärrettäviä kenelle tahansa kielen oppijalle. Tämän lisäksi luotiin jonkin verran myös uusia kuvaimia, jotta saatiin riittävän laajasti katettua kielitaidon eri aspektit ja tasot. Alkuperäiset englanninkieliset kuvaimet käännettiin itsearviointia koskevan kehittämistyön yhteydessä 13 kielelle, sillä tarkoituksena oli, että jokainen oppija voi tehdä itsearviointia sillä kielellä, jonka itse tunsivat hallitsevansa parhaiten. Käännettyjä kuvaimia myös pilotoitiin testitehtävien pilotoinnin yhteydessä, analysoitiin tilastollisesti ja muutettiin itsearviointikuvaimiksi DIALANG-

järjestelmän käyttöön (Huhta & Figueras 2004, 67). Lopputuloksena oli 18 kuvainta sisältävä itsearviointityökalu kolmeen eri osataitoon (lukeminen, puheen ymmärtäminen ja kirjoittaminen).

DIALANG-järjestelmässä vapaaehtoinen itsearviointi suoritetaan heti sanastotestin jälkeen, joka sekin on vapaaehtoinen. Järjestelmä varoittaa oppijaa, jos tämä ei aio tehdä sanastotestiä ja/tai itsearviointia, sillä oppijalle esitettävien tehtävien vaikeustaso perustuu näiden molempien yhdistettyihin tuloksiin. Myöskään testituloksen ja itsearvioinnin vertailu ei ole mahdollista, mikäli oppija ei suorita DIALANG-järjestelmän itsearviointiosuutta. Sanastotestin ja itsearvioinnin tarkoituksena on tehdä oppijan testiin osallistumiskokemuksesta mahdollisimman mielekäs erityisesti vaikeustason osalta (Alderson & Huhta 2005, 305). Oppija valitsee ensin haluamansa osataidon ja sen jälkeen vastaa kyseistä taitoa koskeviin 18 itsearviointikuvaimen joko ”kyllä” tai ”ei”. ”Kyllä”-vastaus valitaan, jos oppija on sitä mieltä, että hän selviytyy kuvaimessa kuvatusta tilanteesta ja ”ei”-vastaus taas silloin, kun oppija ei usko selviytyvänsä kuvaimen esittämästä tilanteesta. Tämän jälkeen järjestelmä laskee oppijan Viitekehysten mukaisen taitotason, raportoi sen testitehtävien suorittamisen jälkeen ja vertaa taitotasoa testisuorituksen tulokseen (Alderson & Huhta 2005, 303).

DIALANG-järjestelmän itsearviointi liittyy erottamattomana osana järjestelmän toiseen innovatiiviseen osaan eli oppijalle annettavaan palautteeseen. Itsearviointia pidetään järjestelmässä oppimistyökaluna, kyseessä ei siten olekaan pelkkä kielitaidon testaustyökalu (Alderson 2005, 97). Itsearviointi tehdään järjestelmässä ennen varsinaisten testitehtävien suorittamista ja session lopuksi oppijalle annetaan Viitekehysten kuuteen taitotasoon perustuvaa palautetta siitä, vastasiko hänen tekemänsä itsearvioinnin tulos sitä tulosta, jonka hän sai testitehtävien suorittamisesta. Mikäli testitulos ja itsearvioinnin tulos eroavat toisistaan, oppijalle annetaan ohjaavaa palautetta, jonka tarkoituksena on saattaa oppija tietoiseksi oman oppimisensa eri aspekteista. Tässä vaiheessa oppijalle kerrotaan myös oman kielitaidon yli- tai aliarvioimiseen liittyvistä riskeistä ja häntä kehoitetaan tutustumaan tarkemmin ”Tietoa itsearvioinnista” -osuuteen, jossa selostetaan mahdollisen testituloksen ja itsearvioinnin vastaimattomuuden syitä sekä annetaan yleistä kielitaitoon ja oppimiseen liittyvää informaatiota. Itsearvioinnin tarkoituksena onkin DIALANG-järjestelmässä parantaa oppijan itsetuntemusta (*self-awareness*), mitä nykyisin pidetään kielen oppimisen olennaisena osatekijänä. (Alderson 2005, 30–31.)

DIALANG-järjestelmä antaa myös muunlaista kuin testituloksen ja itsearvioinnin vastaavuutta koskevaa palautetta: suoritettujen testitehtävien tulokseen pohjautuvan kielitaidon taitotason Viitekehysten asteikolla, mahdollisuuden tarkastella yksityiskohtaisesti omien vastausten oikeellisuutta tai vääryyttä sekä sanastotestin tuloksen. Osallistujalle annetaan mahdollisuus tutustua saamansa taitotason kuvauksiin vertaamalla sitä kyseistä taitotasoa ylemmän ja alemman taitotason kuvaukseen. Palautteeseen sisältyy myös osio, jossa annetaan oppijalle vinkkejä siihen, kuinka hän voi edetä saamaltaan taitotasolta seuraavalle tasolle. DIALANG-järjestelmän ideologiaan kuuluu ajatus siitä, että

oppija on itse vastuussa arviointiprosessista DIALANG-arviointityökalua käyttäessään, ja siten arvioinnin tulokset ja arviointia koskeva palaute on tarkoitettu yksinomaan oppijan omaan käyttöön (Alderson & Huhta 2005, 302).

DIALANG-järjestelmässä toteutuvat Eurooppalaisen kielisalkun kaksi päätarkoitusta: pedagoginen ja raportoiva tehtävä. Pedagoginen tehtävä tarkoittaa sitä, että tuetaan oppilaan itsenäistä ja sosiaalisesti vastuullista kielten opiskelua sekä edistetään jatkuvaa kielenoppimista (Kohonen 2000, 22). DIALANG-järjestelmässä pedagoginen tehtävä tarkoittaa ainakin sitä, että oppija pyritään saamaan tietoiseksi itsestään oppijana ja kielen oppimiseen liittyvästä toiminnastaan muun muassa itsearvioinnin onnistumiseen liittyvän palautteen sekä muunkin järjestelmässä annettavan palautteen avulla (esimerkiksi neuvot siitä, kuinka kehittää kielitaitoaan eteenpäin). Kielisalkun raportoiva eli hallinnollinen tehtävä taas liittyy siihen, että oppijan on helpompi kertoa erilaisissa muutostilanteissa (esimerkiksi oppilaitoksesta toiseen siirryttäessä) omasta kielitaidostaan ulkopuolisille, jotka eivät vielä tunne oppijaa (oppijan kielitaidon dokumentointi ja arviointi; Kohonen 1997, 15). DIALANG-järjestelmän raportoiva tehtävä on huomattavasti rajallisempi kuin pedagoginen tehtävä, sillä DIALANG ei ole sellainen kielitutkintojärjestelmä, josta saa virallisen todistuksen. DIALANG-järjestelmän tarkoituksena on olla yksittäisen kielenoppijan työväline, joka ohjaa ja auttaa oppijaa pääsemään yhä pidemmälle kielitaidossaan.

### 3.2 Kielitaidon itsearvioinnin käyttötarkoituksia

Itsearviointia suositellaan usein muiden arviointitekniikoiden rinnalle täydentäväksi arviointimuodoksi (esim. Smith 1997). Tätä näkemystä tukevat käsitykset, joiden mukaan itsearviointi pohjautuu oppijan sisäiseen kokemukseen eikä pelkästään ulkoapäin tehtyyn havainnointiin tai esimerkiksi testaukseen. Itsearvioinnin avulla on mahdollista hankkia ainutlaatuista lisäinformaatiota arvioitavana olevasta piirteestä, oppimisesta tai osaamisesta. (Esim. Shrauger & Osberg, 1981.) Itsearvioinnin pitämistä muuta arviointia täydentävänä arviointina pidetään kuitenkin nykynäkemyksen mukaan jossakin määrin rajoittuneena, sillä itsearvioinnin avulla voidaan päästä tarkastelemaan sellaisia kielen prosessoinnin alueita, joita muuten voitaisiin havainnoida vain epäsuorasti. Itsearviointi voidaan aivan yhtä hyvin nähdä arviointimuotona, joka on - ainakin periaatteessa - ylivertainen sisältönsä kattavuuden kannalta verrattuna esimerkiksi johonkin kielitestiin. (Huhta 2003, 8.) Tämänkaltaisten näkemysten puolesta puhuvat sellaiset kannanotot, joissa korostetaan, että itsearviointi on tärkeää, koska siinä on kyseessä yksilön näkökulma itsestään. Oppijan itsearviointi mahdollistaa Oscarsonin (1989, 3) mukaan sen, että oppija saa hyödyllistä kokemusta arvioinnista. Uuden tiedon avulla päästään yleisellä tasolla havainnoimaan, mitä oppimisprosessissa tapahtuu, mikä taas puolestaan tuottaa arvokasta tietoa arvioinnin tueksi. Itsearvioinnin avulla on siis mahdollista saada

entistä tasapainoisempi kuva oppijan kielellisistä ja viestinnällisistä taidoista (esim. Tudor 1996, 165). Itsearviointiin korostuvat oppimisen laatua koskevat näkemykset ja siinä toteutuvat oppijakeskeiset arviointitekniikat. Itsearviointi tarjoaa siten erilaisen näkökulman siihen, kuinka oppijan kielitaito voidaan nähdä.

Itsearviointin itsenäisen, ei pelkästään muita arviointitapoja täydentävän arviointin puolesta puhuu muun muassa jatkuvasti ajankohtainen tieto siitä, että arvosanoilla ja arvioinnilla on väijäämättä vaikutusta oppijan minäkuvaan ja itsetuntoon. Atjonen (2007, 79–80) toteaa oppilaiden motivoituvan erityyppisten arviointien avulla pohtimaan jo tapahtunutta kehitystä ja oppimista. Tutkija mainitsee monipuolisen arviointin liittyvän eettisyyteen, jolla pyritään siihen, että kaikki oppilaat voivat menestyä, näyttää osaamisensa ja saada mahdollisuuksia itseluottamuksensa lisäämiseen.

Itsearviointia voidaan harjoitella, ja harjoittelun avulla voidaan ymmärtää vähitellen paremmin sitä, kuinka se toimii. Harjoittamisella saadaan mitä todennäköisimmin parannetuksi itsearviointin luotettavuutta ja käyttökelpoisuutta itsenäisen oppimisen työkaluna. Itsearviointia voidaan käyttää moniin eri tarkoituksiin, mutta se sopii kuitenkin parhaiten nk. *low stakes* -tilanteisiin, joissa oppija ei tunne olevansa kilpailutilanteessa jonkun toisen henkilön kanssa. Tällöin hänen ei tarvitse yrittää keinotekoisesti parantaa osaamistaan itsearviointin keinoin. Virallisen sertifiointin yhteydessä itsearviointiin suhtaudutaan perinteisesti epäillen, koska oppijalle voi tulla sitä suurempi kiusaus yliarvoida omaa osaamistaan, mitä enemmän hänelle on hyötyä itsearviointin positiivisista tuloksista. Little (2005, 324) totesikin vuonna 2005, että silloin ei ollut olemassa kirjallista näyttöä siitä, että testien käyttäjät hyväksyisivät itsearviointin niin kutsutuissa *high stakes* -tilanteissa tai -olosuhteissa. Roever (2001, 90) muistuttaa myös, että *high stakes* -arviointin lisäksi itsearviointi ei sovellu kovin hyvin myöskään ns. *medium stakes* -tilanteisiin. Niillä hän tarkoittaa sellaisia tilanteita, joilla on vaikutusta oppijan elämään, mutta ei kuitenkaan radikaalilla tavalla (esimerkiksi lukukausien välille sijoittuvat testit tai kokeet). Itsearviointia toteutettaessa ja dokumentoitaessa onkin huomioitava myös se, että itselle ja toisille kirjoitetut itsearviointit on tehty eri tarkoituksia varten, ja ne ovatkin siten erilaisia (esim. Huusko 2008; Saari 2002).

Kuten edellä selvitettiin, itsearviointi liittyy diagnostiseen, formatiiviseen, dynaamiseen tai joihinkin muihin niin kutsuttuihin vaihtoehtoihin oppijakeskeisiin arviointitekniikoihin. Itsearviointia käytetään oppimisprosessin eri vaiheissa: ennen sitä, sen aikana ja sen jälkeen. Ennen esimerkiksi tietyn kurssin alkua toteutettu itsearviointi edustaa diagnostista arviointia tyypillisimmillään; tarkoituksena on kartoittaa oppijan vahvuuksia ja heikkouksia jo ennen kurssin tai koulutuksen alkua. Oppimisprosessin aikana toteutettava itsearviointi taas on formatiivista arviointia, joka antaa mahdollisuuden pysähtyä ja tarkastella oppijan oppimisen kehitystä sekä sitä, mitä vielä pitäisi oppia, jotta päästäisiin edeltä käsin asetettuihin oppimistavoitteisiin. Oppimisprosessin aikana toteutettulla itsearviointilla on myös oppijan motivaatiota lisäävä tehtävä. Formatiiivisen ja muidenkin niin kutsuttujen vaihtoehtoisten arviointitapojen maksimaali-

nen hyödyntäminen ei synny muusta opetus- ja oppimiskontekstista erillään. Näiden arviointitapojen ja -ideologioiden tulee olla osa oppijan liittämistä opetuksen uudistamiseen harjoittamalla sekä itsearviointia että jatkuvaa palautetta oppijan ja opettajan välillä. (Black ym. 2003.)

### 3.3 Itsearviointimateriaaleja ja -metodeja

Lähtöoletuksena on, että itsearviointimateriaalien tulee olla selkeitä ja kohde-ryhmälle sopivia, toisin sanoen ne eivät saa olla liian vaikeita, jotta niiden käyttö olisi tehokasta (Oscarson 1980, 19–20). Konkreettisten itsearviointimateriaalien ja -metodien kirjo on melko laaja. Esittelen seuraavaksi lyhyesti joitakin erityyppisiä itsearviointimateriaaleja.

Itsearviointiin käytettävistä materiaaleista yleisimpiä ovat monitasoiset arviointiasteikot, väittämä- ja kuvainlistat sekä erityyppiset päiväkirjat (Huhta 2003). Introspektiivisten materiaalien, kuten esimerkiksi oppimispäiväkirjojen käyttö on usein liittynyt kielen oppimisen kognitiivisten prosessien tutkimukseen. Schärer (1983) muistuttaa kuitenkin oppijoiden, ja aivan erityisesti nk. ”heikkojen” oppijoiden, tarvitsevan paljon apua oppimispäiväkirjojen laatimisessa. Joissakin tutkimuksissa todetaan, että pojat eivät välttämättä ole aina kovin innokkaita osallistumaan reflektiiviseen enemmän tai vähemmän säännölliseen oppimispäiväkirjojen täyttämiseen. Samassa tutkimuksessa kuitenkin mainitaan, että jos oppimispäiväkirjojen täyttäminen on osa kielisalkkua, oppijoiden työskentely muuttuu tarkemmaksi ja reflektiivisemmäksi (Jaakkola ym. 2002, 49).

Eräs itsearvioinnin ruotsalaisista uranuurtajista on Oscarson. Hän kehitti jo vuonna 1978 kokeilumateriaalin, joka sisälsi erilaisia itsearviointilomakkeita. Yksinkertaisimmillaan itsearviointimittari voi olla sellainen, jossa kuvataan sanallisesti molemmat tietyn taidon ääripäät (esimerkki puhumisesta: ”Osaan sujuvasti englannin kieltä” ja ”En osaa puhua englantia lainkaan”; Oscarson 1980, 37). Tämän tyyppisen materiaalin heikkoutena on tietenkin se, että se jättää avoimeksi ääripäiden välillä olevien mittarin tasojen kuvaukset. Oscarsonin kehittämät materiaalit koostuvat muun muassa vieraan kielen käyttöön liittyvistä puhumisen tilannekuvauksista. Niissä oppijan on tarkoitus vastata, samoin kuin tässä tutkimuksessa, kysymykseen siitä, kuinka hyvin hän selviäisi kuvatussa puhumiseen liittyvästä kielenkäyttötilanteesta. Oppijan lomakkeessa on 10-portainen asteikko, jonka ääripäänä asteikon alapäässä on väittämä ”en pystyisi sanomaan yhtään mitään” ja asteikon yläpäässä ”pystyisin kuvaamaan yksityiskohtaisesti, mitä sanomalehden etusivulla on”. Materiaali sisältää myös lomakkeen, jossa on tämän tutkimuksen yleistasoarvion valintaan liittyvän kaltaisen asteikko. Sen tasoista oppijan tulee valita se taso, joka vastaa parhaiten hänen osaamistaan tietyssä kielen osataidossa. Myös ”kyllä”- ja ”ei”-tyyppisiä Osaan-itsearviointiväittämiä (*Can do*) on mukana Oscarsonin materiaaleissa. Kokeilumateriaalit toimivat hyvin, ja niiden pohjalta kehitettiinkin itsearvioin-



timateriaaleja edelleen myöhemminä vuosina (Oscarson 1997, 177). Myös Van Ek (1975) esitteli 1970-luvulla yksityiskohtaisen itsearviointityypin, joka testasi 25 väittämän avulla, oliko oppija jo saavuttanut *Threshold Level* -tasaisen kielitaidon.

Monet itsearvioinnissa käytettävät lomakkeet sisältävät asteittaista samanmielisyyttä osoittavia vastausvaihtoehtoja, kuten esimerkiksi "ei", "vähän", "kyllä" ja "paljon" tai "ei", "joskus" tai "usein". Väittämät voivat olla melko lyhyitä, kuten esimerkiksi "Englannin tunnilla pidän lukemalla oppimisesta" tai "Minusta on tärkeää kirjoittaa muistiin kaikki tunnilla" tai "Pidän parityöskentelystä". Tällaisia lomakkeita on käytetty esimerkiksi oppimisstrategioiden tarkastelussa (esim. Willing 1989). Myös monet muut ovat kehittäneet erityyppisiin itsearviointitilanteisiin soveltuvia materiaaleja. Näistä voidaan mainita esimerkiksi Tudor (1996, oppijan tarpeiden ja tavoitteiden kartoitus) sekä Harris (1997, kouluopetukseen tarkoitetut itsearviointimateriaalit).

Euroopan neuvoston eri projekteissa on kehitetty vuosikymmenien aikana monia erityyppisiä itsearviointimateriaaleja. Euroopan neuvoston projektityöryhmät ovat osallistuneet mm. *Can do* -itsearviointikuvainten laatimiseen. Kuvaimet oli tarkoitus liittää Eurooppalaisen kielisalkun sveitsiläiseen kokeiluversioon (EVK 2004, 295). Myös ALTE<sup>7</sup> siihen kuuluvine kielitutkintojärjestelmiseen on kehittänyt empiirisesti validoituja niin kutsuttuja *Can do* -väittämiä sisältäviä asteikkoja. ALTE:n asteikot on tarkoitettu sellaisille käyttäjille, jotka eivät ole kielitaidon arvioinnin asiantuntijoita. Asteikkoja voidaan ALTE:n mukaan käyttää mm. oppijoiden opettamisessa ja testaamisessa hyödyllisinä apuvälineinä, kun selvitetään esimerkiksi oppijan tietynhetkistä taitotasoa. ALTE:n *Can do* -väittämien määrä on tällä hetkellä noin 400 ja ne jakautuvat seuraavaan kolmeen yleiskategoriaan: sosiaalinen viestintä ja matkailu (*Social and Tourist*), työ (*Work*) ja opiskelu (*Study*). Yleiskategoriat jakautuvat vielä tarkempiin alakategorioihin, joiden tilanteet ALTE:n mukaan kattavat suurimman osan kielinoppijoiden tarvitsemista kielenkäyttötilanteista. Esimerkiksi sosiaalinen viestintä -kategoriassa ja matkailu-kategoriassa on ostosten tekemiseen, ravintoloissa ruokailuun ja hotellimajoittumiseen liittyviä alakategorioita. ALTE:n puhumista koskevat *Can do* -väittämät eivät ole erillisiä puhumisen osataidon väittämiä, vaan ne esiintyvät yhdessä puheen ymmärtämistä koskevien *Can do* -väittämien kanssa eri kategorioissa (ALTE).

Tunnetuimpia itsearviointiin käytetyistä materiaaleista on Eurooppalainen kielisalkku. Eurooppalaisen viitekehyksen kanssa samanaikaisesti kehitetyn ja sen ideologiaan nojautuvan Eurooppalaisen kielisalkun tarkoituksena on kehittää arviointikulttuuri, joka helpottaisi ja hyödyntäisi mahdollisimman hyvin oppijan itsearviointia (Little 2005, 321). Itsearviointiin soveltuvat materiaalit, kuten Eurooppalainen kielisalkku, tukevat jo edellä mainittua reflektiivistä oppimista, oppijan metakognitiivisten taitojen kehittymistä ja oppijan autonomiaa. Näissä kaikissa itsearviointi on keskeistä (Little 2005, 323; 2009).

Kielisalkkuarviointi edistää oppijan autonomisuutta mahdollistamalla opetuksen linkittämisen arviointiin, edistämällä reflektointia, auttamalla oppi-

<sup>7</sup> Association of Language Testers in Europe

joita ottamaan vastuuta omasta oppimisestaan, sekä mahdollistamalla sen, että oppijat huomaavat oppimisensa puutteet ja oppivat ottamaan riskejä (Ekbatani 2000, 6–7). Kielisalkun näkökulma oppimiseen on kokemuksellinen, aktiivinen ja sosiaalinen (Kantelinen & Hildén 2012, 2). Eurooppalaiseen kielisalkkuun tulee sisältyä kahdenlaista itsearviointia: yhteenvetomaista ja summatiivista (kielipassi-osa) sekä formatiivista (kielenoppimiskertomus ja työkansio -osat; Little 2005, 325–326). Itsearvioinnin lisäksi kielisalkkutyöskentely koskee kielitaidon vertaisarviointia (Kantelinen & Hildén 2012, 2). Kuten jo edellä esitettiin, Eurooppalaisella kielisalkulla on sekä pedagoginen että raportoiva tehtävä. Pedagogisen tehtävän tarkoituksena on tehdä oppimisesta oppijalle läpinäkyvää, edistää oppijan autonomiaa ja ohjata oppijan opiskelua. Raportoivan tehtävän tarkoitus on helpottaa oppijan kielitaidon esittelemistä muille raportoimalla oppijan oppimisprosessia ja sen tuloksia eri vaiheissaan ja muodoissaan (esimerkiksi ulkomailla osaaminen voidaan osoittaa käyttämällä tunnettua kriteeriviitteistä asteikkoa kansallisen asteikon sijaan; Kohonen 2000, 23; Kantelinen & Hildén 2012, 2).

### 3.4 Luotettavuudesta kielitaitoa koskevissa itsearviointitutkimuksissa

Tässä luvussa keskitytään tarkastelemaan lyhyesti sitä, millaisiin seikkoihin kielitaidon itsearviointia käsittelevät tutkimukset ovat yleensä painottuneet. Aluksi on todettava, että suurin osa tehdystä itsearviointitutkimuksesta koskee aikuisia kielenoppijoita. Tutkimuksista valtaosa sisältää tietoa puheen- ja tekstin ymmärtämisestä (Brantmeier ym. 2012, 145). Keskeisimpänä tutkimusaiheena on ollut itsearvioinnin luotettavuus eli se, missä määrin itsearvioinnista saatavat tulokset muistuttavat muiden arviointitapojen avulla saatuja tuloksia. Luotettavuus onkin tietynlainen avainkäsite, jonka valossa itsearviointia tarkastellaan sekä tässä luvussa että luvun alakohdissa.

Itsearvioinnin luotettavuus ja paikkansapitävyys on herättänyt keskustelua koko itsearvioinnin olemassaoloajan. Yleisesti ottaen vain ulkopuolisen toimijan, kuten esimerkiksi opettajan tai koulutetun kielitaidon arvioijan, tekemää arviota on totuttu pitämään ainoana oikeana ja luotettavana (Oscarson 1989, 2). Luotettavuutta tarkasteltaessa itsearviointia on ollut pakko verrata johonkin, ja siten vertailukohtana on yleensä ollut joko opettajan arvio tai esimerkiksi jostain kielitestistä saatu tulos. Itsearviointia, jota usein epäillään liiasta subjektiivisuudesta ja oman osaamisen yliarvioimisesta omaa etua silmällä pitäen, on kuitenkin käytetty menestyksekkäästi jopa sertifiointitilanteissa. Esimerkkinä tästä mainittakoon mm. englannin kielen opettajien jatkoopintoihin liittyvät tuutorien ja opiskelijoiden väliset keskustelut (Dickinson 1987) ja tiettyjen kouluasteiden pääsykokeet.

Kun itsearviointia koskevissa tutkimustuloksissa puhutaan korrelaatioista, viitataan yleensä siihen, miten oppijoiden suorittama itsearviointi vastaa jotain

ulkoista arviointia (opettajan arviointi tai kielitestin tulos). Kielitestin tai opettajan arvio voivat olla joko yksi- tai moniosaisia, ts. edustaa esimerkiksi vain yhtä kielitaidon aluetta tai montaa osa-aluetta. Jos tutkimuksissa saatu korrelaatio on korkea, se yleensä tarkoittaa sitä, että itsearviointi ja ulkoinen arviointi ovat samansuuntaisia, ja tällöin itsearviointia pidetään toimivana tai onnistuneena. Toisaalta tässäkin yhteydessä on syytä muistaa, että korrelaatioiden suuruusluokka on aina suhteellinen käsite; esimerkiksi korrelaatioarvo .70 tarkoittaa, että vielä noin puolet itsearvioinnin ja ulkoisen arvioinnin mittaamasta kielitaidon erosta johtuu jostain muusta kuin niiden samansuuntaisuudesta, esimerkiksi mittausvirheestä, huonosti toimineesta mittarista tai siitä, että itsearviointi ja ulkoinen arviointi mittaavat osittain eri asioita. Lisäksi myös itsearvioinnissa tutkimuksen tuloksiin voivat vaikuttaa yleiset tutkimuksen tekemiseen ja menetelmiin liittyvät tekijät, joita ovat ainakin käytetyt itsearviointi- ja analyysimenetelmät (esimerkiksi tietokoneella tai paperilla tehty itsearviointi).

Monet vuosikymmenien aikana tehdyt tutkimukset osoittavat, että oppijat onnistuvat tietyin ehdoin itsearvioimaan osaamistaan validilla ja luotettavalla tavalla (ks. esim. Evers 1981; von Elek 1982, Bachman & Palmer 1989; Blanche 1986; Oskarson 1981; Blanche & Merino 1989; Von Elek 1985; Smith 1997, Dragemark 2009). Kielitaitoa koskevista tutkimuksista useat ovat olleet nk. validointitutkimuksia, joilla on pyritty selvittämään itsearvioinnin luotettavuutta (Oscarson 1997, 177). Eräät itsearviointia sisältävistä tutkimuksista ovat olleet laajoja yhteishankkeita (kuten esim. ammattikorkeakoulukontekstiin liittyvä suomalainen AMKKIA-hanke), joissa on selvitetty ruotsin kielen opetukseen liittyviä asioita ja pyritty kehittämään opetusta ja arviointia; hankkeen tarkoituksena ei siis olekaan ollut pelkästään tuottaa tietoa itsearvioinnin luotettavuudesta. AMKKIA-hankkeen esitesteissä verrattiin ammattikorkeakouluopiskelijoiden itsearvioita ja esitesteissä saatuja arvosanoja. Tulos oli se, että opiskelijoilla oli melko realistinen käsitys omasta ruotsin kielen taidostaan. Esitestien pohjalta kuitenkin todettiin, että puhumisen osalta opiskelijoilla oli taipumusta aliarvioida osaamistaan (Juurakko & Takala 2001, 163–164; Juurakko-Paavola 2005, 246). Oscarson totesi jo vuonna 1978 tutkiessaan englantia vieraana kielenä opiskelleiden aikuisopiskelijoiden itsearviointitaitoa, että aikuisopiskelijat pystyvät arvioimaan melko tarkasti oman kielellisen osaamisensa, kunhan heillä oli käytössään valmiita skaaloitettuja itsearviointikuvauksia (Oscarson 1997, 178). Oscarsonin mukaan aikuisopiskelijat yleensäkin pystyvät arvioimaan vieraan kielen taitonsa suhteellisen oikeansuuntaisesti (Oscarson 1981, 234). Hän muistutti myös, että vaikka itsearviointi näyttää erityisesti aikuisilla antavan tyydyttäviä tuloksia suhteessa ulkoiseen arviointiin, oppijat tarvitsevat kuitenkin koulutusta itsearviointiin, jotta heidän arviointinsa olisivat tarkkoja. Suullisen taidon itsearviointia tutkittaessa on itsearvioinnin vertailukohtana lisäksi usein ollut jokin suullista kielitaitoa mittaava haastattelutesti (esim. FSI). Esimerkiksi Smith & Baldauf (1982) tutkivat haastattelun ja itsearvioinnin avulla saatujen arviointien validiteettia ja totesivat niiden olevan hyvin samansuuntaisia.

Itsearviointin luotettavuus on kytköksissä oppijan subjektiiviseen käsitykseen ja kuvaan omasta itsestään oppijana sekä hänen realistisiin käsityksiinsä omista oppimistuloksistaan. Kielen oppiminen on myös monitahoinen prosessi, johon vaikuttavat monet subjektiiviset tekijät, kuten affektiivisuus ja persoonallisuuspiirteet, joiden kaikkien sanotaan liittyvän oppijoiden itsearviointin luotettavuuteen (Blanche 1988, 75). Monet asiantuntijat pitävätkin yhtenä oppimisprosessin tärkeimmistä seikoista sitä, että oppija oppii tekemään luotettavia ja valideja päätelmiä omasta oppimisestaan (ks. Trim 1980; Atjonen 2007). Itsearviointin luotettavuuteen voivat vaikuttaa tutkimusten mukaan monenlaiset tekijät, kuten esimerkiksi oppijan aikaisempi opiskeluhistoria, uratavoitteet, ryhmän tai vanhempien taholta tulevat odotukset ja kulttuuritausta (Oscarson 1984; Strong-Krause 1997). Lisäksi voi hyvinkin olla mahdollista, että esimerkiksi oppijoiden opiskelemalla aineella voisi olla vaikutusta itsearviointin tuloksen tarkkuuteen.

Atjonen (2007, 83) toteaa, että itsearvioinnissa tärkeää on arvioinnin osuvuus eli validiteetti eikä niinkään pysyvyys (reliabiliteetti). Blanche & Merino (1989) toteavat itsearviointitutkimuksia koskevan selvityksensä pohjalta oppijoiden itsearviointin luotettavuuden vaihtelevan kielitaidon ja itsearviointiin käytettyjen materiaalien mukaan. Yleisesti he kuitenkin esittävät, että itsearviointin ja ulkoisen arvioinnin välinen luotettavuus on joko hyvä tai erittäin hyvä.

1990-luvulla kehitetyn DIALANG-kielitaidon diagnostisen arviointijärjestelmän ansiosta itsearviointia koskevan tutkimuksen kohteet ovat monipuolistuneet Oscarsonin mainitsemista. DIALANG-järjestelmän tietokonepohjaiseen ympäristöön sijoittuu muun muassa Yangin (2003) tutkimus, joka käsittelee oppijoiden järjestelmän palautteen käyttöä. Tutkimus osoittaa, että oppijat eivät välttämättä ymmärrä tietokonepohjaisen DIALANG-arviointijärjestelmän sanastotestin, itsearviointin ja osittain näiden pohjalta saadun palautteen ideologiaa. Oppijat näyttävät vertaavan DIALANG-järjestelmää tuntemiinsa virallisiin kielitutkintojärjestelmiin, joiden pääasiallinen tehtävä on osoittaa kielitaito virallisen tutkintotodistuksen avulla. Toisaalta tutkimuksesta selviää, että DIALANG-järjestelmän antaman ohjaavan palautteen ansiosta oppijat alkavat pohdita omaa oppimisprosessiaan ja niitä tekijöitä, jotka liittyvät heidän itsearviointiinsa. DIALANG-järjestelmän osalta on tutkittu ja tehty kehittämistyötä monesta muustakin aiheesta, joista voidaan mainita ainakin seuraavat: innovatiivisten tehtävätyyppien kehittäminen puhumisen ja kirjoittamisen osataitoihin, käytettyjen itsearviointikuvainten ja muiden kielitaidon mittaustapojen välinen yhteys sekä kieliasiantuntijoiden näkemykset itsearviointin hyödyistä.

Luvussa 3.4.1 tarkastellaan tutkimusten avulla itsearviointin luotettavuuteen liittyviä tekijöitä ja piirteitä. Tutkimukset on ryhmitelty sen mukaan, mihin itsearviointin piirteeseen ja piirteen luotettavuuteen ne keskittyvät. Puhumiseen tai suulliseen tuottamiseen liittyviä tutkimuksia käsitellään ensimmäisten esiteltävien tutkimusten joukossa kunkin tarkasteltavan piirteen otsikon alla. Luvun alaluvuissa on mukana myös tutkimuksia, jotka eivät koske puhumisen itsearviointitutkimuksia. Ne ovat mukana kuitenkin siitä syystä, että niistä selviää joitain yleisesti hyödyllisiä itsearviointin luotettavuuteen liittyviä seikkoja.

### 3.4.1 Itsearviointimenetelmät

Blanchen (1986) mukaan oppijat onnistuvat paremmin arvioimaan yleistä, missä tahansa yhteydessä opittua suullista kielitaitoaan kuin tiettyjen käytyjen kielikurssien sisältöjen hallintaan liittyvää taitoa tai osaamista (Blanche 1986).

Bachmanin ja Palmerin (1989) tutkimus tuo esille itsearviointimateriaalien suunnitteluun ja laatimiseen liittyvän kiinnostavan seikan osoittamalla, että itsearviointitilanteissa on merkitystä sillä, missä muodossa itsearviointiin tarkoitettu kysymys tai väittämä esitetään. Jos oppijoilta kysytään, mitä oppijat eivät osaa kielellä tehdä, se antaa korkeampia korrelaatioita suhteessa ulkoisiin tai muihin arvioihin, kuin jos heiltä kysytään, mitä he osaavat kielellä tehdä. Tällä huomiolla on tärkeä merkitys sen vuoksi, että jälkimmäinen tapa eli se, jossa kysytään, mitä oppija osaa kielellä tehdä, on yleisesti käytössä itsearviointilomakkeissa olevissa kuvauksissa ja kuvaimissa tai väittämässä (esimerkiksi *Can do* -väittämät; ALTE 2002). Tästä havainnosta huolimatta suurin osa itsearviointiin liittyvistä kuvaimista tai väittämistä on edelleen laadittu muotoon, joka korostaa sitä, mitä oppija osaa kielellä tehdä. Selitys sille, miksi käytäntöä ei ole muutettu, on luultavasti se, että itsearvioinnin oppijan kehittymistä ja kannustamista korostavaan ideologiaan eivät sovi kielteisesti muotoillut itsearviointikuvaimet tai -väittämät.

LeBlanc ja Painchaud (1985) laativat kaksi erilaista itsearviointilomaketta Ottawan yliopiston pääsykokeita varten. He toteavat, että itsearviointilomakkeen kuvainten tai väittämien muodolla on olennaisesti vaikutusta siihen, miten oppijat onnistuvat itsearvioinneissaan. Oppijat vastasivat yleisiin itsearviointikuvaimiin (esimerkiksi ”Ymmärrän lyhyitä ja yksinkertaisia tekstejä, kuten postereita, aikatauluja ja kuulutuksia”), minkä jälkeen heidän vastauksiaan verrattiin standarditestin tuloksiin. Korrelaatio tässä tapauksessa oli .53. Kun oppijat vastasivat itsearviointikuvaimiin, joiden voitiin olettaa edustavan heille tuttuja kielenkäyttötilanteita (esim. keskustelu ranskankielisen opiskelijan kanssa), korrelaatio nousi standarditestin tuloksiin verrattaessa jopa .80:een ja sen ylikin. Tutkijat toteavatkin, että oppijat näyttävät itsearvioivansa tehokkaammin toisen kielen osaamistaan silloin, kun heitä pyydetään itsearvioimaan taitoaan sellaisissa tilanteissa, jotka ovat lähellä heidän arkikokemuksiaan.

### 3.4.2 Tietokoneen avulla tehtävä itsearviointi

Tietokoneella tehtävä itsearviointi voidaan lukea itsearviointimenetelmien joukossa sen verran erityylyiseksi toiminnaksi kuin paperilla toteutettu itsearviointi, että näitä itsearviointitutkimuksia käsitellään niiden ominaislaadun vuoksi omana lukunaan.

Malabongan, Kenyonin ja Carpenterin tutkimus (2005) edustaa tietokoneen avulla suoritettua itsearviointitutkimusta (Malabonga ym. 2005). Tutkimuksessa käytettiin hyväksi tietotekniikkaa, kun verrattiin toisiinsa COPIa (*Computerized Oral Proficiency Instrument*) ja SOPIa (*Simulated Oral Proficiency Interview*). Tietokonepohjaisessa COPI-testissä oppijan itsearvioinnilla on tärkeä funktio, sillä kone ehdottaa tutkinnon aluksi tehdyn itsearvioinnin pohjalta tut-

kinnon suorittajalle taitotasoa, jolta suorittajan kannattaisi aloittaa tutkinnon suorittaminen (Kenyon & Malabonga 2001). Kenyonin (1996) varhaisemmassa COPI:n käyttöön liittyvässä itsearviointitutkimuksessa oli jo tätä ennen käynyt ilmi, että oppijoiden itsearvioinnin ja heidän SOPI:ssa arvioitujen taitotasojensa väliset korrelaatiot jäivät kohtuullisiksi (.78). Yliopistoympäristöön sijoittuvan monikielisen (arabia, kiina ja espanja) tutkimuksen tarkoitus oli tarjota oppijoille mahdollisuus kontrolloida omaa testin suorittamistaan ja tehdä suorituskokemuksesta henkilökohtaisempi koneiden avulla toteutetussa testisuorituksessa, jota oppijat usein pitävät 'kylmänä' ja 'ei-inhimillisenä' (Kenyon & Malabonga 2001). Tutkimuksessa tehtiin itsearviointia sekä paperilla (ennen tutkimusta täytettävä itsearviointilomake) että tietokoneen avulla. Tutkimuksessa jokainen 55 osallistujasta suoritti vähintään 4 tehtävää itse valitsemaltaan tasolta ja vähintään 3 tehtävää vierekkäisiltä tasoilta. Tutkimuksen molemmat itsearviointiosat osoittivat erittäin hyvää sisäistä konsistenssia: Cronbachin alfa oli paperilla tehdyssä itsearvioinnissa .98 ja tietokoneella tehdyssä itsearvioinnissa .99. Tutkimuksessa paperilla toteutetun itsearvioinnin (1. itsearviointi) ja COPI-tuotoksen välinen korrelaatio oli .79, kun taas COPI-ohjelman yhteydessä toteutetun itsearvioinnin (2. itsearviointi) välinen korrelaatio oli jopa .88. Tutkijat totesivatkin, että erityyppisten ja eri aikaan käytettyjen itsearviointi-instrumenttien tulokset tuottavat oppijoiden itsearvioinneista samanlaisia tuloksia. Tutkimuksen tuloksiin kuului myös sellainen havainto, että korkeimmille taitotasolle kielitaitonsa puolesta sijoittuvat oppijat käyttävät yleensä vähemmän aikaa vastaustensa suunnittelemiseen, mutta heidän vastauksensa ovat pidempiä kuin alemmille taitotasolle kielitaitonsa puolesta sijoittuvien oppijoiden.

Kenyonin ja Malabongan (2001) tutkimuksessa selvitettiin 4-portaisen Likert-asteikon avulla myös oppijoiden suhtautumista SOPI:n ja COPI:n. Tässä tutkimuksessa paljastui muun muassa se, että SOPIa pidettiin vaikeampana kuin COPIa ja että osallistujat tunsivat itsensä hermostuneemmiksi suorittaessaan SOPIa kuin COPIa. Osallistujat pitivät COPIa parempana esimerkiksi kielitaidon heikkouksia ja vahvuuksia esille tuotaessa sekä myös tehtävien suorittamiseen liittyvien ohjeiden selkeydessä (Kenyon & Malone 2010, 11–12).

DIALANG-hankkeeseen kytkeytyvää kirjoittamista käsittelevää tutkimusta ovat tehneet Luoma ja Tarnanen (2003) kehittäessään innovatiivisia tietokonepohjaisia tehtävätyyppejä projektin käyttöön. Projektissa jouduttiin kuitenkin rajaamaan tehtävätyyppien määrää, ja tässä vaiheessa monet innovatiiviset tehtävätyypit jäivät varsinaisesta järjestelmästä pois. Tutkimuksessa ulkomaalaiset suomen kieltä toisena kielenä opiskelevat opiskelijat kirjoittivat kaksi kirjoitus-tehtävää ja vertasivat sitten omaa suoritustaan eritasoisin maamerkkisuoritukseen (*benchmark*). Myös ulkoiset arvioijat arvioivat tutkimukseen osallistuneiden kirjoituksia. Tämän lisäksi tutkimukseen osallistujia haastateltiin ja heidän tietokoneensa näyttöruudun tapahtumia kuvattiin. Tutkimus osoitti, että oppijat pystyvät arvioimaan oman kielitaitonsa suhteellisen tarkasti, kun apuna käytettiin tarkoitukseen laadittua itsearviointitehtävää.

### 3.4.3 Kokemus itsearvioinnista

Itsearvioinnin luotettavuuteen liittyvänä vakavimpana ongelmana pidetään usein itsearviointikokemuksen puutetta, minkä vuoksi itsearviointi saattaa olla epäjohdonmukaista tai sattumanvaraista (Oscarson 1997; Ross 1998). Viitekehys muistuttaa monen muunkin itsearviointiin liittyvän tutkimuksen tavoin, että itsearviointi on luotettavaa (ts. se vastaa ulkoista arviota kielitaidosta), jos oppijat ovat saaneet koulutusta siihen (EVK 2004, 261; Oscarson 1989, 3; Tudor 1996, 167; Janssen-van Dieten 1989, McDonald & Boud 2003). Tutkijat korostavatkin itsearvioinnin käyttöönotossa niin kutsutun käynnistysvaiheen tärkeyttä, jotta oppijoiden asenteet itsearviointia kohtaan saataisiin muokattua positiivisiksi. Varsinaisen itsearviointikoulutuksen tärkeyden lisäksi oppijat tarvitsevat ohjausta perusopiskelutaidoissa, kuten ajan hallinnassa ja itsereflektiossa, jotta oppijat voisivat saada mahdollisimman suuren hyödyn omien taitojensa itsearvioinnista (Alderson & MacIntyre 2006, 304). Oppijan työskentelytaitojen paraneminen onkin mainittu yhtenä tehokkaan itsearvioinnin tuloksena itsearvioinnin alkuajoista lähtien (esim. Holec 1988). Tutkimukset ovat osoittaneet myös, että on väärin olettaa aikuisten oppijoiden hallitsevan opiskelutaidot ja oppimisstrategiat ikään kuin automaattisesti vain ikänsä ansiosta (Galloway & Labarca 1990). Joissakin tutkimuksissa korostetaan myös, että itsearviointiin kouluttamisen pitäisi sisältää muun muassa arvioinnissa käytettävien asteikkojen tulkintaa, sillä opettajilla ja opiskelijoilla voi olla koulutuksesta huolimatta hyvinkin erilainen käsitys tietyn arviointimittarin tulkinnasta (Orsmond ym. 1997, 2000).

De Saint-Léger (2009) tutki, kuinka oppijoiden suullisen kielitaidon itsearviointitaidot paranivat lukuvuoden edetessä. Tutkimuksen tulos osoitti, että kokemus auttaa myös itsearvioinnissa, sillä oppijat oppivat arvioimaan omaa suullista kielitaitoaan tarkasti tarkkailujakson aikana.

Oppijoiden halukkuus ja itsevarmuus ilmaista itseään suullisesti luokassa kasvoi itsearviointikokemuksen lisääntyessä (De Saint-Léger & Storch 2009). Molemmissa tutkimuksissa itsearviointivälineinä käytettiin itsearviointilomakkeita, jotka sisälsivät monivalintatehtäviä, itsearviointiasteikkoja ja avokysymyksiä.

DIALANG-hankkeen yhteydessä tehtiin jonkin verran tutkimustyötä, jossa tarkasteltiin itsearviointikuvaimien laatua ja suhdetta muihin kielitaidon mittareihin. Tässä yhteydessä tehtyjen pienimuotoisten tutkimusten (Desroches ym. 2005; Demeuse ym. 2006) päätuloksena voidaan pitää havaintoa, jonka mukaan itsearviointi on ilmeisen vaikeaa niille oppijoille, jotka eivät ole saaneet siihen koulutusta (Huhta 2010, 93).

Janssen-van Dieten (1992) suorittamassa tutkimuksessa, joka koski maa-hanmuuttajien suorituksia hollanti toisena kielenä -testissä ja vastaavassa itsearviointitestissä, ei vastaavuutta näiden kahden testin välillä löytynyt. Lisäksi ei löytynyt minkäänlaista yhteyttä arviointien ja oppijoita koskevien taustamuuttajien välillä (ikä, maassa oleskeluaika, alkuperämaa, hollannin kielen opintojen määrä). Tutkija toteaaakin tutkimuksensa tulosten selittyvän ainakin osittain sillä, että itsearviointia tehneet henkilöt jätettiin yksin suoriutumaan

tehtävästä eikä heitä koulutettu itsearviointiin. Siten myös Janssen-van Dieten korostaa itsearviointikoulutuksen välttämättömyyttä.

Carter (1992) tutki itsearviointia kirjoittamisen alalla ja totesi tutkimuksensa pohjalta, että itsearviointikyky paranee harjoittamalla itsearviointia. Hänen tutkimuksensa osoitti, että oppijat pystyvät arvioimaan omaa tekstiään lukuisien piirteiden osalta.

Blanchen (1990) tutkimus ei suoraan liittynyt oppijan edeltävään itsearviointikokemukseen. Hän tutki, löytyisikö todisteita siitä, että oppijan vieraiden kielten käyttötottumus vaikuttaisi siihen, miten oikeansuuntaisesti hän onnistuu arvioimaan omaa kielitaitoaan. Tutkimuksessa ei kuitenkaan voitu havaita käyttötottomusten vaikutusta.

#### 3.4.4 Arvioitava taito

Tähän lukuun valittujen esiteltävien tutkimusten valintakriteerinä on ollut pääosin se, että ne ovat kohdistuneet kielen eri osataitoihin liittyvään itsearviointiin. Luvussa raportoidaan aluksi niitä tutkimuksia, joissa itsearvioinnin käytöstä on saatu myönteisiä tuloksia ja sen jälkeen mainitaan joitakin tutkimuksia, joissa itsearviointi ei ole onnistunut.

Rossin (1998) tekemä itsearviointia koskeva meta-analyysitutkimus esittelee itsearvioinnin käytön luotettavuutta kielitaidon eri osataitojen arvioinnissa. Korrelaatiot itsearvioinnin ja ulkoisen arvioinnin välillä ovat korkeimpia tekstin ymmärtämisen osataidossa (0.61), mikä näyttäisi kertovan siitä, että itsearvioinnin käyttö tekstin ymmärtämisen osataidon yhteydessä on validimpaa kuin muissa kielitaidon osataidoissa. On kuitenkin huomattava, että tekstin ymmärtämistä koskeva korkea korrelaatio on siinä mielessä odotuksenmukainen, että tutkimuksiin osallistujina oli vieras- tai toisenkielisten tekstien lukemiseen tottuneita yliopisto-opiskelijoita (Ross 1998, 6). Myös puheen ymmärtämisen osataidossa itsearvioinnin luotettavuutta kuvaava analyysi osoittaa eri tutkimusten keskimääräisten löydösten olevan korrelaatioiltaan korkeita (.65), mutta tässä osataidossa esiintyy huomattavasti enemmän variaatiota osallistujien itsearviointien välillä (Ross 1998, 6). Puhumisen osalta Rossin itsearviointia koskeva meta-analyysitutkimus osoittaa, että sen itsearvioinnin korrelaatiot ovat ymmärtämistaitoja alempia (0.55), mikä kertoo siitä, että oppijat olisivat vähemmän valmiita itsearvioimaan puhumiseen liittyviä taitojaan (Ross 1998, 8). Samassa yhteydessä tutkimuksessa todetaan, että erityisesti puhumisen arviointi olisi altis ulkoapäin tulevien tekijöiden vaikutuksille. Kirjoittamisen itsearviointi osoittautuu Rossin tutkimuksessa sellaiseksi osataidoksi, jonka keskimääräiset korrelaatiot ovat alhaisimpia kaikista neljästä tutkitusta osataidosta (0.52; Ross 1998, 9). Rossin tutkimukseen liittyvät korrelaatioarvojen tasojen erot toimivat pohjana sille, että tutkija toteaa oppijoiden olevan hyvin eritasoisia toisen kielen eri osataitojen itsearviointitaidoissaan (Ross 1998, 5).

Aiemmin tuottamistaitojen eli puhumisen ja kirjoittamisen itsearvioinnin osalta on ilmennyt, että oppijat onnistuivat tarkemmin itsearvioimaan puhumisen kuin kirjoittamisen osataitoa (Rasch 1979).



Käsillä olevaan työhön läheisesti liittyvä Dandenaault'n (1997) Concordian yliopistossa Kanadassa toteutettu tutkimus käsittelee suullisen kielitaidon itsearviointia ranskankielisten henkilöiden kertomia tarinoita hyväksikäyttäen. Kertomuksen jälkeen jokaisesta tuotoksesta otettiin erilleen 30 sekunnin jakso, ja äidinkielenään englantia puhuvat asettivat jokaisen jakson 7-portaiselle taitotasolle. Näiden arvioiden pohjalta puhejaksot asetettiin tilastollisesti tasoille 2, 4 ja 6, minkä jälkeen jokaista tasoa edustamaan valittiin kaksi puhenäytettä jokaiselta kolmelta tasolta sekä ylimmältä ja alimmalta taitotasolta yksi puhenäyte. Tämän jälkeen ranskankieliset oppijat, joiden toisena kielenä oli englanti, kuuntelivat nauhalta englanninkielisiä puhenäytteitä ja vertasivat omaa suullista taitoaan nauhalta kuunneltaviin näytteisiin valiten joko "parempi kuin", "suunnilleen samantasoinen kuin" tai "huonompi kuin" -vaihtoehtojen väliltä. Itsearvioinnin jälkeen suoritettu Raschin analyysi osoitti, että oppijat onnistuivat hyvin sijoittamaan oman kielitaitonsa oikealle tasolle. Dandenaault'n tutkimus osoittaa, että tämänkaltainen itsearviointi on uskottavaa ja johdonmukaista. Tutkimuksensa rohkaisevien tulosten jälkeen Dandenaault suosittelee, että tämänkaltaista tutkimusta pitäisi tehdä lisää. Hän arvelee myös, että tulevissa tutkimuksissa tulisi joko lisätä kuunneltavien näytteiden määrää tai oppijoiden tehdä tehtävät ja sen jälkeen kuunnella omaa tuotostaan ja verrata niitä nauhalta kuultaviin suorituksiin. Käsillä oleva tutkimus vastaakin Dandenaault'n jälkimmäiseen jatkotutkimushaasteeseen, sillä tutkimuksen tutkimusasetelma on suunniteltu juuri tämäntyyppisen itsearvioinnin tekemiseen.

Lee-Ellis (2009) kehitti ja validoi Rasch-analyysiä käyttäen 30 minuutin pituisen C-testin korean kieleen. Tutkimuksen tarkoituksena oli luoda korean kielen taitoa mittaava jo olemassa olevia testejä (kuten esim. DLPT ja OPI) halvempi, vähemmän aikaa ja työtä vaativa korean kielen testi. Tutkimukseen sisältyi validointielementtinä 15 Can do -väittämää sisältävä puhumista koskeva itsearviointilomake. Tutkimukseen osallistujia pyydettiin arvioimaan omaa suullista korean kielen taitoaan 5-portaisella asteikolla. Tutkimuksessa selvitettiin kehitetyn testin reliabiliteetin lisäksi sitä, kuinka hyvin tutkimukseen osallistuneiden 37 henkilön C-testin tulokset korreloivat käytetyn itsearviointilomakkeen kanssa. C-testin ja itsearvioinnin tuloksien välillä vallitsi selvä positiivinen lineaarinen suhde, mikä on merkinä C-testin ja itsearvioinnin välisestä korkeasta korrelaatiosta (0.825). Lee-Ellis huomauttaa, että korrelaatio on korkea, vaikka tutkimuksessa tehty itsearviointi kohdistui suulliseen kielitaitoon, kun taas C-testi oli kirjallinen testi.

Blanchen (1986) kommunikatiivista ranskan kielen kurssia koskeva tutkimus puolestaan osoittaa, että oppijoiden ja opettajan yleistä kielen tarkkuutta koskevat arviot ovat hämmästyttävän samankaltaisia. Tähän tutkimukseen pohjautuen itsearvioinnin ja testitulosten on todettu useimmiten korreloivan keskenään tilastollisesti merkittävästi, mutta suuruus vaihtelee. Blanchen mukaan (1988, 81) eri osataitoja mittaavat kielitestit korreloivat yleensä yhtä hyvin keskenään kuin samaa taitoa mittaava itsearviointi ja testi.

Clarkin (1981) tutkimuksessa testattiin yliopisto-opiskelijoiden kiinnostusta ja taitoa vieraisissa kielissä. Tutkimuksessa verrattiin puhumisen, puheen

ymmärtämisen ja tekstin ymmärtämisen osakokeita FSI:n<sup>8</sup> haastattelujen ja vastaavien ymmärtämistaitojen osakokeiden tuloksiin. Korrelaatiot olivat kauttaaltaan melko korkeita (noin .60).

Edelleen Wilsonin tutkimuksessa (1996) pyrittiin empiirisesti osoittamaan itsearvioinnin validiteetti englannin-, saksan- ja ranskankielisten FSI/ILR<sup>9</sup> -järjestelmään perustuvien suullisen kielitaidon asteikkojen ja objektiivisten virallisten testien välillä. Tutkimuksen tuloksena oli, että Sveitsissä koulutetut sekä natiivit että ei-natiivit ranskan tai saksan kielen puhujat pystyivät arvioimaan kielitaitonsa suunnilleen samalle tasolle, kuin koulutetut kielitaidon arvioijat tekivät käyttäen apuna FSI-tyyppistä haastattelutestiä.

Mainittu LeBlanc'n ja Painchaud'n (1985) kanadalainen yli 3000:ta opiskelijaa koskeva tutkimus keskittyi tekstin ja puheen ymmärtämisen taitojen itsearvioinnin ja standardoitujen testien tulosten välisten korrelaatioiden tarkasteluun. Korrelaatiot osoittautuivat tutkimuksessa näissä osataidoissa niin korkeiksi (.80), että Ottawan yliopisto otti käyttöön systemaattisen itsearvioinnin toisen kielen ohjelmien pääsykokeissa. LeBlanc'n ja Painchaud'n tutkimuksesta saatiin tutkijoiden odotuksien vastaisesti muutakin hyödyllistä tietoa kuin vain tuloksia itsearvioinnin ja stardoitujen testien välisistä korkeista korrelaatioista. Kun itsearviointi otettiin käyttöön, opiskelijoiden siirtymiset ryhmästä toiseen vähenivät huomattavasti. Lisäksi itsearvioinnin käyttöönotosta seurasi monia käytännön etuja, kuten esimerkiksi ajansäästö omaa kielitaitoa arvioitaessa itsearviointilomakkeen (ei standardoidun lomakkeen) avulla. Itsearviointia käytettäessä ei myöskään enää tarvittu standardoituun testiin liittyvää, aikaa vievää hallintokoneistoa eikä testeissä mahdollisten vilppitapausten varalta paikalla olevia valvojia.

Itsearvioinnin toimivuudesta on saatu tietoa myös sellaisissa tutkimuksissa, joissa kielitaidon itsearviointi on ollut vain yksi tarkasteltava asia muiden joukossa. Tällaisesta tutkimuksesta esimerkkinä voidaan mainita Tudorin ja Nivelles'n (1991) tekemä kaksiosainen tutkimus, jossa tarkasteltiin varsinaisen itsearvioinnin (osa 1) lisäksi oppijoiden oppimisprioriteetteja sekä oppijoiden käyttämiä opiskelutekniikoita (osa 2). Itsearviointiosassa oppijoita pyydettiin ensin arvioimaan englannin kielen yleistasonsa ja oma suullinen kielitaitonsa neljän kieleen liittyvän piirteen osalta (ääntäminen, sujuvuus, laajuus ja ilmaisen tarkkuus). Tähän tarkoitukseen käytettiin sanallisesti määriteltä skaalaa, johon oppijat sijoittivat arvionsa välille 1–20. Oppijoiden itsearviointia verrattiin sitten opettajan suorittamaan arviointiin ja oppijoiden kurssin loppukokeessa tekemään suoritukseen. Kurssin loppukoe koostui kahden opettajan suorittamasta suullisesta haastattelusta (aiemmin mainittu luokan opettaja ei ollut haastattelua tekemässä) sekä kielitaidon arviointiin tarkoitettua testipatteristosta. Tutkimuksen tulokset osoittivat suullisessa taidossa huomattavan suurta vastaavuutta oppijoiden itsearvioinnin ja sekä opettajan että loppukokeen tuottamien arviointien välillä (itsearviointi–opettajan suorittama arviointi .70; itsearviointi–loppukokeen tuottama arviointi .55). Kielen yleistaitotason korrelaati-

---

<sup>8</sup> Foreign Service Institute

<sup>9</sup> Interlanguage Roundtable

tiot olivat hieman alempia, vaikkakin vielä melko hyviä (itsearviointi-opettajan suorittama arviointi .67; itsearviointi-loppukokeen tuottama arviointi .68). Tutkimuksen johtopäätös itsearvioinnista olikin, samoin kuin yleisesti itsearviointia koskevien tutkimusten mukaan se, että oppijoilla näyttäisi olevan tietynlaista tietoa ja ymmärrystä osaamisestaan, minkä pohjalta he pystyvät arvioimaan omaa kielitaitoaan. Tämä yleisemminkin itsearviointitutkimuksia koskeva tulos (ks. Blanche & Merino 1989) onkin perusteena sille, että itsearviointia suositellaan käytettäväksi opettajan suorittaman arvioinnin ja standardoidun testiarvioinnin rinnalla (Tudor 1996, 168–169.)

Huhta (2010, 93–94) tutki DIALANG-hankkeessa käytettyjen itsearviointiväittämien ja muiden kielitaidon mittaustapojen liittymistä toisiinsa englannin kielen pilottidatan avulla ja suhteutti tulokset oppijoiden kielitaidon tasoon, äidinkieleen ja koulutustasoon. Tässä tutkimuksessa kävi ilmi, että oppijat yliarvioivat lukemisen ja kuullun ymmärtämisen taitojaan, mutta osoittivat pientä aliarviointitaitumusta kirjoittamisen taitojensa arvioinnissa. Lukemisessa taitojen yliarviointi korostui A1- ja B1-tasoisilla kielenoppijoilla, joista suuri osa yliarvioi osaamistaan sijoittaen sen jopa kaksi tai useampaakin tasoa paremmaksi kuin mitä todellinen osaaminen oli. Kuullun ymmärtämisessä alimpien tasojen (A1–A2) oppijat yliarvioivat myös suuresti taitojaan, kun taas ylimpien tasojen (C1–C2) oppijat näyttivät aliarvioivan englannin kielen kuullun ymmärtämisen tasoaan. Nämä esimerkit yliarvioinnista ovat samansuuntaisia kuin tietyissä aiemmissa tutkimuksissa. On ollut tutkimuksia, joissa nk. hyvät oppijat ovat aliarvioineet suoritustensa tasoa ja päinvastoin heikot oppijat taas puolestaan yliarvioineet suorituksiaan (esim. Dochy ym. 1999). Merkille pantavaa tässä tutkimuksessa verrattuna muihin mainittuihin on se, että taitotasosteikon keskivaiheille (B1–B2) sijoittuvat oppijat onnistuivat parhaiten itsearvioinneissaan. Kirjoittamisessa toistui sama suuntaus kuin kuullun ymmärtämisessäkin, ts. alimpien taitotasojen oppijat yliarvioivat osaamistaan ja ylimpien tasojen oppijat taas aliarvioivat osaamistaan.

Itsearviointitutkimusten joukossa on myös tutkimuksia, joissa ei löydetty kovin hyvää vastaavuutta itsearvioinnin ja ulkoisen arvioinnin välillä. Yksi tunnetuimmista on kanadalainen tutkimus, jossa löydettiin vain heikko yhteys 8. luokan ranskan kielen kielikylpyopiskelijoiden itsearviointitulosten ja heidän virallisesta testistä saamiensa tulosten välillä (puheen ymmärtäminen, tekstin ymmärtäminen, puhuminen ja kirjoittaminen; Peirce, Swain & Hart 1993). Yhdeksi syyksi huonolle vastaavuudelle tutkijat arvelivat sitä, että tutkimukseen osallistuneilla oppijoilla ei olisi useinkaan ollut mahdollisuutta verrata omaa ranskan kielen taitoaan autenttisten ranskan kielen puhujien kielitaitoon. Tämän oletuksen pohjalta tutkijat arvelivat, että itsearvioinnin ja ulkoisen arvioinnin huono vastaavuus tai vastaamattomuus voi päteä missä tahansa koulutuskontekstissa olevaan oppijaan, jota pyydetään arvioimaan omaa kielitaitoaan. Tämän vuoksi tutkijat suhtautuivatkin varauksellisesti itsearvioinnin käytön hyödyllisyyteen.

Patrin (2002) suullisen esitystaidon itse- ja vertaisarviointiin liittyvä tutkimus osoitti melko matalia korrelaatioita verrattaessa mainittuja arviointeja

opettajan arviointeihin (kontrolliryhmä .50 ja testiryhmä .46). Sen sijaan testi-ryhmän vertaisarviointien ja opettajan suorittaman arvioinnin väliset korrelaatiot olivat korkeita (.85).

Myöskään Campbellin ym. (2001) tutkimus, jossa selvitettiin itse- ja vertaisarvioinnin hyödyllisyyttä opettajan arviointiin verrattuna suullisen kaupallisen esityksen arvioinnissa, ei osoittanut itsearvioinnin vastaavan opettajan arviointia (holistinen arviointi, korrelaatio .41) kovinkaan hyvin. Sen sijaan tutkimus osoitti vertaisarvioinnin vastaavan opettajan arviointia monilta osin (holistinen arviointi, korrelaatio .58). Tutkijoiden johtopäätös itsearvioinnin osalta on siten se, että vertaisarviointi on hyödyllisempää suullisen kaupallisen esityksen arvioinnissa kuin itsearviointi. Tutkimuksessa todettiin myös suullisten esitysten itsearviointiin liittyen, että esiintyjillä oli taipumusta yliarvioida visuaalisten apukeinojen tärkeys ja aliarvioida esityksen sisältö sekä esiintymiseen liittyvät tekijät kuten katsekontakti, kehonkieli ja äänen laatu.

### 3.4.5 Oppijan taitotaso

Monissa itsearviointia koskevissa tutkimuksissa on selvitetty sitä, onko itsearviointia tekevän oppijan taitotasolla merkitystä hänen itsearviointinsa paikkansapitävyyteen.

Yksi mainitsemisen arvoinen tutkimus on Falchikovin ja Boudin (1989) tutkimus, joka koskee eri aineita opiskelevien korkeakouluopiskelijoiden itsearviointia. Tutkimus osoitti, että oppijat, joiden kielitaito sijoittui taitotasoasteikon ylimmille tasoille, onnistuivat arvioimaan oman kielitaitonsa tarkemmin kuin ne oppijat, joiden kielitaito sijoittui alemmille taitotasoille. Muidenkin tutkimusten tulokset ovat osoittaneet, että pidemmälle kielen hallinnassaan edenneiden oppijoiden itsearviointi on tarkempaa ja oikeampaa suhteessa johonkin ulkopuoliseen mittariin kuin alimmilla taitotasoilla olevien oppijoiden itsearviointi (esim. Dochy ym. 1999; Caouette 1990).

Myös Lewkowiczin ja Moonin (1985) tutkimus, jossa selvitettiin eri kielitaidon tasoilla olevien oppijoiden oman kielitaitonsa arvioinnin paikkansapitävyyttä, antoi samansuuntaisia tuloksia kuin edempänä esitelty Huhdan (2010) tutkimus. Tendenssi oli, että ylemmäksi kielitaidon asteikolla sijoittuvilla oppijoilla oli taipumusta oman kielitaitonsa aliarviointiin, kun taas alemmille taitotasoille sijoittuvien kielitaidon itsearvioinnissa oli havaittavissa taipumus yliarviointiin (ks. myös Heilenman 1990; Oscarsson 1984; Blanche-Merino 1989).

Matsunon (2009, 75) japanilainen yliopistotasolla toteutettu tutkimus, jossa verrattiin englannin vieraan kielen kirjoittamisen itse- ja vertaisarviointia opettajan suorittamaan arviointiin, tuotti samanlaisia tuloksia kuin edellä. Tutkimuksen eräs tulos oli se, että erityisen monet kirjoittamistaidoiltaan hyvätaoiset oppijat arvioivat oman kirjoittamisen taitonsa alemmaksi kuin oli odotettavissa. Matsuno huomatti kuitenkin, että itsearvioivien opiskelijoiden opettaja seurasi, miten opiskelijat arvioivat oman kirjoittamisen tuotoksensa. Tutkija totesi, että tällä seikalla saattoi olla vaikutusta siihen, että oppijat pikemminkin aliarvioivat kuin yliarvioivat omaa suoritustaan (Matsuno 2009, 88–89).

Jaakkolan ym. (2002, 49) tutkimuksessa, jossa keskityttiin selvittämään koululaisten oppimaan oppimista ja sen edistämistä, todettiin, että opinnoissaan hyvin suoriutuvat käyttävät hyväkseen itsearviointissa tarvittavia meta-kognitiivisia strategioita useammin ja tehokkaammin kuin opinnoissaan heikommin menestyvät oppijat.

DIALANG-hankkeeseen liittyvästä Desroches'n ym. (2005) tutkimuksesta selvisi puolestaan se, että itsearviointi on vaikeaa niille oppijoille, joiden kieli-taito sijoittuu taitotasoasteikon alimmille tasoille.

Huhdan DIALANG-hankkeeseen liittyvässä toisessa tutkimuksessa selvitettiin muun muassa sitä, kuinka DIALANG-järjestelmää käyttäneet opiskelijat suhtautuivat järjestelmästä saamaansa itsearviointipalautteeseen ja itsearviointiin käyttöön yleensäkin. Tutkimus osoitti, että oppijan taitotasolla ei näytännyt olevan yhteyttä siihen, kuinka he kokivat saamansa itsearviointipalautteen suhteessa saamaansa testitulospalautteeseen (Huhta 2007).

### 3.4.6 Oppijan ikä ja sukupuoli

Kuten edellä selostettiin, DIALANG-hankkeessa on tehty tutkimusta muun muassa siitä, kuinka järjestelmää käyttäneet opiskelijat suhtautuivat järjestelmästä saamaansa itsearviointipalautteeseen ja itsearviointiin käyttöön yleensäkin. Huhta (2007) tutki, kuinka DIALANG-käyttäjien tausta (ikä, sukupuoli, taitotaso) vaikutti siihen, missä määrin käyttäjät pitivät järjestelmästä saamaansa itsearviointia koskevaa palautetta kiinnostavana ja hyödyllisenä. Tutkimustulosten valossa näytti siltä, että nuoret alle 18-vuotiaat suhtautuivat saamaansa itsearviointipalautteeseen negatiivisemmin kuin vanhemmat käyttäjät (Huhta 2007). Minkäänlaisen tiukan johtopäätöksen tekeminen siitä, miten eri-ikäiset oppijat suhtautuvat itsearviointiin tai siitä saatavaan palautteeseen, on haastavaa. Mitä luultavammin oppijan iällä on tietty merkitys erityisesti siksi, että vanhemmilla oppijoilla on kokemusta monenlaisista tilanteista, kun taas nuorten kokemusmaailma on vielä rajallisempi. Jaakkolan ym. (2002, 46) edellä mainitusta tutkimuksesta selvisi, että 7. luokkalaisilla oli vaikeuksia arvioida sekä oppimistavoitteitaan että arvosanojaan realistisesti. Yleensä oppilaat yliarvioivat sekä oppimistavoitteitaan että arvosanojaan. Toisaalta tutkimukseen osallistuneiden opettajien mukaan metakognitiiviset taidot kehittyvät iän ja kypsyyden karttuessa. Itsearviointin harjoittamisella oli suora vaikutus 9. luokan oppilaiden metakognitiivisten reflektointitaitojen kehittymiseen.

Sukupuolen vaikutusta itsearviointin onnistumiseen on tutkittu joissakin tutkimuksissa vaihtelevin tuloksin. Joidenkin tutkijoiden mukaan juuri oppijan sukupuolella on suuri merkitys muun muassa sosiaalistumisprosessiin, rooliodotuksiin ja kielenoppimisstrategioihin (esim. Oxford 1990), joilla kaikilla saattaa olla yhteytensä siihen, kuinka onnistunutta oppijan mahdollinen itsearviointi on.

Caouetten (1990) tutkimus osoitti tyttöjen selviytyvän poikia paremmin itsearviointitehtävistä, kun taas Smithin tutkimus (1997) ei osoittanut eroja tyttöjen ja poikien itsearviointien välillä. Myöskään Coomben tutkimuksessa (1992) sukupuolten välillä ei tuntunut olevan eroa.

Thomson (1996) teki itseohjautuvuuteen keskittyvän itsearviointitutkimuksen, jossa mukana olleet oppijat olivat äidinkieleltään englannin-, kiinan- ja koreankielisiä. Tutkimuksen tuloksia oli muun muassa se, että korealaiset miesopiskelijat arvioivat oman kielitaitonsa parhaimmaksi kaikista mukana olleista ryhmistä. Toisena ääripäänä tutkimuksessa olivat kiinalaiset naisopiskelijat, jotka sitä vastoin arvioivat oman osaamisensa paljon huonommaksi kuin mitä heidän saamansa keskitasoiset kurssiarviot olivat (Thomson 1996, 83).

Mainittu Huhdan tutkimus (2007) siitä, kuinka DIALANG-käyttäjien tausta vaikuttaa siihen, missä määrin he pitävät saamaansa itsearviointipalautetta kiinnostavana ja hyödyllisenä, osoitti, että naiset pitävät saamaansa itsearviointipalautetta kiinnostavampana kuin miehet.

### 3.4.7 Oppimiskulttuuri ja koulutustausta

Tutkimuksissa on tuotu esille myös oppijan kulttuuri-, koulutus- ja kehitystasojen vaikutus hänen oppimista koskeviin uskomuksiinsa ja asenteisiinsa (esim. Thomson 1996, 77).

Holec (1990, 43) toteaa, että oppijan aikaisempien oppimiskokemusten kuluessa omaksuma oppimiskulttuuri vaikuttaa paljon siihen, miten hän toteuttaa arviointia myöhemmin.

Coomben (1992) maahanmuuttajien toisen kielen (englanti) oppimista koskevissa tutkimuksissa kehitettiin Likert-asteikkoinen itsearviointiväline, jonka tuloksia verrattiin virallisen kielitestin (BEST)<sup>10</sup> tuloksiin. Tämän tutkimuksen mielenkiintoinen tulos on se, että luku- ja kirjoitustaidon sekä itsearvioinnin välillä vallitsee korkea korrelaatio ( $r=.83$ ).

Latomaan (1996) maahanmuuttajia Pohjoismaissa koskeva itsearviointia koskeva tutkimus kuuluu sosiolingvistisen tutkimuksen alaan. Tutkimuksessa selvitettiin taustatekijöiden ja oppijoiden saaman itsearviointitason muuttujia. Tutkimuksessa todettiin, että toisen kielen itsearvioinnin kannalta tärkeimpiä muuttujia ovat maahanmuuttajan koulutustausta ja maahanmuuttoikä. Yleisellä tasolla Latomaan tutkimuksen tulokset tukevat käsitystä itsearvioinnin ja ulkoisen testauksen vastaavuudesta.

ToLP<sup>11</sup>-projektin kysely ja DIALANG liittyvät yhteen siten, että ToLP-projektin kyselyssä selvitettiin yläkoululaisten asenteita itsearviointia kohtaan esittämällä heille samantyyppisiä kysymyksiä, joita esiintyy DIALANG-järjestelmän itsearviointiosuudessa. Tässä yhteydessä varmistui yleinen käsitys, jonka mukaan sillä kokemuksella ja palautteella, jota oppijat saavat virallisen kielenopetusympäristön, kuten koulun välityksellä, on suuri vaikutus siihen, miten huonona tai hyvänä he pitävät omaa kielen osaamistaan. On kuitenkin myös mainittava, että suomalaiset yläkoulun oppilaat pitävät tärkeimpänä yksittäisenä merkinä omasta osaamisestaan sitä, kuinka hyvin he selviävät todellisista kielenkäyttötilanteista esim. matkustaessaan (Huhta 2010, 119).

<sup>10</sup> Basic English Skills Test

<sup>11</sup> ToLP = Towards Future Literacy Pedagogies

### 3.4.8 Kulttuuritausta

Kulttuuritaustalla saattaa olla vaikutusta erityisesti itsearvioinnissa havaittavaan oman osaamisen yli- tai aliarviointiin. Tämä taipumus tulee esille monissa itsearviointitutkimuksissa (ks. alla), joihin osallistuneet opiskelijat edustavat eri kulttuuritaustoja ja joiden välistä itsearvioinnin luotettavuutta verrataan toisiinsa tutkimuksessa. Tässä pelkästään suomalaisia oppijoita sisältävässä tutkimuksessa kulttuuritaustaan pohjautuva vertailu ei ole mahdollista, sillä tämän tutkimuksen pohjalta ei luonnollisestikaan voida tehdä päätelmiä eri kulttuuritaustaisten henkilöiden itsearvioinnin luotettavuudesta. Sen sijaan suomalaisen kulttuuritaustan vaikutus on havaittavissa myös tämän tutkimuksen oppijoiden tietyissä kommentteissa ja tendensseissä. Tästä syystä kulttuuritausta on haluttu mainita yhtenä itsearvioinnin luotettavuuteen liittyvänä seikkana tässä luvussa.

Kulttuuristen tekijöiden ja oppijan kulttuuritaustan vaikutus itsearvioinnin tulosten tarkkuuteen on tuotu esille monissa tutkimuksissa (esim. Lewkowicz & Moon 1985; Blue 1988; Strong-Krause 1997; Thomson 1996, Matsuno 2009). Lewkowiczin ja Moonin tutkimuksessa tuotiin kulttuurisidonnaisuuden lisäksi esille fyysiset ja sosiaaliset seikat, jotka voivat vaikuttaa oppijakeskeisen näkemyksen käyttöönottoon tietyssä kontekstissa ja sitä kautta itsearvioinnin onnistumiseen.

Blue (1988) puolestaan on todennut tutkimuksiinsa nojautuen, että eri kulttuuriryhmille ominainen oman kielitaidon joko yli- tai aliarviointi on monikulttuuristen opiskelijaryhmien opettajille suuri haaste. Kulttuuritaustalla näyttää olevan vaikutusta oppijan itsearvioinnin hyväksymiseen ja sen oikeellisuuteen.

G.D. Deckert (1990) toteutti itsearviointitutkimuksen, jonka osallistujina oli 107 eritaustaista opiskelijaa. Tutkimuksessa oli oppijoita Lähi-idästä (55 arabia), Latinalaisesta Amerikasta (22), Itä-Aasiasta (16, joista 11 japanilaisia) ja Kaakkois-Aasiasta (14). Tutkimus osoitti, että itseluottamukseen perustuvan arabikulttuurin edustajista melkein puolet yliarvioi oman kielitaitonsa jopa 10 pisteellä tai ylikin. Tendenssi oli vastakkainen japanilaisilla, ts. heistä vain yksi yliarvioi itsensä yli 10 pisteellä ja 4 aliarvioi oman osaamisensa samassa määrin.

Edellä mainitussa Thomsonin (1996) tutkimuksessa todettiin, että aasialaisilla naisopiskelijoilla on suuria vaikeuksia antaa opettajalleen ”suoraa” palautetta siitä, mitä he ajattelivat kurssiin liittyvästä itsearvioinnista. Jos ja kun he uskaltautuivat antamaan palautetta opettajalle, palaute oli joko ilmaistu ylikohdella tavalla tai sen sijoituspaikka oli erikoinen muuhun oppijan kirjoittamaan tekstiin nähden (esimerkiksi loppukirjoitus varsinaisen muun opettajan luettavaksi tarkoitetun tekstin jälkeen; Thomson 1996, 87).

DIALANG-hankeeseen liittyvässä tutkimuksessa Floropoulou (2002b) tutki viiden kreikkalaisen ja viiden kiinalaisen opiskelijan suhtautumista itsearviointiin ja DIALANG-järjestelmään. Hänen mukaansa henkilön kulttuuritausta on suhteessa heidän näkemyksiinsä: suurin osa tutkimuksen kohteena olevista henkilöistä pitää itsearviointia kannatettavana toimintana ja suhtautuu siten siihen myönteisesti. Kiinalaisten oppijoiden asenne DIALANG-järjestelmää

kohtaan on kuitenkin myönteisempi kuin kreikkalaisten, jotka osoittavat välinpitämättömiä tai jopa kielteisiä asenteita järjestelmää kohtaan.

DIALANG-hankkeen pilottivaiheen yhteydessä tehdyissä tutkimuksissa selvitettiin muun muassa sitä, kuinka DIALANG-järjestelmässä käytetyt itsearviointikuvaimet ja muut kielitaidon mittaustavat liittyvät toisiinsa. Niissä tarkasteltiin myös muiden tekijöiden suhdetta itsearviointiin. Huhta (2010) käytti englannin kieleen liittyvässä tutkimuksessaan muuttujina oppijan kielitaidon tasoa, äidinkieltä ja koulutustasoa. Tutkimuksessa ei havaittu eroa itsearvioinnin tarkkuudessa suomalaisten ja saksalaisten oppijoiden välillä. Tutkimustulokset eivät tue käsitystä, jonka mukaan eri kulttuuri- ja kielitaustaiset oppijat eroaisivat kyvyiltään tehdä itsearviointia.

### 3.4.9 Oppijan persoonallisuus ja minäkuva

Monissa tutkimuksissa on todettu, että toisen kielen taidon yliarviointi on yleisin itsearviointiin liittyvä epätarkkuus (esim. Oscarson 1984, Blanche & Merino 1989; Heilenman 1990; Sinharay ym. 2009). Itsearvioinnin luotettavuuteen liittyvistä tutkimuksista ja luotettavuuteen vaikuttavista tekijöistä voidaan mainita myös itsearviointiin ja kielelliseen ahdistukseen (*language anxiety*) liittyvät tutkimukset. Muun muassa MacIntyre, Noels ja Clément (1997) tutkivat oppijoiden ahdistuksen vaikutusta heidän itsearviointeihinsa ja totesivat, että ahdistuneilla oppijoilla on taipumusta kielitaitonsa aliarvioimiseen, kun taas vähemmän ahdistuneet oppijat usein yliarvioivat kielitaitoaan (MacIntyre ym. 1997, 266). Kieleen liittyvä ahdistus näyttää tutkijoiden mukaan aiheuttavan omien taitojen aliarvioinnin lisäksi mm. sen, että ahdistuneet oppijat kommunikoiivat vähemmän kuin rennot oppijat ja että ahdistuneiden oppijoiden tuotokset eivät ole laadullisesti niin hyviä kuin muiden oppijoiden (MacIntyre & Gardner 1994). Tutkijat muistuttavat, että erityisesti puhuminen on julkista toimintaa, jossa kommunikoidaan toisten kanssa, mikä osaltaan voi lisätä ahdistuneisuutta (MacIntyre & Gardner 1994). MacIntyre ym. (1997, 266) totesivat niin ikään, että vaikka itsearviointi usein kehittääkin oppijoiden kielellistä kykyä, se saattaa olla erittäin tuskallista ja ei-motivoivaa kovin ahdistuneille oppijoille, jotka itsearvioinnin kautta tulevat tietoisiksi puutteistaan.

On myös syytä huomioida, että eräät muutkin oppijan persoonaan liittyvät piirteet kuin ahdistus saattavat vaikuttaa itsearvioinnin luotettavuuteen. Esimerkiksi sillä, onko oppija luonteeltaan sisään- vain ulospäinsuuntautunut saattaisi olla vaikutusta hänen itsearviointia koskeviin vastauksiinsa ja siten itsearvioinnin luotettavuuteen. On todennäköistä, että monet suulliseen vuorovaikutukseen liittyvät tilanteet, joita kuvataan erityisesti vierasta kieltä koskevissa puhumisen itsearviointiväittämässä, saattavatkin olla hankalia sellaiselle oppijalle, joka ei tunne olevansa kovin puhelias eikä sujuvasanainen edes omalla äidinkielellään. Joissakin tutkimuksissa onkin todettu, että tietynlaisilla luonteenpiirteillä, kuten esimerkiksi varmuudella, on yhteyttä toisen kielen suulliseen hallintaan (Ockey 2011, 983). Tässä tutkimuksessa ei kuitenkaan voida paneutua tähän asiaan, mutta maininnan arvoiseksi tämän seikan tässä kohdassa tekee se, että tutkimukseen osallistuneiden oppijoiden itsearviointia koskevissa



kommenteissa on tietyn tyyppiseen luonteeseen liittyviä kommentteja (ks. oppijoiden kommentit luvun 5.4.1 alakohdassa ”Oppijan luonteen ja/tai toimintatavan vaikutus itsearviointiin”).

Itsearvioinnin luotettavuudesta puhuttaessa on niin ikään paikallaan muistaa, että joillakin oppijoilla on taipumusta aliarvioida systemaattisesti omia taitojaan (ks. esim. Ferguson 1978; Holec 1979). Kyseessä näyttäisi siten olevan tietynlainen persoonallisuuden piirre, joka aiheuttanee sen, että oppijalla on taipumusta joko oman osaamisensa aliarviointiin tai yliarviointiin. Oppijan persoonaan liittyvät seikat, kuten esimerkiksi hänen itseluottamuksensa vaikuttavat todennäköisesti oppijan itsearviointiin ja sen onnistumiseen kulttuuristen seikkojen lisäksi. Sinharay ym. (2009) raportoivat kolmessa Etelä-Amerikan maassa toteutetussa TOEIC Bridge -testiä koskevassa tutkimuksessaan, että oppijat, jotka uskoivat selviytyvänsä hyvin englannin kielellä, onnistuivat paremmin suorituksissaan kuin ne oppijat, jotka eivät näin ajatelleet. Tutkimuksessa todettiin myös, että itsearvioinnissa käytettyjen Can do -väittämien ja TOEIC Bridge -testissä saatujen arvosanojen välinen korrelaatio vastasi hyvin aiempien validointitutkimusten tasoa.

### 3.5 Itsearviointi ja asenteet

Oppijan omalla asenteella itsearviointia kohtaan on ratkaiseva merkitys sekä itsearvioinnin onnistumisen että sen luotettavuuden kannalta. Itsearvioinnin harjoittamisessa tarvitaan Viitekehyksessä mainittujen deklaraatiivisen tiedon (*savoir, kokemusperäinen tieto ja muodollisen opiskelun tuottama tieto*), taitojen ja taitotiedon (*know-how, savoir-faire, proseduraalinen tieto eli kyky toimia*) sekä oppimiskyvyn (*savoir-apprendre*) hallinnan lisäksi erityisesti elämäntaitojen (*eksistentiaalinen kompetenssi, savoir-être*) hallintaa. Elämäntaitotaidot tarkoittavat yksilöllisten ominaisuuksien, persoonallisten piirteiden ja asenteiden kokonaisuutta, joihin sisältyy käsitys itsestä ja muista sekä halukkuus toimia sosiaalisessa vuorovaikutuksessa muiden ihmisten kanssa. Kielenkäyttäjien kommunikatiiviseen toimintaan vaikuttavat muun muassa erilaiset asenteet, motivaatio, arvot, uskomukset, kognitiiviset tyylit ja persoonallisuustekijät. Myös kieleen ja kielenkäyttöön liittyvä herkkyys (*kieli- ja viestintätietoisuus*) sekä erilaisten opiskelutaitojen, jotka on Viitekehyksessä luokiteltu oppimiskyvyn alle, riittävä hallinta, liittyy tärkeänä osana itsearviointiin. Mikäli itsearviointia halutaan harjoittaa tehokkaasti, opiskelutaidoissa mainitut taidot, joissa korostuvat oppijan kyky itsenäiseen työskentelyyn, oman oppimisensa suunnitteluun ja arviointiin suhteessa hänen omaan persoonallisuuteensa ja voimavaroihinsa, ovat välttämättömiä (EVK 2004, 30–32, 146–154.) Kantelinen ja Hildén (2012, 3) toteavatkin, että vaikka esimerkiksi Suomessa on hiljattain otettu perusopetukseen käyttöön kansallinen versio Eurooppalaisesta kielisalkusta, salkun käyttöönotto ja hyödyntäminen vaativat ennen kaikkea asenteellisia valmiuksia.

On selvää, että kaikkien oppijoiden asenteet omaa oppimistaan kohtaan, itsearviointi mukaan lukien, eivät aina ole helppoja muuttaa. Osa oppijoista saattaa takertua perinteiseen opettajatoiseen näkemykseen, jossa vain opettajalla on ollut valta arvioida (Cram 1995). Tällaiset oppijat tuntuvat jäävän omien asenteidensa vangiksi, koska uusien näkökulmien ja uuden vastuunjaon hyväksyminen on heille vaikeaa. Eri kulttuureissa vallalla olevat kulttuuriset arvostukset voivat siten vaikuttaa suuresti siihen, miten oppijoiden autonomisuutta voidaan parhaiten edistää. Monissa tapauksissa tarvitaan radikaalia asenne-muutosta, mikäli halutaan edistää oppijakeskeisiä oppimismetodeja, kuten esimerkiksi itsearviointia ja sen avulla oppijoiden autonomisuutta. Oppijoiden asenteilla on itsearvioinnissakin tärkeä merkitys, sillä ne vaikuttavat ensinnäkin siihen, haluaako oppija lainkaan tehdä itsearviointia, ja toiseksi siihen, onnistuuko itsearviointi antamaan todellisen kuvan oppijan kielitaidosta. Toisaalta ei tule myöskään unohtaa, että on olemassa itse- ja/tai vertaisarviointia käsitteleviä tutkimuksia, joissa oppijoiden asenteet näitä arviointimuotoja kohtaan ovat olleet erittäin positiivisia (ks. esim. Lindblom-Ylänne ym. 2006).

Oppijoiden asenteiden muutoksen lisäksi myös opettajien asenteiden tulee muuttua. Kun puhutaan jonkin uuden menetelmän käyttöönottoon liittyvästä vastustuksesta, ajatellaan liian usein vastustuksen tulevan oppijoiden suunnasta. Myös opettajilla saattaa olla kielteisiä näkemyksiä itsearvioinnista ja sen luotettavuudesta. Monesti oppijoilta voidaan vaatia melkoista muutosta, jotta he kehittyisivät autonomisiksi oppijoiksi, mutta ei muisteta itse opettajan tehtävässä toimiessa pohtia omaa suhtautumista uusien ajatuksiin ja menetelmiin kohtaan. Opettajalta odotetaan oppimista koskevassa vastuunjaossa ammattitaidon kasvua ja kehittymistä, mikä asettaa vaatimuksia totutuille ja mahdollisesti oman opettajakoulutuksen aikana opituille opetusmenetelmille ja -opetustyyliille sekä omalle ammatilliselle persoonalle. Oppijakeskeisissä oppimissuuntauksissa opettaja toimiikin pikemminkin neuvonantajan, koordinoijan, ryhmän vetäjän, mentorin tai oppaan kuin perinteisen opettajan roolissa (esim. Université Laval 1996). Opettajan monien eri roolien lisäksi jotkut tutkijat ovat pohtineet sitä, pystyykö opettaja uudessa roolissaan edistämään oppijan itsearviointitaitoa. Tässä tehtävässä haastavaa on riippuvuuden ja automian välisen tasapainon löytäminen, sillä jokainen oppija on erilainen (Barbot 1990, 78). Kohonen (1997, 10) tähdentää, että itsenäisiä oppijoita kouluttavat itsenäiset opettajat ja että opettajan ammatillisen kasvun ja arviointiin liittyvän uudistamisen välinen yhteys on välttämätön niissä oppimisympäristöissä, joissa mielitään kasvattaa omasta oppimisestaan vastuuta kantavia oppijoita. Useat tutkijat painottavatkin, että opettajien koulutusta tulee uudistaa siten, että siihen sisällytetään oppilaskeskeiseen arviointiin liittyviä tekniikoita ja perehtymistä Eurooppalaisen kielisalkun kaltaiseen yhteisesti hyväksytyyn välineeseen (Oscarson 1989; Kantelinen, Hildén 2012).

LeBlanc ja Painchaud (1985) totesivat tutkimuksessaan, että itsearviointiin kuuluva oletamus siitä, että jokainen oppija ottaisi vastuun omasta oppimisestaan, voi aiheuttaa ongelmia sellaisille oppijoille, joiden mielestä kaiken oppimisen arviointi pitää jättää asian hallitsevien auktoriteettien hoidettavaksi. Jotkut oppijat ovat voineet tottua sellaiseen luokkahuonekulttuuriin, että itsearvioin-

nin kaltaisen arviointimuodon käyttöönotto saattaa tuntua heistä rooli-dotusten rikkomiselta, mistä taas voi seurata vastahakoisuus tai jopa vastustus itsearviointia kohtaan (Tudor 1996, 174). Osa oppijoista ei siis välttämättä ole kiinnostunut harjoittamaan itsearviointia ja siten jakamaan opetukseen ja sen arviointiin liittyvää vastuuta. Vastahakoisuus saattaa johtua itsearviointikokemuksen puutteen lisäksi muun muassa siitä, että oppija ei luota itseensä ja kykynsä tehdä oikeanlaisia arvioita kielitaidostaan. Erityisesti niissä maissa, joissa on vielä vallalla autoritaarinen ja opettajakeskeinen opetuskuulttuuri, itsearvioinnin käyttöönotto voi viedä aikaa. Huuskon (2008) mukaan itsearvioinnissa ongelmallista onkin juuri defensiivinen suhtautuminen silloin, kun henkilö käsittelee omia puutteitaan. Hänen mukaansa reflektiivinen itsearviointi edellyttää rohkeutta tarkastella oman toiminnan lähtökohtia, arvoja ja edellytyksiä sekä muuntautumiskykyä (Huusko 2008, 133). Joidenkin tutkimusten mukaan itsearviointiin todellakin liittyy voimakkaita tunteita. Itsearviointia saatetaan pitää hyödyllisenä ja paljastavana tai sitten turhana lomakkeiden täyttämisenä tai jopa ahdistusta aiheuttavana toimintana (Alanen & Kajander 2011, 74). Muita syitä oppijan haluttomuuteen tehdä itsearviointia voivat olla muun muassa huono lukutaito, edeltävät oppimiskokemukset ja itsearvioinnin kokeminen uhkana tai jollakin tavalla itselle kaukaisena asiana.

Kun puhutaan kielteisistä asenteista itsearviointia kohtaan, on syytä muistaa, että joidenkin tutkijoiden mukaan vastustus itsearviointia kohtaan itsearviointiprosessin alussa on normaalia. Heidän mukaansa vastustusta ei voi siten myöskään täysin välttää. Esimerkiksi Underhill (1987, 296) toteaa, että jotta itsearviointia voitaisiin käyttää tehokkaasti hyväksi, on lähdettävä tietoisesti muuttamaan oppijoiden tottumuksia, uhmattava heidän erilaisia oppimista koskevia uskomuksiaan, kyseenalaistettava normeja ja saatettava julkiseksi oppijoiden arvoja tai piileviä asenteita.

Asenteiden, uskomusten, mielipiteiden ja tunnereaktioiden yhteydestä kriittiseen reflektointiin, joka on itsearvioinnin peruskiviä, on julkaissut tutkimuksia muun muassa Mezirow (1991). Jos halutaan muuttaa jo olemassa olevia oppijan sisäistämiä merkitysskeemoja, on asetettava kyseenalaisiksi monenlaisia psykologisia, kulttuuriin liittyviä ja epistemologisia tekijöitä. Esimerkiksi psykologiset tekijät ovat tiedostamattomia toimintaa estäviä sääntöjä, joiden rikkominen aiheuttaa ahdistuneisuutta. Osa oppijoista saattaa määritellä kyvyn (*ability*) olevan jotakin muuttumatonta ja pysyvää, kun taas toiset saattavat yhdistää kyvyn olevan jotakin muutokseen ja kasvuun liittyvää (Dweck 1999). Epistemologiset käsitykset liittyvät uskomuksiin tiedon luonteesta, lähteestä ja käytöstä. Tutkijat ovat osoittaneet, että oppijoiden välillä on suuria eroja siinä, mitä he uskovat tiedon olevan. Tiedon määrittelyyn ja hahmottamiseen liittyvillä uskomuksilla on vaikutuksensa tapaan, jolla oppijat oppivat ja tekevät arvioita (esim. Baxter Magolda 1992). Uskomuksiin liittyviä variaatioita on kutsuttu "tietämistavoiksi" (*ways of knowing*). Kutsumme näitä tapoja tässä tutkimuksessa tiedon hahmottamistavoiksi. Baxter Magolda on luokitellut tiedon hahmottamistapoja neljään laadullisesti erilaiseen luokkaan. Ensimmäinen tavoista on *absoluuttinen* tiedon hahmottamistapa, jota luonnehtii selkeä mustavalkoisuus

(jokin on oikein tai väärin tai kaikkien ongelmien sisältö on selkeästi määriteltävissä, ja lisäevidenssi on siis tarpeetonta). Toinen tapa hahmottaa tietoa on *siirtymään liittyä*. Tässä tavassa hyväksytään se, että tieto on varmaa joillakin alueilla, kun taas toisilla alueilla se on epävarmaa. Siten lisäevidenssin avulla voidaan päästä hyväksyttävään lopputulokseen. Kolmas eli *itsenäinen* tiedon hahmottamistapa sisältää uskomuksen tai oletuksen siitä, että tieto on epävarmaa. Esimerkiksi opettajia ei enää pidetä heidän asemansa ansiosta ainoina mahdollisina tietolähteinä, vaan oppijan oma mielipide voi olla yhtä arvokas kuin opettajankin mielipide. Tutkijoiden mukaan tähän hahmottamistapaan liittyy kuitenkin kielteisenä puolena kaiken hyväksyvä ”anything goes” - asenne, joka ilmenee muun muassa siten, että lisäevidenssin hankkimisen tarve koetaan vähäiseksi. Neljättä tiedon hahmottamistapaa kutsutaan *kontekstuaaliseksi*, koska siinä korostetaan, että tietoa arvioidaan tiettyyn kontekstiin liittyvän evidenssin avulla. Itsearviointiin voidaan todeta liittyvän kolmanteen ja neljanteen tietämistapaan, joissa korostuvat toisaalta oppijan ja tietoa auktoriteetin välinen jaettu vastuu ja toisalta tiedon ja oppimisen sijoittaminen tiettyyn kontekstiin. Tämän alan uudempaa tutkimusta edustavat King ja Kitchener (2004), jotka kutsuvat tutkimusaihettaan reflektiivisen arvioinnin kehittämiseksi (*development of reflective judgement*).

Jotta oppijat voisivat sitoutua kriittiseen reflektiivisyyteen, jota itsearviointinissa tarvitaan, heidän tulee muuttaa tietoon liittyviä käsityksiään ja uskomuksiaan. Vaadittu muutos edellyttää, että oppijat oppivat tunnistamaan ja olemaan kriittisiä niillä alueilla, joita usein pidetään niin itsestään selvinä, ettei niitä osata edes kyseenalaistaa (esimerkiksi koulu ja siellä vallitsevat valtasuhteet ja käytänteet). Koska tiedon määrittelyyn ja hahmottamiseen liittyvillä käsityksillä on vaikutuksensa oppijoiden oppimiseen, sillä on merkitystä, mihin tiedon hahmottamisluokkaan oppija sijoittuu käsitystensä perusteella. Jotta oppija voisi sitoutua kriittiseen reflektiivisyyteen, hänen käsityksensä tiedosta tulisi muuttua absoluuttisesta hahmottamistavasta kontekstuaaliseen tiedon hahmottamistapaan. (Lucas & Tan 2013, 105–106.)

Lukuisissa itsearviointiin liittyvissä tutkimuksissa on tutkittu lähinnä oppijoiden, opettajien tai esimerkiksi kieliassiantuntijoiden asenteita ja/tai näkemyksiä joko yleensä itsearviointia tai sen antamia tuloksia kohtaan.

Smith (1997) suoritti englanti vieraana kielenä kontekstiin sijoittuvan itsearviointitutkimuksen Israelissa. Tutkimustulokset osoittivat, että 12. luokan oppilaat pitivät itsearviointia validimpana kuin opettajan suorittamaa arviointia. Opettajat suhtautuivat hyvin luottavaisesti oppilaidensa itsearviointikykyihin muissa kuin nk. high stakes -tilanteissa, joissa opettajat epäilivät oppilaidensa rehellisyyden joutuvan koetukselle.

Ericksonin (1996) tutkimus liittyy suoraan oppijoiden asenteisiin, joita heillä on itsearviointia kohtaan. Erickson vertasi tyttöjen ja poikien asenteita, kun he olivat osallistuneet kielitestiin ja olivat juuri aloittamassa itsearviointia. Tyttöjen asenne kielitestissä onnistumista kohtaan oli paljon pessimistisempi kuin poikien.

Suomessa vuosina 2006–2009 toteutetussa suomalaisessa *Muuttuvat teksti- ja mediamaisemat* -hankkeessa (ToLP-hanke) tarkasteltiin tekstikäytänteitä käy-

tön, oppimisen ja opetuksen sekä arvioinnin ja identiteetin näkökulmista suomalaisessa yläkoulussa. Vaikka Suomessakin oppijakeskeinen pedagogiikka on ollut esillä jo vuosia, muun muassa kyseisen hankkeen tutkimustulokset oppilaiden roolista koulujen arviointikäytänteissä kertovat siitä, että opetus on kuitenkin edelleen hyvin opettajajohtoista ja opettajan määräämää (Luukka ym. 2008). Arviointi- ja palautekäytänteisiin liittyvät tulokset vahvistavat samaa näkemystä sekä sitä, että itsearvioinnin harjoittaminen ja muun muassa sitä koskevat näkemykset voivat vaihdella suurestikin opettajien ja oppilaiden välillä. Kun ToLP-hankkeessa selvitettiin kyselylomakkeen avulla opetuksen ja arvioinnin käytänteitä sekä oppilaiden roolia arviointikäytänteissä, kävi selville muun muassa se, että 93–94 % yhdeksännen luokan opettajista vastasi, että heidän oppilaansa harjoittivat itsearviointia ainakin joskus. Sitä vastoin saman tutkimuksen kyselyyn vastanneista oppilaista vain 63 % (äidinkieli) ja 65 % (vieras kieli) oli sitä mieltä, että itsearviointia harjoitettiin kouluissa ainakin joskus (Huhta & Tarnanen 2009, 7). Tutkimuksessa tarkasteltiin myös, millaisia käsityksiä ja arvoja sisältyy mm. opetus- ja arviointikäytänteisiin sekä itsearvioinnin vaikutusta oppilaan saamiin arvosanoihin. Kaikki tutkimuksen kohteena olleet opettajat ja 87 % kohteena olleista oppilaista vastasi, että opettajan arviointi vaikutti huomattavasti sekä äidinkielen että vieraiden kielten arvosanoihin. Vain pieni vähemmistö, 3–5 % opettajista, oli sitä mieltä, että itsearvioinnilla on suuri vaikutus oppilaan saamiin arvosanoihin. Huomattavaa on, että kyselyyn vastanneista oppilaista 10 %, ts. suurempi osa kuin heidän opettajistaan, uskoi, että itsearvioinnilla todella on tärkeä vaikutus heidän saamiinsa arvosanoihin. (Huhta & Tarnanen 2009, 9).

ToLP-hankkeen kyselylomakkeen väittämällä pyrittiin selvittämään asenteiden ja mielipiteiden muodossa, missä määrin oppilaat ja opettajat olivat samaa mieltä siitä, kuinka realistisesti omien taitojen itsearviointi oppilailta sujui. Merkille pantavaa oppilaiden itsearvioinnin hallintaan liittyvien asenteiden suhteen oli se, että 44 % äidinkielenopettajista ja 42 % vieraan kielen opettajista oli joko osittain tai täysin eri mieltä väittämän sisällön kanssa. Näyttää siis siltä, että melkein puolet molempien ryhmien opettajista ei luota siihen, että heidän oppilaansa osaisivat arvioida omaa osaamistaan realistisesti. Tästäkin asiasta tutkimuksen kohteena olleiden oppilaiden asenteet oman osaamisen itsearvioinnin hallintaa kohtaan olivat jonkin verran myönteisemmät kuin heidän opettajiensa asenteet: 31 % oppilaista oli väittämän kanssa ainakin osittain eri mieltä (Huhta & Tarnanen 2009, 12). Oppilaiden ja opettajien asenteista on lisäksi todettava, että vaikka itsearviointia harjoitetaankin yläkoulussa jonkin verran, se ei välttämättä tarkoita sitä, että itsearviointia pidettäisiin tärkeänä ja yleisesti hyödyllisenä. Itsearvioinnin toteuttamistavoissa voi olla myös toivomisen varaa: joissakin tapauksissa itsearviointi on toteutettu pinnallisesti ja ei-reflektoivasti muodikkaana toimintana, jolla ei ole yhteyttä opettamiseen ja oppimiseen. (Tarnanen & Huhta 2011, 12.)

Myös Lindblom-Ylänne ym. (2006, 60) totesivat itse- ja vertaisarviointia sisältävässä tutkimuksessaan, joka tehtiin yliopistossa 15 opiskelijalle oppimis-

päiväkirjojen avulla, että samansuuntaisten tulosten lisäksi oppijoiden asenteet näitä arvointimuotoja kohtaan olivat hyvin myönteisiä.

Oppijoiden suhtautumista itsearviointia kohtaan ovat tutkineet myös Andrade ja Du (2007). Heidän tutkimustuloksensa osoitti, että oppijoiden asenteet muuttuivat myönteisemmiksi, kun itsearviointia harjoitettiin säännöllisesti tietyn ajanjakson kuluessa. Samalla myös oppijoiden motivaatio kasvoi ja oppiminen lisääntyi.

Huhta (tulossa) tutki sitä, millaisina tietyn ryhmän edustajat, tässä tapauksessa koulutetut kieliasiantuntijat ja ei-kieliaineiden opiskelijat, pitivät DIALANG-arviointityökalun itsearvioinnin hyötyjä. Asiantuntijat pitivät suurimpana hyötynä sitä, että järjestelmä yhdistää itsearvioinnin ja testituloksen ja mahdollistaa siten näiden kahden vertaamisen. Seuraavaksi tärkeimpänä hyötynä pidettiin DIALANG-arviointityökalun itsearvioinnin ja käytettyjen itsearviointikuvainten yleistä laatua ja laajaa kielivalikoimaa. Hyötyinä mainittiin myös mm. itsearvioinnin itsetuntemusta ja tietoisuutta kasvattavaa vaikutusta, motivoimista, tietokonepohjaisuutta sekä järjestelmän itsearvioinnin nimettömyyttä. Myös opiskelijat pitivät suurimpana itsearvioinnin hyötynä mahdollisuutta verrata itsearvioinnin ja testin tulosta. Opiskelijoiden vastauksissa korostuivat jokseenkin samat seikat kuin kieliasiantuntijoidenkin, mutta niiden tärkeysjärjestys vaihteli hieman. Opiskelijoiden vastauksista selvisi kuitenkin ehkä yllättävästi, että he eivät tuntuneet pitävän DIALANG-järjestelmän käyttöön liittyvää itsenäisyyttä, joustavuutta, tietokonepohjaisuutta tai omasta oppimisesta tietoiseksi tulemistä maininnan arvoisina seikkoina. Sitä vastoin tutkimukseen osallistuvat kieliasiantuntijat olivat sitä mieltä, että kaikki nämä elementit olivat tärkeitä järjestelmää käyttävän oppijan näkökulmasta (Huhta tulossa, 10).

### **3.6 Itsearviointi opettamisen ja opetussuunnitelmien näkökulmasta**

Oppijakeskeinen opetus tuottaa monipuolisia opetusmenetelmiä. Näitä ovat esimerkiksi oppilaan itsenäisen työskentelyn muodot, oppilasryhmän työskentely sekä työskentely, jossa oppilas toimii ”opettajana”. Aktiivinen toimintaan osallistuminen voi olla motivoivaa ja haastavaa. Tutkijat muistuttavat kuitenkin, että tämäntyyppisen opiskelun onnistunut toteuttaminen edellyttää välttämättä ainakin seuraavia asioita:

- oppimiselle on asetettu selkeitä kasvatuksellisia ja sisällöllisiä tavoitteita.
- oppilas on sisäistänyt ja ymmärtänyt oppimisen tavoitteet ja kohteet.
- opettajan asiantuntemus on riittävää opettamisen liittyvien kulttuuristen, psykologisten ja pedagogisten tekijöiden osalta (Krokkfors & Patrikainen 2006, 35).

Jos edellä mainitut perusasiat eivät ole kunnossa, saatetaan ajautua tilanteeseen, jossa itsearvioinnin harjoittaminen muuttuu itseisarvoksi – puuhasteluksi ja tekemiseksi tekemisen vuoksi, minkä seurauksena oppijat kokevat koko asian turhauttavaksi ja oppimista hidastavaksi tekijäksi (Krokkfors & Patrikainen 2006, 35.)

Itsearviointia harjoitetaan nykyisin kouluissa, ja erityisesti ala-asteella korostetaan oppijan oman sanallisen arvioinnin tärkeyttä (esim. Huhta 2003). Itsearvioinnin arvostus näkyy myös oppimateriaaleissa, jotka nykyisin sisältävät paljon aiheeseen liittyvää ainesta. Suomalaisessa koulujärjestelmässä itsearviointia siis pidetään tärkeänä ja siten sen kehittäminen ja käyttöönotto kouluissa onkin otettu huomioon jo lainsäädännön tasolla. Perusopetuslaissa säädetään siitä, että oppilaan oppimista, työskentelyä ja käyttäytymistä tulee arvioida monipuolisesti. Oppilaan oman itsearvioinnin voidaankin todeta toimivan opettajan suorittamaa arviointia monipuolistavana. Perusopetuslaki säättää myös siitä, että oppilaan arvioinnilla pyritään ohjaamaan ja kannustamaan opiskelua sekä kehittämään oppilaan edellytyksiä itsearviointiin (Perusopetuslaki 628/1998, 22 §).

Opetussuunnitelmien perusteet pohjautuvat funktionaaliseen näkemykseen kielen oppimisesta, ja niissä kannustetaan arvioinnin monimuotoisuuteen. Perusteissa korostetaan myös oppilaan aktiivista roolia oppimis- ja arviointiprosessissa, minkä tarkoituksena on se, että oppilas pystyisi arvioimaan omaa osaamistaan (POPS 2004, 262). Itsearvioinnin tukemisella pyritään siihen, että oppilas toisaalta oppisi tuntemaan itsensä paremmin ja luottamaan itseensä oppijana, ja toisaalta siihen, että hän oppisi hallitsemaan ja arvioimaan opiskelutaitojaan myöhempää elämää varten (mm. oman edistymisen ja asetettujen tavoitteiden seuraaminen sekä oman oppimisprosessin säätöminen). Säännöllisellä palautteen annolla on tärkeä merkitys sille, että oppilas oppii pohtimaan omaa oppimisprosessiaan ja ottamaan itsearvioinnin haltuunsa (POPS 2004). Oppilaan itsearviointiedellytysten kehittäminen mainitaan yhtenä perusopetuksen tehtävänä kansallisissa perusopetuksen opetussuunnitelman perusteissa. Opetussuunnitelman perusteissa mainitaan myös kielenoppimisen tavoitteena *oppimaan oppiminen*, johon kuuluu se, että oppilas oppii muun muassa arvioimaan omaa työskentelyään (POPS 2004, 260).

Myös lukiolaissa mainitaan perusopetuksen yhteydessä jo mainitut arvioinnin tavoitteet, ts. opiskeluun kannustaminen ja sen ohjaaminen sekä opiskelijan itsearviointiedellytysten kehittäminen (Lukiolaki 629/1998, 17 §). Kyseisessä lain kohdassa säädetään niin ikään oppimisen ja työskentelyn monipuolisesta arvioinnista. Lukion opetussuunnitelman perusteissa mainitaan opettavien oppiaineiden arvioinnista, että opetuksessa tulee kehittää opiskelijan itsearviointitaitoja. Lukiokoulutuksen opetussuunnitelman perusteet sisältävät myös viittauksia oppilaan itsearviointiin. Kurssisuoritusten arviointi -kohdassa itsearvioinnista todetaan muun muassa, että kurssin arvioinnin tulee olla monipuolista ja että opiskelijan oma itsearviointi voidaan ottaa huomioon käyttäen hyväksi muun muassa kurssin arviointikeskusteluja (LOPS 2003, 220).

### 3.7 Vertaisarviointi

Tämän tutkimuksen aiheena ei varsinaisesti ole kielitaidon vertaisarviointi. Tämä vaihtoehtoisin arviointitekniikoihin lukeutuva arviointi liittyy käsillä olevaan tutkimukseen epäsuorasti kielistudiossa tehdyn kuunneltaviin puhenäytteisiin perustuvan itsearviointin kautta. Kun tutkimukseen osallistujat vertasivat omaa suoritustaan nauhalta kuultuihin puhenäytteisiin, he suorittivat samalla tietynlaista vertaisarviointia. Vertaisarvioinniksi voidaan näet varsin hyvin ymmärtää juuri se, että oppijat vertasivat oman tuotoksensa piirteitä kuunneltavan näytteen piirteisiin ja vastasivat tutkimuksen kysymykseen omasta suoriutumisestaan verrattuna näytteen puhujaan. Lisäksi monet oppijat antoivat vielä kirjallista palautetta kuuntelemansa näytteen sisällöstä ja sen kielellisistä piirteistä. Oikeastaan vertaisarvioinnista jää tämän tutkimuksen osalta puuttumaan vain se, että oppijat eivät päässeet antamaan palautetta kuunnelluista näytteistä näytteiden tuottajille (jotka siis olivat viralliseen Yki-testiin osallistuneita henkilöitä). Sen sijaan oppijat pääsivät näytteiden avulla vertaamaan omaa suoritustaan, erittelemään sekä oman että kuunnellun näytteen ominaispiirteitä ja siten tulemaan tietoisemmiksi oman kielitaitonsa vahvuuksista ja heikkouksista suhteessa muihin kielenkäyttäjiin. Käsittelen seuraavassa vertaisarviointia näiden yhteyksien vuoksi lyhyesti ottamalla esimerkinomaisesti esille muutamia tutkimuksia, joiden tuloksia pidän mainitsemisen arvoisina.

Useissa tutkimuksissa on korostettu jo pitkään, että vertaisarviointi voidaan laskea kuuluvaksi nk. vaihtoehtoisten arviointimuotojen kuten itsearvioinnin ja portfolioarvioinnin joukkoon (esim. Boud 1990; Tudor 1996; Mondock 1997). Vertaisarviointi luokitellaan usein itsearviointia täydentäväksi arviointiksi (Lussier & Turner 1995, 150) tai itsearvioinnin yhdeksi muunnelmaksi (Atjonen 2007, 85). Viitekehys luokittelee vertaisarvioinnin kuuluvan muiden suoritettamaan arviointiin (EVK 2004, 261). Se eroaa varsinaisesta yksin harjoitettava itsearvioinnista siten, että vertaisarviointia harjoittaessaan oppija saa todellista arviointituntumaa toisten oppijoiden suorituksista, mikä sitten omalta osaltaan saattaa kehittää oppijan omaan suoritukseen tarvittavia metakognitiivisia strategioita (Lussier & Turner 1995, 150). Vertaisarvioinnissa korostuu eri toimijoiden välinen yhteistoiminnallisuus (Christie & Ferdos 2006). Ellis (2005) korostaa, että sosiokulttuurisesta näkökulmasta asiaa tarkasteltaessa yhteistoiminnallinen arviointi on välttämätöntä, jos halutaan saada tietoa siitä, mitä oppijat todella osaavat. Ellisin mukaan kyky osata jotain liittyy näet siihen, mitä oppijat pystyvät tekemään vuorovaikutuksessa toistensa kanssa eikä niinkään siihen, mitä he pystyvät tekemään yksin (Ellis 2005a, 228). Monista vertaisarvioinnin hyvistä puolista huolimatta osa sekä opettajista että opiskelijoista suhtautuu siihen varauksellisesti. Elsisen ja Juurakko-Paavolan mukaan (2006, 65) osa KORU<sup>12</sup>-hankkeeseen osallistuneista opettajista ei halunnut käyttää ver-

<sup>12</sup> Korkeakoulujen ruotsin kielen suoritusten arvioinnin yhteismitallisuuden kehittämishanke



taisarviointia sen vuoksi, että arviointi oli heidän mielestään epäluotettavaa ja jopa yksipuolisen ylistävää. Syynä vertaisarvioinnin arvosteluun oli tuolloin se, että opiskelijatoverin osaamista ei haluttu arvioida kriittisesti. Joidenkin opettajien mukaan vertaisarviointi onnistuisikin opiskelijoilta paremmin kirjallisesti kuin suullisesti.

Vertaisarvioinnissa harjaantuvat monet elämässä tarvittavat hyödylliset taidot, esimerkiksi yhteistyö- ja kommunikaatiotaidot. Vertaisarvioinnin hyvät puolet ovat osittain samoja kuin itsearvioinnin. Toisin sanoen sen sanotaan kehittävän pohdinta- ja päättelykykyä sekä korkeamman kognitiotason ajattelua, edistävän joustavaa, aktiivista ja pikemmin syvä- kuin pintatason oppimista (esim. Birdsong & Sharplin 1986; Gipps 1992). Tudor (1996, 182) toteaa osuvasti, että vertaisarviointi on käytännön tason oppijan valmennusta, joka kehittää oppijan ymmärrystä kielenkäytöstä ja tietyntyyppisistä vaikeuksista, joita oppija itsekin saattaa kohdata kieltä käyttäessään. Vertaisarvioinnin käyttöä suositellaan varauksetta itsearvioinnin opettamiseen, koska oppijat pohtivat yhdessä sekä oppimistavoitteita että arviointikriteereitä ja oppivat siten kriittistä pohtivaa toimintaa (Alanen & Kajander 2011, 77).

Vertais- eli toveriarviointia on käytetty jonkin verran erityisesti puhumisen arvioinnissa. Hong Kongiin sijoittuva Millerin ja Ngin (1994) suullisen kielitaidon arviointiin liittyvä tutkimus käsitteli muun muassa sitä, miten oppijat suhtautuivat vertaisarviointiin. Monilla oppijoilla oli hyvin negatiivinen käsitys vertaisarvioinnista. He pitivät sitä subjektiivisena ja epäreiluna tehtävänä ja pelkäsivät muun muassa kasvojensa menetystä opiskelutovereidensa edessä. Hieman yllättävä on tutkimuksen havainto, jonka mukaan kielten opiskelijat suhtautuivat vertaisarviointiin kielteisimmin. Yleisesti ottaen oppijat onnistuivat arvioimaan hyvin oman ryhmänsä jäsenten suullista kielitaitoa tietyissä olosuhteissa: arvioijien oma kielitaito sijoittui korkealle taitotasoasteikossa, ryhmä oli homogeeninen, ryhmän jäsenillä oli jonkinlainen aiempi käsitys toinen toistensa suullisen kielitaidon tasosta, testausympäristö ei ollut uhkaava ja oppijat osallistuivat testin suunnitteluun (ks. myös Miller & Ng 1996, 145).

Cheng ja Warren (2005) tekivät Hong Kongissa muiden kuin kielten opiskelijoiden vertaisarviointiin liittyvän tutkimuksen, johon kuului suullisen kielitaidon arvioinnin lisäksi myös muita arvioinnin kohteita, esimerkiksi ryhmässä suoritettavien tehtävien organisointi. Tutkimus osoitti, että siihen osallistuneet oppijat pitivät epämiellyttävämpänä sitä tutkimuksen osaa, jossa heidän piti antaa arvio tovereidensa suullisesta kielitaidosta (ks. myös Lindblom-Ylänne ym. 2006). Oppijat kertoivat tehtävän epämiellyttävyyteen syitä kysyttäessä, että toverin suullisen tai kirjallisen kielitaidon arviointi oli vaikeaa, koska he eivät tieneet, miten sitä pitäisi arvioida. Syynä osaamattomuuteen tuntui olevan se, ettei oppijoilla ollut selvää käsitystä siitä, millaista osaamista edustaa korkealle tai hyvin matalalle taitotasoasteikkoon sijoittuva kielitaito. Opiskelutoverin suullisen kielitaidon arvioinnin epämiellyttävyydestä huolimatta kielitaidon arviointi sujui oppijoilta aivan yhtä hyvin kuin muidenkin tutkimuksessa mukana olleiden arvioinnin kohteiden, kun sen tuloksia verrattiin opettajien antamiin arvosanoihin. Vertaisarviointia on käytetty tutkimusten mukaan pal-

jon myös kirjallisiin töihin: oppijat ovat korjailleet toisen tekstiä auttaen siten tekstin muokkaamisessa (esim. Rainey 1990; Murau 1993). Yleisesti voidaankin todeta, että monet vertaisarviointiin liittyvät tutkimukset osoittavat vertaisarvioinnin tuottavan samansuuntaisia tuloksia kuin ulkoisen arvioinninkin (esim. Forde 1996; Topping 1998; Campbell ym. 2001).

Suomessa Elsinen (2007, 363) on tutkinut oppijoiden suhtautumista vertaisarviointiin ja itsearviointiin. Elsinen toteaa, että oppijoiden asenteet vertaisarviointia kohtaan ovat yleensä hyvin myönteisiä, vaikka vain harvoilla heistä oli kokemusta kyseisestä arviointimuodosta. Tutkimuksessa mainitaan, että vertaisarviointi tuo esiin vuorovaikutuksen kautta toisen osapuolen näkemyksen ja siten laajentaa omalta osaltaan kuvaa arvioitavan henkilön kielitaidosta.

Yleisesti on todettava, että vertaisarvioinnin käyttöönotto saattaa vaatia paljon aikaa yhteiskunnassa, jossa suorittaminen, vertaaminen ja kilpailu ovat jokapäiväisiä ilmiöitä. Oppijoiden voi olla vaikea tottua toisaalta antamaan ja toisaalta vastaanottamaan palautetta omasta osaamisestaan luokkatovereiltaan. Atjonen muistuttaa, että voi olla yleensäkin vaikea antaa palautetta toiselle jostain sellaisesta asiasta, jota on itsekkin vasta opiskelemassa. (Atjonen 2007, 87).

Joidenkin tutkimusten tavoitteena on ollut verrata itse-, vertais- ja ulkoista arviointia toisiinsa. Matsunon (2009, 95) jo edellä mainitun tutkimuksen tulos oli, että vertaisarviointi toimi erittäin hyvin, sillä vertaisarvioijat olivat työssään sisäisesti yhdenmukaisia. Lisäksi heidän arviointinsa eivät olleet riippuvaisia heidän omasta kirjoittamissuorituksestaan. Tutkimus osoitti myös, että vertaisarvioinnissa oli vähemmän poikkeavuuksia kuin itsearvioinnissa tai opettajan suorittamassa arvioinnissa.

Lapsetkin pystyvät tekemään vertaisarviointia – ja yleensä pitävät sitä myönteisenä kokemuksena. Chalkia (2012) kokeili kolmen kuukauden ajan luokassaan englannin suullisen taidon itse- ja vertaisarviointia 11–12-vuotiaiden oppilaidensa kanssa. 18 oppilaasta 16 kertoi hyötynensä vertaisarviointiin liittyvästä vastuunottamisesta ja osallistumisesta. Suurin osa lapsista, jotka eivät aiemmin olleet kokeilleet itsearviointia, kertoivat tullessaan paremmin tietoisiksi oman kielitaitonsa heikkouksista ja vahvuuksista, parantaneensa suullista kielitaitoaan sekä kasvattaneensa itseluottamustaan. Itse- ja vertaisarvioinnin todettiin vaikuttaneen myönteisesti oppilaiden kriittisen ajattelukyvyyn kehittymiseen, kun he joutuivat miettimään mitä he tekivät, miten ja miksi he tekivät jotain, ja myös, kun he ilmaisivat mielipiteensä parinsa edistymisestä. (Chalkia 2012, 229, 232.) Jaakkolan ym. (2002, 46) mukaan vertaisarviointi vaikutti epäsuorasti 9. luokkalaisten lasten metakognitiivisiin taitoihin. Tutkijoiden mukaan oppilaat oppivat itsestään tarkkailemalla oppilastovereitaan.

### 3.8 Yhteenveto itsearvioinnista

Itsearviointia koskeva tutkimus- ja kehittämistyö on tuottanut monenlaisia tuloksia vuosikymmenien aikana. Valtaosa tutkimuksista tähdentää sitä, että oppijalle on hyötyä metakognitiivisten taitojen kehittymisestä sekä opintojensa aikana että luultavasti myöhemmin elämässään muutenkin. (esim. Oscarson 1989; EVK 2004; Atjonen 2007; Perrenoud 2011). Enemmistössä tehdyistä tutkimuksista on saatu todisteita itsearvioinnin ja ulkoisen arvioinnin välisestä selvästä, joskus vahvastakin yhteydestä (esim. LeBlanc & Painchaud 1985; Blanche 1986; Blue 1988; Lee-Ellis 2009). Joissakin tutkimuksissa näin ei ole kuitenkaan käynyt, vaan selvä yhteys on jäänyt osoittamatta (esim. Peirce, Swain & Hart 1993; Orsmond, Merry & Reiling 1997; Patri 2002; Matsuno 2009).

Itsearvioinnin käyttöä on puolustettu muun muassa siitä näkökulmasta, että se tekee opetus- ja arviointitilanteesta tasapuolisemman ja toteuttaa elinikäisen oppimisen kasvatustavoitetta (esim. Boud 1995; Atjonen 2007; Hildén 2011). Itsearviointi on eettistä toimintaa, sillä se laajentaa arviointipohjaa ja pienentää sen riskiä, että osa taidoista jäisi arvioinnin ulkopuolelle ja että jotkut oppijat tai oppijaryhmät joutuisivat marginaaliseen asemaan (Dragemark 2009, 234).

Kaikesta edistyksestä huolimatta on kuitenkin syytä kysyä, miksi kielten opetus on kuitenkin edelleen hyvin opettajajohtoista ja opettajan määräämää (Luukka ym. 2008). Mihin mahtaakaan perustua perinteiseen ulkopuoliseen asiantuntijan kompetenssiin luottaminen ja oppijan itsearvioinnin huomiotta jättäminen? Miksi monet itsearviointia vastustavat tai sen toimivuutta epäilevät ovat sitä mieltä, että ulkopuolinen auktoriteetti, kuten esim. opettaja, osaa paremmin arvioida oppijoiden kielitaitoa kuin oppijat itse? Itsearvioinnin käyttöä vastustavia näkemyksiä on monissa tapauksissa verraten vaikea ymmärtää, kun ottaa huomioon, että osa kielten opettajista ei ole vielä saanut varsinaista arviointikoulutusta. Heidän kielitaidon arviointia koskevat arvionsa saattavat siis perustua omiin vuosien varrella hyväksi koettuihin arviointikäytänteisiin, jotka eivät kuitenkaan aina perustu minkään systemaattisen ja todennettavan arviointikriteeristön varaan. Suomessa on kuitenkin viime vuosina järjestetty kielten opettajille runsaasti suullisen kielitaidon arviointiin liittyviä koulutuksia ja kehittämishankkeita (esim. kieltenopetuksen monipuolistamis- ja kehittämishanke KIMMOKE, seitsemää yliopistoa koskeva kielikasvatuksen tutkimus- ja kehittämishanke OSKU ja ammattikorkeakoulujen kielten opetuksen kehittämiseen liittyvät AMKKIA- ja KORU-hankkeet). Myös Eurooppalaisen viitekehyyksen julkaiseminen ja laaja käyttöönotto on lisännyt muun muassa arviointikäytänteiden monipuolistamiseen liittyvää tietoa (EVK 2004).

Itsearvioinnin kaltaisessa subjektiivisessa ja monipolvisessa operaatiossa sen tulokseen vaikuttavat lukuisat sosiaaliset, kulttuuriset ja psykologiset tekijät (Holec 1990, 45). Eri kulttuureja edustavien henkilöiden itsearviointia koskevassa suhtautumisessa näyttää siis olevan eroja. Esimerkkinä voidaan mainita joihinkin aasialaisiin kulttuureihin liittyvät taipumukset aliarvioida omaa

osaamista (ks. esim. Thomson 1996; Deckert 1990; Matsuno 2009). Ilmiö liittyyne ainakin osittain siihen, että kulttuurista riippuen opettajan ja oppijan välistä hierarkista eroa saatetaan korostaa voimakkaasti. Onkin oletettavaa, että itsearviointin käyttöönotto tämänkaltaisissa kulttuureissa vaatii tavallista suurempaa panostusta ja asenteiden muokkausta.

Selvää on myös, että muuttuakseen jatkuvaksi toimintatavaksi itsearviointi edellyttää aina itseohjautuvuuteen harjaantumista. Tällainen oppijoiden koulutus tietoisuuden tason nostamiseksi tunnetaan ranskankielisessä kontekstissa nimellä *évaluation formatrice* (kouluttava arviointi). Sen pääperiaatteena on subjektiivisen vaikutelman vertaaminen todellisuuteen (EVK 2004, 254). Onkin muistettava, että oppijan itsearviointitaitojen kehittymisen pitäisikin olla asteittain kehittyvä reflektointi- ja suunnitteluprosessi, joka kuuluu osana laajempaan kokonaisuuteen: oppija vastuutetaan omista (kieli)opinnoistaan (Tudor 1996, 170–171).

## **4 TUTKIMUSKYSYMYKSET, -AINEISTO JA -MENETELMÄT**

Luvussa 4 keskitytään esittelemään tutkimuksen tutkimuskysymykset ja tutkimuksen kulku. Tutkimuskysymykset esitellään aluksi luvussa 4.1. Luku 4.2 sisältää tutkimusmateriaalin valmisteluun liittyviä seikkoja, kuten esimerkiksi käytettyjen puhetehtävien esittelyn ja kuunneltavien näytteiden valinnan. Luvun 4.3 pääfokus on tutkimusaineiston keruuta koskevissa toiminnoissa. Tutkimusaineiston käsittely esitellään lopuksi luvussa 4.4.

### **4.1 Tutkimuskysymykset**

Tutkimuksella pyritään löytämään vastauksia seuraaviin kysymyksiin:

1. Kuinka toimivia ovat oppijoiden käyttämät erilaiset itsearviointimenetelmät?
2. Missä määrin erilaiset itsearvioinnit vastaavat toisiaan?
3. Missä määrin oppijoiden tekemät itsearvioinnit vastaavat ulkoista arviointia?
4. Mitä käsityksiä oppijat esittävät paperilla esitettyjen itsearviointiväittämien toimivuudesta?
5. Mikä on tyypillistä niille puhenäytteille, joissa oppijoilla oli eniten vaikeuksia verrata omaa puhetuotostaan kuunneltuihin näytteisiin? Entä mikä oli tyypillistä niille puhenäytteille, joissa esiintyy vähiten vaikeuksia?
6. Mihin kielen piirteisiin oppijat kiinnittävät huomiota arvioidessaan kielitaitoaan uuden itsearviointimenetelmän avulla?

Ennen tutkimuskysymysten tarkempaa esittelyä on syytä kommentoida lyhyesti tutkimuskysymyksissä 1 ja 4 sekä muuallakin tässä tutkimuksessa esiintyvää

toimivuus-käsitettä. Tässä tutkimuksessa toimivuus-käsitteellä tarkoitetaan sekä tietyn osion tai mittarin teknistä toimivuutta että sen laadullista dimensiota. Kun kyse on esimerkiksi jostakin tietystä osiosta, jota oppijat käyttävät itsearviointiin ja jonka toimivuutta he kommentoivat, toimivuus-käsitteellä voi olla myös käytettävyys-merkitys. Siten tässä tutkimuksessa toimivuus-sanana merkitys on ymmärrettävä melko väljästi.

Ensimmäinen tutkimuskysymys keskittyy tarkastelemaan tutkimuksessa käytettyjen eri itsarviointimenetelmien toimivuutta. Toimivuudella tarkoitetaan sitä, miten toisaalta perinteisen yleensä paperilla tehtävän itsearvioinnin kirjalliset itsearviointiväittämät ja toisaalta uudenlaisen puhenäytteiden kuuntelemiseen perustuvan itsearvioinnin puhenäytteet toimivat. Ensimmäisessä kysymyksessä keskitytään siis selvittämään sitä, miten kukin käytetty mittari toimii kokonaisuutena ja miten mittarin yksittäiset osiot toimivat. Vastauksia ensimmäiseen kysymykseen saadaan pääosin tilastollisten analyysien tulosten avulla. Eri osioiden hyvin tai huonosti toimimiseen voidaan kuitenkin saada lisävalaistusta myös väittämien laadullisista analyyseistä.

Toisen tutkimuskysymyksen avulla tarkastellaan sitä, missä määrin oppijoiden tekemät itsearviointit vastaavat toisiaan. Tässä tutkimuksessa kokeillaan uudentyyppistä kuunneltaviin puhenäytteisiin perustuvaa itsearviointia perinteisen paperilla toteutettavan itsearvioinnin rinnalla. Siten tähän tutkimuskysymykseen saatujen vastausten avulla tarkastellaan myös tulosten luotettavuuteen liittyviä seikkoja. Tällöin tarkastellaan toisaalta sitä, näyttävätkö erilaiset itsearviointitavat antavan samansuuntaisia tuloksia ja toisaalta sitä, näyttääkö uudentyyppinen itsearviointi antavan niin luotettavia tuloksia, että sitä voidaan varauksetta suositella vieraiden kielten opetus- ja oppimiskäyttöön. Toinen tutkimuskysymys liittyy läheisesti kolmanteen tutkimuskysymykseen, jossa tarkastellaan oppijoiden itsearviointien vastaavuutta ulkoisen arvioinnin kanssa. Tutkimuksen perustan vankka rakentaminen pohjautuu siihen, että työn aluksi analysoidaan käytettävät mittavälineet ja niiden luotettavuus ja toimivuus. Useisiin aiempiin eri itsearviointimenetelmien välistä positiivista suhdetta tutkineisiin tutkimuksiin viitaten (mm. Wood 1991, 231) voidaan olettaa, että tässä työssä käytettävät itsearviointimenetelmät tuottaisivat samansuuntaisia tuloksia ja että eri menetelmät korreloisivat siten positiivisesti keskenään.

Kolmannella tutkimuskysymyksellä pyritään siis saamaan näyttöä siitä, miten oppijoiden tekemät itsearviointit vastaavat ulkoista arviointia. Tutkimuksen lähtöoletus ulkoisen arvioinnin luotettavuuden osalta on se, että tehtävänsä koulutettujen arvioijien suorittama arviointi on lähtökohtaisesti luotettavaa ja että oppijoiden itsearviointeja voidaan verrata siihen. On kuitenkin täysin mahdollista olettaa myös, että oppijoiden itsearviointi olisi yhtä luotettavaa tai jopa luotettavampaa kuin ulkoinen arviointi. Tutkimuksessa ei oteta eksplisiittisesti kantaa siihen, kumpi arviointi, ulkoinen vai itsearviointi, tuottaisi luotettavampia tuloksia. Jotta tutkimus voidaan toteuttaa, itsearviointeja on voitava verrata johonkin konkreettiseen. Tässä tutkimuksessa konkreettista osaa edustaa ulkoinen arviointi, ja vertaamalla itsearviointia siihen voidaan selvittää näiden kahden arvioinnin välillä mahdollisesti vallitseva yhteys. Itsearvioinnin luotettavuuden yhteydessä todetaan usein, että vieraan tai toisen kielen oppijoi-

ta tulisi etukäteen kouluttaa itsearviointiin, jotta voidaan olettaa itsearvioinnin ja ulkoisen arvioinnin vastaavan toisiaan. Kuitenkin esitetään myös sellaisia näkökantoja, että suurimmalla osalla kielen oppijoista on tietynlainen kyky määritellä oman kielitaitonsa taso. Tällöin oppijoilla tulisi olla käytettävissään tätä tarkoitusta varten suunniteltu mittaväline, jonka avulla he voivat ilmaista sisäisiä havaintojaan ja näkemyksiään (Oscarson, 1978). On oletettavaa, että eri itsearviointien tulokset korreloisivat ulkoisen arvioinnin kanssa ainakin jossain määrin samansuuntaisesti. Tämän tutkimuksen tulosten pohjalta tullaan ottamaan kantaa myös näihin itsearviointia koskeviin näkemyksiin.

Neljäs tutkimuskysymys, joka edustaa laadullista analyysia, keskittyy selvittämään sitä, millaisia käsityksiä oppijat esittävät paperilla esitettyjen itsearviointiväittämien toimivuudesta. On selvää, että paperilla suoritettavassa itsearvioinnissa kirjallisten itsearviointiväittämien tulee olla kielellisesti riittävän selkeitä ja yksitulkintaisia, jotta niitä voidaan pitää tehokkaina ja luotettavina itsearviointitilanteessa. Oppijoiden kokemus kirjallisten itsearviointiväittämien toimivuudesta tai toimimattomuudesta on tärkeää tämäntyyppisten itsearviointivälineiden jatkokehittelyn näkökulmasta.

Viides tutkimuskysymys on luonteeltaan laadullinen, ja se on kaksiosainen. Kysymysten avulla pyritään tarkastelemaan sitä, mikä on tyypillistä niille puhenäytteille, joissa oppijoilla on eniten vaikeuksia verrata omaa puhetuotostaan kuunneltuihin näytteisiin. Edellisen vastakysymyksen avulla valotetaan niin ikään sitä, mikä on tyypillistä niille puhenäytteille, joissa esiintyy vähiten ongelmia sillä hetkellä, kun oppijat vertaavat omaa suoritustaan kuunneltuihin näytteisiin.

Kuudennen tutkimuskysymyksen avulla pyritään saamaan vastauksia siihen, mihin kielen piirteisiin oppijat kiinnittävät huomiota arvioidessaan kielitaitoaan uuden itsearviointimenetelmän avulla. "Vaikea päättää" -rastien pohjalta tehty analyysi toteutettiin siten, että tarkempaan analyysiin valittiin molemmista tehtävistä kuusi näytettä, jotka tuottivat eniten (3 kpl) tai vähiten (3 kpl) vaikeuksia silloin, kun oppijat vertasivat omaa suoritustaan kyseisiin puhenäytteisiin. Lähtökohtana tälle laadun tarkastelulle on oletus, että luultavimmin oppijoiden kommentit keskittyisivät tiettyihin kielen piirteisiin kuten esimerkiksi ääntämiseen. Tämän tutkimuskysymyksen työhön mukaan ottamisella pyritään kartoittamaan yleisiä tendenssejä siitä, mitkä kielen piirteet näyttävät nousevan esille heterogeenisen oppijaryhmän suorittamassa itsearvioinnissa, jossa omaa suullista suoritusta verrataan konkreettisiin puhenäytteisiin. Kuudennella tutkimuskysymyksellä on muita tutkimuskysymyksiä pienempi painoarvo, mutta se on mukana tutkimuksessa, sillä sen avulla voidaan hankkia lisätietoa siitä, millaisia huomioita ja kommentteja oppijoilla on kuunneltavista puhenäytteistä. Kuudes tutkimuskysymys tasapainottaa osittain myös tutkimuksen paperilla tehdyn ja kuunneltaviin näytteisiin perustuvan itsearvioinnin pohjalta saatuja tuloksia.

## 4.2 Tutkimusmateriaalin valmistelu

Työn aineiston keruu ajoittuu marraskuun 1999 ja maaliskuun 2000 väliseen aikaan. Aineistosta tehdyt analyysit ovat pääosin kvantitatiivisia, mutta työ sisältää myös merkittäviä kvalitatiivisia osia. Kvalitatiivisten osien tarkoituksena on ollut täydentävän tiedon hankkiminen oppijoiden itsearvioinnista sekä siihen liittyvistä argumenteista ja tendensseistä. Työn aineiston keruu edustaa nk. strukturoitua aineistonkeruuta, jossa analysoitavaksi tarkoitettu aineisto on kerätty kaikilta osallistujilta samalla tavalla käyttäen tähän tarkoitukseen suunniteltuja lomakkeita. Lisäksi kaikkien tutkimukseen osallistuneiden oppijoiden osallistumisolosuhteet kielistudiossa on pyritty pitämään täysin samanlaisina ryhmästä riippumatta siten, että tutkimuksen tekijä on henkilökohtaisesti vastannut kaikkien kielistudioryhmien istuntojen ohjaamisesta.

Tutkimuksen konkreettiset toteuttamisvaiheet voidaan jakaa karkeasti kuuteen pääkohtaan:

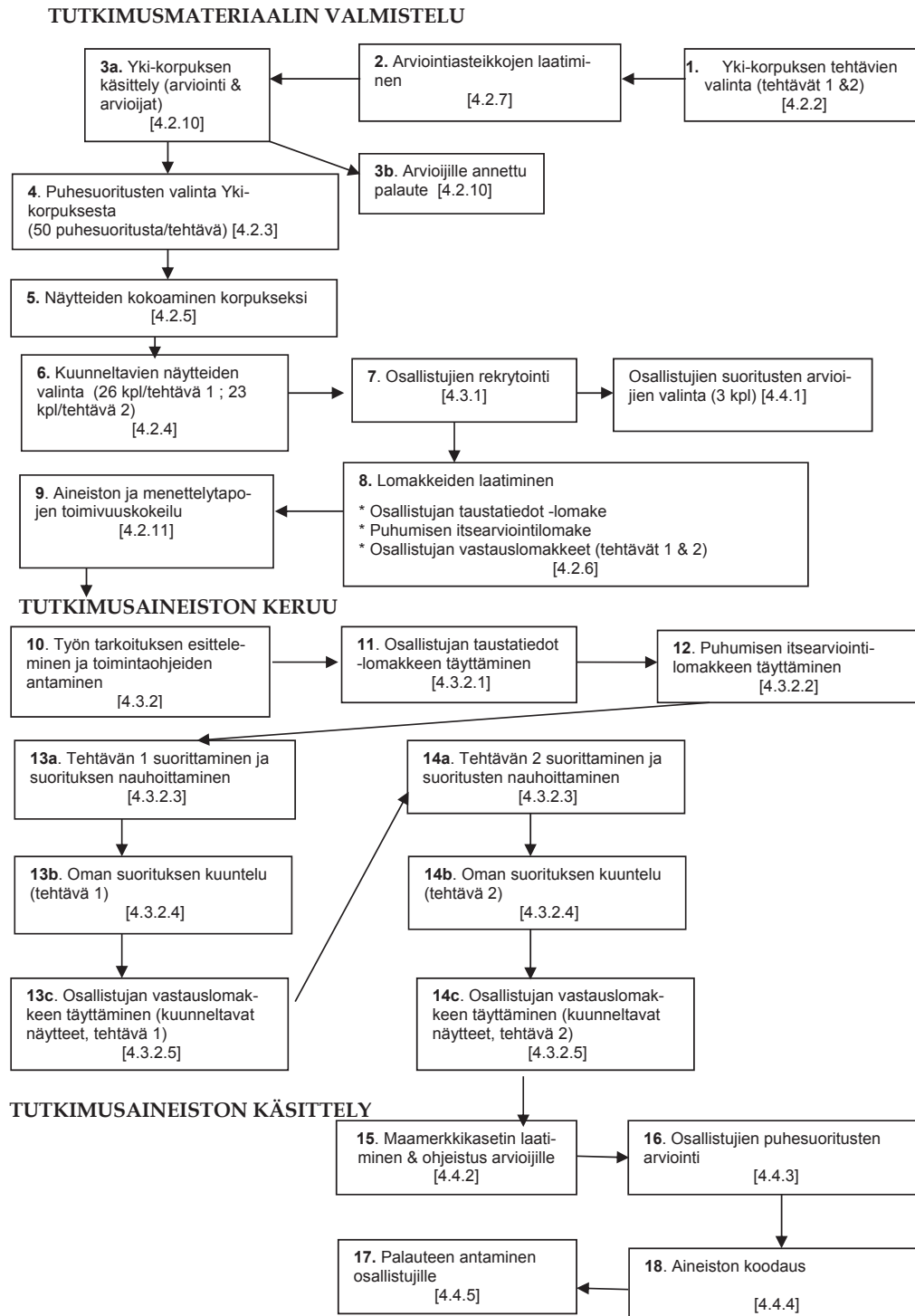
1. Yleisten kielitutkintojen ranskan kielen puhumisen osakokeista koottiin 50 puhe-suoritusta<sup>13</sup>/tehtävä sisältävä alkuperäiskorpus (vast'edes Yki-korpus), jonka suoritukset ulkopuoliset arvioijat asettivat taitotasoille.
2. Valittiin tietty määrä näytteitä tutkimukseen osallistuvien oppijoiden kuunneltaviksi (tehtävät 1 ja 2).
3. Oppijat suorittivat paperilla kielitaitonsa yleisitsearvioinnin ja 42-kohtaisen tilannekohtaisen itsearvioinnin.
4. Oppijat suorittivat tehtävän 1 ja kuuntelivat oman puhe-suorituksensa.
5. Oppijat vertasivat omaa puhe-suoritustaan nauhalta kuunneltaviin tehtävää 1 koskeviin näytteisiin.
6. Kohdat 4 ja 5 toistettiin samoin tehtävän 2 osalta.

Kuviossa 1 selostetaan tutkimusmateriaalin valmistelua, tutkimusaineiston keruuta ja tutkimuksen eri vaiheita kronologisessa järjestyksessä. Tutkimuksen kronologinen etenemisjärjestys on merkitty numeroin jokaisen kuviossa olevan laatikon alkuun. Laatikossa olevan tekstin jälkeen hakasulkuihin on merkitty numeroviite siihen tekstin kohtaan, jossa kyseistä tutkimuksen vaihetta selostetaan sanallisesti. Kuvion laatikoissa olevat otsikot voivat olla selostavassa tekstissä joko yläotsikkotasolla tai jonkin suuremman kokonaisuuden alla. Kuviossa olevat laatikosta toiseen etenevät nuolet (→) kuvaavat tutkimuksen kronologisia vaiheita.

Tutkimuksen kulkua selostavassa tekstissä esitellään aluksi tutkimuksessa käytetyt puhumistehtävät 1 ja 2, minkä jälkeen seuraa tutkimuksen eri vaiheiden selostaminen. Teksti ei etene välttämättä kaikilta osin kronologisesti tutkimuksen eri vaiheiden mukaan, vaan tekstissä on pyritty esittelemään tiettyyn kokonaisuuteen liittyviä toimenpiteitä ja asioita.

<sup>13</sup> Studiossa kuunneltavista puhe-suorituksista käytetään myöhemmin näyte -sanaa





KUVIO 1 Tutkimusmateriaalin valmistelun, aineiston keruun ja tutkimuksen eri vaiheet.

#### 4.2.1 Tutkimuksessa käytettyjen puhumistehtävien esittely

On tunnettua, että kielitaidon arviointia ohjaillaan sillä, minkätyyppisiä tehtäviä valitaan puhujien suoritettaviksi, koska tietyyppisissä tehtävissä mahdollistuu tietyyppinen puhe (esim. Luoma 2004; McNamara 1996; Fulcher 2003; Taylor 2011; Weir 1990, Weir 1993). Bachman ja Palmer (1996) korostavat, että kielitaidon arvioinnissa käytettävien tehtävien ominaisuuksia on pystyttävä erittelemään ja määrittelemään, jotta voidaan saada evidenssiä siitä, että suoritus tietyissä tehtävissä liittyy selviytymiseen vastaavankaltaisessa tilanteessa todellisessa elämässä. Muun muassa edellä mainitut tutkijat ovat kehittäneet yleisiä malleja, joiden avulla kielitaidon testaamisessa käytettyjä tehtäviä voidaan luokitella tehtävien sisältämien erilaisten piirteiden (esim. muoto, odotetun vastauksen kieli, vastavuoroisuus) perusteella (Fulcher 2003, 56). Weir on kehittänyt vastaavanlaisen luokittelumallin puhumisen arviointiin; mallissa korostuvat puhe-suorituksen tuottamisolosuhteisiin liittyvät asiat kuten esimerkiksi puheen tuottamiseen käytetty suunnitteluaika, puheen tarkoitus, keskustelijat sekä puhujien roolit ja aiheet (Weir 1993, 39). Puhumisen arvioinnin luotettavuuden kannalta on oleellista, että puheen tuottajalta saadaan puhenäytettä mahdollisimman monenlaisista puhetilanteista. Siten on tärkeää, että virallisissa puhumistaitoa arvioivissa testeissä samoin kuin puhumista koskevissa tutkimuksissa käytetään erityyppisiä ja -pituisia tehtäviä, jotka soveltuvat aiotuun tarkoitukseen. Tästä syystä tähän tutkimukseen valittiin kaksi erityyppistä Yleisten kielitutkintojen ranskan kielen puhumisen osakokeen tutkintotehtävää, joiden oletettiin antavan tutkimuksen kannalta riittävän monipuolisen kuvan tutkimukseen osallistuvien oppijoiden kielitaidosta.

##### 4.2.1.1 Tehtävä 1: Kuvasta kertominen

Tehtävä 1 edustaa tutkimukseen valituista kahdesta tehtävästä pidempää. Kyseinen tehtävä on mustavalkoiseen kuvaan perustuva kuvailutehtävä (liite 1), jossa François-niminen nuori mies istuu maassa pallo vieressään. Puhumistehtävän suoritusajaksi tutkimuksessa määriteltiin 1,5 minuuttia. Kyseessä ei ole spontaani puhetuotos, koska tutkimukseen osallistujilla oli käytettävissään 1,5 minuutin valmistelu-aika ennen puhe-suoritusta.

Brownin, Andersonin, Shillcockin ja Yulen (1984, 39) laatiman erityyppisten tehtävien vaikeusastetta koskevan luokituksen mukaan helpointa tehtävätyyppiä edustaa staattinen tehtävä, jossa puhuja kuvaa esimerkiksi diagrammia. Tehtävän 1 kaltainen mielikuvituksen käyttöä edellyttävä tehtävä voidaan tulkita dynaamiseksi kuvailutehtäväksi, koska puhujan tulee koostaa tarina tai koherentti ja yhtäjaksoinen puhuttu teksti, jolta odotetaan myös tiettyä pituutta. Tehtävien vaikeusasteeseen vaikuttaa Brownin ja Yulen (1983) mukaan myös esimerkiksi se, kuinka monta kuvattavaa asiaa tehtävässä on ja miten monipuolisia kuvattavat tapahtumat ovat. Tutkimuksessa käytetyssä tehtävässä 1 on siis yksi kuvattava nuori mies. Sen sijaan Yki-tutkinnossa olleessa alkuperäisessä tehtävässä tutkinnon suorittajalle annettiin neljän eri henkilön kuvat, joista hä-

nen tuli valita kaksi haluamaansa kuvattavaksi suullisesti. Tutkimuksessa käytetty tehtävä on siis helpotettu versio alkuperäisestä virallisesta Yki-tutkintotehtävästä, koska tutkimuksessa riittää vain yhden henkilön kuvaileminen. Tehtävän suorittaminen ei vaadi esimerkiksi monipolvisten tapahtumien kuvaamista, vaan osallistujalle jätetään verrattain vapaat kädet suullisen kuvailun sisällön suhteen.

Kuvasta kertomistehtävien hyvänä puolena voidaan Weirin (1990, 74) mukaan pitää sitä, että niiden avulla saadaan riittävän pitkä puhenäyte testattavilta ja siten ne mahdollistavat monien eri kriteereiden käyttämisen puhe-tuotoksen arvioinnissa. Tutkimusaineiston keräämisen ja suoritusten arvioinnin yhdenmukaisuuden näkökulmasta on syytä mainita kuvasta kertomistehtävän hyvänä puolena myös se, että tilanne, jossa puhe tuotetaan ja kuva, jota kuvailaan, on kaikille puheen tuottajille sama.

Kuvasta kertomistehtävien huonona puolena Weir (1990, 77) mainitsee muun muassa sen, että tehtävässä annettu kuva ei välttämättä kiinnosta testattavaa ja että kuvasta kertominen ei tehtävänä ehkä ole sellainen toiminto, mitä todellisessa kielenkäyttötilanteessa tehtäisiin. Onkin totta, että kuvailutehtävä, jossa kerrotaan kuvasta määrittelemättömälle henkilölle, joka ei edes ole paikalla kyseisessä puhetilanteessa, on melko kaukana todellisesta kielenkäyttötilanteesta. Weir huomauttaa myös, että mitä avoimempi kuvasta kertomistehtävä on, sitä enemmän suoritusten onnistumiseen saattavat vaikuttaa mielikuvituksen ja luovuuden sekä testattavan / puheen tuottajan taustan ja kulttuurin tapaiset asiat. (Weir 1990, 77.) Liikkumaton ja stabiili kuva saattaa myös tuntua puheen tuottajasta ei-motivoivalta, mikä taas voi vaikuttaa alentavasti puhe-tuotoksen sisältöön ja pituuteen. Kognitiivisia aktiviteetteja tarkasteltaessa tehtävän 1 kaltainen tehtävä rajoittuu ainoastaan kuvailuun. Puheen tuottajan ei myöskään tarvitse kuvailutehtävässä ponnistella esimerkiksi kysymysten esittämiseksi tai niihin vastaamiseksi tai aloitteellisena osapuolena olemiseksi vuorovaikutustilanteessa. Lisäksi kuvailutehtävät on mahdollista suorittaa suurelta osin käyttäen hyväksi ulkoa opittuja fraaseja. Tämä seikka saattaa johtaa siihen, että puheen tuottajaa ei missään puhetuotoksen vaiheessa haasteta yrittämään parastaan ja näyttämään, mihin asti hänen kielitaitonsa todella riittää.

Fulcher esittelee (2003, 57) puhumisen taitoa kuvailevan viitekehysten, jonka luokittelupiirteiden pohjalta on analysoitu tehtävän 1 ominaisuuksia taulukossa 1. Tähdellä (\*) merkityt ominaisuudet ovat peräisin Weirin (1993, 39) suullisen kielitaidon suoritusolosuhteisiin liittyvästä luokittelusta.

TAULUKKO 1 Tehtävän 1 ominaisuuksien luokittelua Fulcherin (2003) ja Weirin (1990) mallien pohjalta.

Luokittelupiirre		Kommentit
Tehtävän orientaatio	Avoin	Tuotos riippuu pitkälti puhujasta
Vuorovaikutussuhde	Ei-vuorovaikutuksellinen	Monologipuhetta
Päämääräorientaatio	Ei määritelty	Tehtävän avulla ei pyritä saavuttamaan mitään esim. kommunikaatiotilanteessa
Keskustelijoiden välinen status ja tuttuus	Ei keskustelijoita	Ei keskustelijoiden erilaisesta tai samanlaisesta asemasta johtuvia vaikutuksia
*Rekisteri	Epävirallinen	
Aihepiiri(t)	Urheilu, ihmiset	Melko tuttu useimmille kielenoppijoille
Tilanteet	Vapaa-aika, harrasteet	Melko tuttu useimmille kielenoppijoille

Taulukossa 1 esiteltyjen luokittelupiirteiden osalta voidaan vielä todeta, että tehtävän 1 aihepiiri ja epävirallinen rekisteri ovat suurimmalle osalle oppijoista tuttuja. Sen tähden on oletettavaa, että puheen tuottaminen tehtävään 1 liittävän kuvan pohjalta ei olisi kovin hankalaa ainakaan siitä syystä, että aihe olisi liian etäinen tai vähän tunnettu. Vuorovaikutustilanteeseen liittyvien statusroolien puuttuminen saattaa myös helpottaa jossain määrin joidenkin oppijoiden puheen tuottamista. Sen sijaan tehtävää 1 koskevan päämääräorientaation epämääräisyys voi hyvinkin olla puhemotivaatiota ja puhetuotoksen pituutta vähentävä ominaisuus.

#### 4.2.1.2 Tehtävä 2: Tilanteessa reagointi

Tehtävä 2 (liite 2) on kompleksisempi kuin tehtävä 1, koska siinä painottuu vuorovaikutuksen hallinnan lisäksi kielen sopivuus tilanteeseen. Tehtävän ala liittyy ammattikontekstiin, sillä tehtävässä oppijaa pyydetään esittelemään toisilleen puhujan kollega Kristiina Saukkonen-Riipinen ja tärkeä ranskalainen vieras, Julien Fabre-Perrault. Varsinaisen esittelyn lisäksi oppijan tulee kertoa ranskalaiselle vieraalleen, että hänen kollegansa puhuu hyvin ranskaa. Ennen varsinaista tehtävänantoa paperissa on selostettu kontekstia, johon puhetilanne sijoittuu. Selostuksessa kerrotaan, että puhuja on hakenut asemalta tärkeän (sana alleviivattu) vieraan, Julien Fabre-Perrault'n ja että puhuja ja vieras ovat jo esittäytyneet toisilleen. Tämän jälkeen he palaavat puhujan työpaikalle, missä tapahtuu tehtävänannossa kerrottu esittelytilanne.

Tehtävässä 2 on kyseessä tilannesimulaatio, jossa korostuu sosiaalinen vuorovaikutus ja jossa puhujan odotetaan reagoivan kommunikaatiotilanteessa mahdollisimman luontevasti. Tehtävää voitaisiin ajatella myös roolileikkinä, koska siinä tulee sijoittaa itsensä kuvitteelliseen kommunikaatiotilanteeseen, jossa on paikalla kolme eri rooleissa toimivaa henkilöä. Tilannesimulaation ja

roolileikin eroa on selitetty muun muassa siten, että simulaatiotilanteessa kielenkäyttäjä on tilanteessa, jossa hän voisi periaatteessa olla todellisessakin elämässä. Simulaatiotilanne on siten kielenkäyttäjälle tuttu, ja hän voi esiintyä tilanteessa enemmän tai vähemmän omana itsenään. Sen sijaan roolileikkitalanne edustaa kielenkäyttötilannetta, jossa kielenkäyttäjä ei yleensä ole, ja jota hän ei tavallisesti kovin hyvin tunne. Tilanteesta selvitäkseen kielenkäyttäjän tulee siten omaksua tietty rooli. (Huhta & Suontausta 1993, 242.) Tehtävän 2 määrittely joko simulaatio- ja roolileikkitehtäväksi askarruttaa erityisesti tutkimukseen osallistuneista lukiolaisista puhuttaessa. Voidaan näet varsin perustellusti kysyä, onko tämän ryhmän kannalta kyseessä simulaatio- vai roolileikkitehtävä. On otettava huomioon, että lukiolaiset ovat iältään sen verran nuoria, ettei heillä voida olettaa olevan kokemusta tehtävässä edellytettävästä ammatillisesta kontekstista. Tämän tosiasian perusteella voidaankin todeta, että nuorimpien tähän tutkimukseen osallistuneiden oppijoiden kohdalla kyseessä on pikemminkin roolileikki- kuin simulaatiotehtävä. Toisaalta on muistettava, että tilanne saattaa olla tehtävässä kuvattuun kontekstiin liittyen samankaltainen myös joillakin iäkkäimmillä tutkimukseen osallistuneilla. Eihän yliopistopöytäopiskelijoilakaan välttämättä ole minkäänlaista kokemusta tehtävässä 2 esille tulevasta tilanteesta. On vaikea määrittellä, onko kyseessä simulaatio- vai roolileikkitalanne, kun otetaan huomioon se, että periaatteessa hieman vanhemmilla osallistujilla voisi olla kokemusta kyseisen kaltaisesta tilanteesta, mutta heillä ei kuitenkaan sitä syystä tai toisesta ole ja siten tehtävässä kuvattu tilanne saattaa olla myös heille outo.

Tehtävä 2 vastaa hyvin Luoman (2004, 31) Bachmanin ja Palmerin (1996, 44) määrittelyn pohjalta Luoman edelleen kehittämää puhumisen tehtävätyyppejä, jossa puhumisen tehtävä nähdään aktiviteettina, jossa puhujat käyttävät kieltä saavuttaakseen jonkin erityisen päämäärän tiettyssä puhetilanteessa. Tehtävässä 2 puhumisella on selvä päämäärä ja tavoite: saada esiteltyä ranskalainen vieras puhujan suomalaiselle kollegalle ja tämän lisäksi kertoa lisätietona suomalaisen kollegan puhuvan hyvin ranskaa.

Taulukossa 2 esitellään tehtävän 2 ominaisuuksia samoin kuin edellä tehtiin tehtävässä 1, ts. taulukkoon on otettu luokittelupiirteitä sekä Fulcherin (2003, 57) että Weirin (1993, 39) malleista (viimeiset merkitty \*-merkillä).

TAULUKKO 2 Tehtävän 2 ominaisuuksien luokittelua Fulcherin (2003) ja Weirin (1990) mallien pohjalta.

Luokittelu		Kommentit
Tehtävän orientaatio	Suljettu	Tuotosta ohjaa tarkka tehtävänanto
Vuorovaikutussuhde	Kaksisuuntainen, mutta kyse vain toisen osallistujan puheenvuorosta	Simuloidaan vuorovaikutustilannetta
Päämääräorientaatio	Ei määritelty	Tehtävän avulla pyritään esittelemään henkilöitä toisilleen
Keskustelijoiden välinen status ja tuttuus	2 keskustelijaa	Keskustelijoina tärkeä ulkomaalainen ranskan kielen natiivivieras ja puhuja
Aihepiiri(t)	Sosiaalinen vuorovaikutus, työ	Tilanne sinällään monille tuttu, mutta voi olla vaikea vieraalla kielellä
*Rekisteri	Virallinen	
*Rooli	Professionaalinen: Kollega/kollega	Molemmilla puhujilla oletettavasti samankaltainen rooli (ei määritellä tehtävännossa)
*Tuttuus	Henkilöt eivät tunne toisiaan	Vaatii tietynlaista muodollisuusastetta
Tilanteet	Työ	Melko tuttu useimmille kielenoppijoille

Tehtävässä 2 edellytetään hyvin erilaista kielen hallintaa kuin kuvailevassa tehtävässä 1. Tehtävä 2 valittiin mukaan tutkimukseen sen vuoksi, että sen avulla pystyttäisiin arvioimaan tutkimukseen osallistuvien ranskan kielen käyttöön liittyvien kohteliaisuusmuotojen ja ammattimaiseen kielenkäyttötilanteeseen liittyvää kielen muodollisuusasteen hallintaa. Onkin selvää, että tämänkaltaisen kielen hallinta asettaa vaatimuksia erityisesti tutkimukseen osallistuneiden luokiolaisten osaamiselle. Suomen kielelle tyypillinen sinuttelumuotojen käyttö lähes kaikentyylisissä tilanteissa ja kaikenikäisten henkilöiden välillä muodostaa selkeän vastakohtan ranskan kielen teitittelyn vaatimukselle tehtävän 2 tilanteessa. Tehtävän 2 voidaan todeta edustavan suoraa tai autenttista testiä, joissa testaustilanne tai -tapa muistuttaa todellista tai luonnollista kielenkäyttötilannetta. Näin todettaessa on kuitenkin muistettava, että tilanne tapahtuu studiossa eikä kasvokkaisessa tilanteessa, mikä aiheuttaa tietynlaista epäluonnollisuutta (Huhta & Suontausta 1993, 228).

#### 4.2.2 Yki-korpuksen tehtävien valinta (tehtävät 1 ja 2)

Yki-korpuksen tehtävien valintaa koskeva työ aloitettiin tekemällä lista niistä tehtävistä, joissa oli ollut riittävä määrä suorittajia ranskan kielen Yki-

tutkinnossa. Riittävänä määränä tämän tutkimuksen kvantitatiivisten analyysin kannalta pidettiin vähintään 50 suorittajaa/tehtävä.

Toisena tehtävien valintakriteerinä toimi se, että valittavasta tehtävästä löytyi eritasoisia puhe-suorituksia. Ykin ylimmän tason suorittajamäärät ranskan kielessä ovat suhteellisen vähäiset, ja tästä syystä tehtävien valinnassa keskityttiin tehtäviin, joissa oli ollut suorittajia perus- ja keskitasolla. Näille kahdelle alimmalle tutkintotasolle keskittyvien suoritusten valinta oli perusteltua myös siitä syystä, että tutkimusta suunniteltaessa oletuksena oli, että tutkimukseen osallistuvien oppijoiden ranskan kielen suullinen taito sijoittuisi pääosin näille kahdelle taitotasolle.

Kolmantena tehtävien valintakriteerinä pidettiin tehtävien erilaisuutta, jonka ansiosta tutkimukseen osallistuvien puhutun kielen kielitaidosta voitaisiin saada laajempi kuva. Kahden eri tehtävän käyttäminen oli perusteltua myös siitä näkökulmasta, että tutkimukseen osallistuneille oppijoille voitiin antaa palautetta heidän puhumisen taitotasostaan kahden erityyppisen puhumistehtävän pohjalta.

Neljäs tehtävien valintakriteeri oli tehtävien pituus. Tehtäviä valittaessa oli otettava huomioon se, että tutkimukseen osallistumisen kokonaisaika rajoitettiin maksimissaan kahteen tuntiin. Tästä syystä tehtiin ratkaisu, että toinen valittavista tehtävistä olisi pidempi, ja toisen tehtävistä tulisi olla lyhyt.

Valittujen tehtävien suoritusten valinta aloitettiin kuuntelemalla läpi kaikki kyseiset tehtävät sisältävät kasetit yksitellen. Tämän jälkeen kaikista puhe-suorituksista laadittiin kommentoitu lista, johon merkittiin mm. kuunnellun puhe-suorituksen selvyys ja kesto, äänityksen laatu, puhenopeus, puhujan sukupuoli, tehtävänannon noudattaminen sekä tutkintotaso, jonka tutkintoon tutkinnon suorittaja oli osallistunut (perus- tai keskitaso). Puhe-suorituksia valittaessa otettiin huomioon myös suorittajan kansalaisuus; valittavista puhe-suorituksista valittiin vain suomalaisten puhe-suoritukset, jotta ne olisivat mahdollisimman vertailukelpoisia.

#### 4.2.3 Puhe-suoritusten valinta Yki-korpuksesta

Kun valittuja tehtäviä koskevat alkuperäiset puhe-suoritukset oli kuunneltu läpi ja valittu niistä yllä mainittujen kriteereiden pohjalta sopivimmat 50 kpl/tehtävä, molempien tehtävien valituista suorituksista koottiin kutakin tehtävää koskevat kasetit arvioijien käyttöön.

Kun valittuja alkuperäissuorituksia sijoitettiin arvioijien kaseteille, puhe-suoritusten järjestys oli muuten sattumanvarainen, mutta huomiota kiinnitettiin siihen, että mies- ja naisäännet vuorottelivat kasetilla aina kun se oli mahdollista ja että perus- ja keskitasolle osallistuneiden suorittajien puhe-suorituksia sijoitettiin jokseenkin tasaisin välein kaseteille. Koska valitut alkuperäissuoritukset molemmissa tehtävissä edustivat ainoastaan perus- tai keskitason puhe-suorituksia, ei niiden joukossa oletettu olevan edistyneemmän kielitaidon tason suorituksia (Ykissä ylin taso). Tutkimuskäyttöön tarkoitettujen puhe-suoritusten tulisi kuitenkin olla riittävän monipuolisia ja edustaa kaikkia kielitaidon tasoja. Tästä syystä päätettiin pyytää erikseen kahta ranskan kielen laudaturopiskelijää

suorittamaan kyseiset puhumistehtävät ja nauhoittaa heidän suorituksensa. Tällä tavalla saatiin tutkimuksessa käytettävien puhe-suoritusten joukkoon myös ylimmän tason puhe-suorituksia.

Seuraavaksi 5 arvioijaa arvioi edellä kuvatulla tavalla muodostetut ja kaseteille kootut suoritukset. Arvioijien valintakriteereistä ja taustasta kerrotaan tarkemmin luvussa 4.2.10.

#### 4.2.4 Kuunneltavien näytteiden valinta (tehtävät 1 ja 2)

Kun arvioijat olivat suorittaneet puhe-suoritusten taitotasolle asettamistyön, koottiin varsinainen tutkimukseen liittyvä kuuntelukorpus, jonka näytteet oli tarkoitettu tutkimukseen osallistuvien oppijoiden kuunneltaviksi kielistudiossa. Kuuntelukorpuksen jätettiin vain ne puhe-suoritukset, jotka vastasivat etukäteen päätettyjä valintakriteereitä, ja sen vuoksi kuuntelukorpuksen näytteiden määrä pieneni huomattavasti 50 alkuperäissuorituksen määrästä. Kahden eri tehtävän kuuntelukorpuksen näytteiden määräksi jäi 26 (tehtävä 1) ja 23 (tehtävä 2) kuunneltavaa näytettä.

Tehtävää 1 ja tehtävää 2 koskevien kuuntelukorpusten näytteitä valittaessa huomiota kiinnitettiin arvioijien yksimielisyyteen puhe-suoritusten tasosta. Toisena tärkeänä valintakriteerinä oli se, että kuuntelukorpuksen tuli sisältää erilaisia näytteitä ja siten sen piti koostua eritasoisista puhenäytteistä. Kolmantena yleisenä valintakriteerinä toimi riittävyden periaate, jonka mukaan näytteitä tuli erilaisuuden ja eritasoisuuden lisäksi olla mukana tarpeellinen määrä.

Tässä vaiheessa puhe-suorituksista tehtiin tilastollisia analyysejä, joissa selvitettiin yksittäisiä suorituksia koskevia tilastollisia tunnuslukuja, kuten keskiarvo, mediaani, moodi, standardipoikkeama ja varianssi. Analysoitavana olleet suoritukset järjestettiin tilastollisten tunnuslukujensa mukaisesti ryhmiin, joissa samanlaiset arvot saaneet suoritukset kuuluivat samaan ryhmään. Tämän ryhmittelyn pohjalta päätettiin sitten, kuinka monta näytettä kustakin ryhmästä valittiin kuuntelukorpuksen ja millä perusteella. Arvioijien välinen yksimielisyys suoritusten tasosta ei voinut olla ainoa suoritusten valintakriteeri, sillä valintapäätöksiä tehtäessä tuli varmistaa muun muassa se, että oppijoiden kuunneltavaksi tarkoitetut näytteet edustaisivat mahdollisimman monipuolista valikoimaa erilaisia puhe-suorituksia. Kuuntelukorpuksen laadun takaamiseksi päätettiin, että ne suoritukset, joiden varianssi oli 0 eli joiden tasosta vallitsi täydellinen yksimielisyys arvioijien välillä, otettaisiin pääsääntöisesti mukaan kuuntelukorpuksen. Täydellisen yksimielisyyden periaatteen noudattamisesta jouduttiin kuitenkin joustamaan tehtävässä 1 taitotasolle A1 arvioitujen puhe-suoritusten kanssa, koska tämätasoisia puhe-suorituksia oli jopa kuusi kappaletta. Jos kaikki kyseiset A1-tasoiset puhe-suoritukset olisi jätetty tehtävää 1 koskevaan kuuntelukorpuksen, olisi kuunneltavien näytteiden yhteismäärä kasvanut 30 näytteeseen, mikä taas olisi aiheuttanut tutkimukseen osallistumisajan pitenemisen yli kahteen tuntiin. Tämän vuoksi arvioijien täydellisestä yksimielisyydestä huolimatta A1-tasoisia näytteitä jätettiin tehtävän 1 kuuntelukorpuksen vain kaksi kappaletta kuudesta. Yleisenä nyrkkisääntöjä suorituksia valittaessa voidaan harvinaisen täydellisen yksimielisyyden lisäksi mainita se, että



mitä suurempi yksimielisyys suorituksen tasosta vallitsi arvioijien välillä, sitä parempi mahdollisuus suorituksella oli päätyä lopullisen kuuntelukorpuksen näytteeksi. Esimerkiksi suurin osa niistä yhteen ryhmään järjestetyistä suorituksista, joiden taitotasosta neljä viidestä arvioijasta oli samaa mieltä, päätyi kuuntelukorpuksen näytteeksi.

Suorituksia valittaessa mielessä jouduttiin siis koko ajan pitämään se, että tutkimukseen osallistumisaika oli rajoitettu. Valinnassa kiinnitettiin huomiota erityisesti kutakin suoritusta koskevaan standardipoikkeamaan ja varianssiin sekä siihen, että valittavat näytteet olisivat mahdollisimman laadukkaita myös äänityksensä puolesta. Osa 50 arvioidusta suorituksesta/tehtävä eliminoitiin välittömästi niiden analyysissä saamien tilastollisten arvojen vuoksi. Näissä suorituksissa esimerkiksi varianssi suoritettujen arviointien välillä oli tyypillisesti erittäin suuri (2 tai jopa 3 tasoa), mikä kertoo siitä, että arvioijien välillä oli huomattavia eroavaisuuksia suoritusten taitotasossa. Näytteiden valinnassa jouduttiin siis tekemään kompromisseja, joiden lähtökohtana oli toisaalta rajoitettu osallistumisaika ja toisaalta kuunneltavia näytteitä koskeva monipuolisuuden vaatimus. Tutkimusaineiston kokoamisen aikaan ei vielä ollut käytössä Facets-ohjelmaa, jonka avulla olisi voitu analysoida arvioinnit ja helpottaa siten lopullisten näytteiden valintaa.

#### 4.2.5 Näytteiden kokoaminen korpukseksi

Kun tutkimukseen osallistuvien oppijoiden käyttöön tarkoitettun kuuntelukorpuksen näytteitä järjestettiin kaseteille, otettiin huomioon osittain samoja seikkoja kuin alkuperäissuorituksia arvioijien käyttöön valittaessa. Mies- ja naispuhujien eritasoiset suoritukset sijoitettiin kaseteille mahdollisimman tasaisin välein, jotta kasettien kuuntelu ei muodostuisi liian monotoniseksi. Miehet suorittavat Ykin ranskan kielen tutkintoja huomattavasti vähemmän kuin naiset, ja tästä syystä miesten määrältään vähäiset puhe-suoritukset sijoitettiin suunnilleen tasavälein kasetin naispuhujien suoritusten joukkoon. Näytteitä kaseteille järjestettäessä kiinnitettiin huomiota myös siihen, että eripituiset puhe-suoritukset vuorottelisivat kaseteilla. Tätä kriteeriä pidettiin tärkeänä, jotta voitaisiin välttää pitkäväteisyyden tuntu kielistudiossa kasettien näytteiden kuunteluvaiheessa.

Kuuntelukorpuksen näytteet numeroitiin juoksevasti (1–26, tehtävä 1 ja 1–23, tehtävä 2) ja jokaisen kuunneltavan näytteen eteen äänitettiin teksti, jossa mainitaan kyseisen näytteen numero (esim. ”näyte 5”). Jokaisen näytteen jälkeen asetettiin äänimerkki, joka oli merkinä noin 12 sekunnin tauon alkamisesta. Tämän tauon aikana tutkimukseen osallistuvien oli tarkoitus verrata omaa suoritustaan nauhalta kuunneltuun näytteeseen ja vastata, selviytyikö hän omasta mielestään tehtävästä paremmin kuin kulloisenkin näytteen puhuja. Vastaaminen tapahtui rastittamalla joko ”kyllä” tai ”ei” itsearviointilomakkeesta jokaisen kuunnellun näytteen jälkeen sekä osallistujan niin halutessa kommentoimalla kuunneltua näytettä.

Koska kuuntelukorpusten tehtävien suoritusaikojen pituudet (tehtävä 1 pidempi kuin tehtävä 2) ja näytteiden kappalemäärät (26 kpl tehtävässä 1 ja 23

kappaletta tehtävässä 2) olivat erilaisia, tehtävän 1 kuuntelukorpuksen pituudeksi muodostui yhteensä 32 minuuttia, kun taas tehtävän 2 korpuksen pituus oli vain 14 minuuttia. Molempien tehtävien korpusten kuuntelu kielistudiossa kesti siis yhteensä 46 minuuttia. Varsinaisen näytteitä koskevan kuunteluajan lisäksi tutkimukseen osallistumisaikaan tulee lisätä tutkimukseen valmistautumiseen, osallistujan taustatietolomakkeen täyttämiseen, paperilla esitetyn itsearviointilomakkeen täyttämiseen sekä tehtävien 1 ja 2 suorittamiseen ja oman suorituksen kertaalleen kuuntelemiseen varattu aika. Näistä toiminnoista muodostuu tutkimukseen osallistumiseen varattu kaksi tuntia.

Molempia tehtäviä koskevista kaseteista tehtiin lisäksi ylimääräiset kopiot, jotka olivat aina käyttövalmiina kielistudiossa mahdollisten teknisten ongelmien varalta.

#### **4.2.6 Lomakkeiden laatiminen**

Tutkimusta varten laadittiin lukuisia erilaisia lomakkeita, jotka tutkimukseen osallistujat täyttivät. Luvun 4.2.6 alaluvuissa esitellään kolme lomaketta, joita tarvittiin tutkimusaineiston keräämiseen ja analysointiin.

##### **4.2.6.1 Osallistujan taustatiedot -lomake**

Tutkimukseen osallistujien taustatietoja koskeva lomake (liite 3) on laadittu tätä tutkimusta varten. Lomakkeella kerätyt tiedot keskittyivät tietoihin, jotka liittyvät osallistujan koulu- ja ranskan kielen opiskeluhistoriaan (esimerkiksi minkä ikäisenä ranskan kielen opinnot on aloitettu, ranskan kielen opintojen pituus eri kouluasteilla, ranskan kielen käytön frekvenssi ja luonne tutkimukseen osallistumishetkellä). Yhdessä lomakkeen kysymyksessä selvitettiin lisäksi tutkimukseen osallistuvan oppijan aiempaa itsearviointikokemusta. Lomakkeessa mainittiin myös, että tutkimukseen osallistuvien henkilöiden henkilö-, tausta- ja suoritustietoja käsiteltäisiin luottamuksellisesti ja että tutkimuksen aikana tehtävien suoritusten ja arviointien käsittely tehtäisiin nimettömänä.

##### **4.2.6.2 Puhumisen itsearviointilomake**

Kuten edellä on selostettu, tämä tutkimus liittyy kiinteästi kahteen kielitaidon arviointijärjestelmään, jotka ovat kansallinen Yleiset kielitutkinnot (Yki) ja yleiseurooppalainen DIALANG. Tutkimuksessa käytetyt alkuperäiset puhe-suoritukset otettiin Yki-tutkinnon puhe-suorituksista ja paperilla tehtyyn itsearviointiin käytettiin puhumisen itsearviointilomaketta (liite 4), joka oli kehitteillä DIALANG-hankkeen suullisen kielitaidon arviointityökaluja Internetiin kehitettäessä vuoden 2000 tienoilla. Hankkeessa laaditun itsearviointilomakkeen englanninkieliset väittämät käännettiin suomeksi Jyväskylän yliopiston Soveltavan kielentutkimuksen keskuksessa, joka tuolloin vastasi DIALANG-hankkeen koordinoinnista.

Puhumisen itsearviointilomake sisältää kolme osaa, joista ensimmäisessä tutkimukseen osallistuja arvioi yleisellä tasolla omaa ranskan kielen puhumisen

taitotasoaan. Osallistujan tuli merkitä rastilla (x) kuudesta puhumisen yleistasokuvauksesta se, joka parhaiten kuvaa hänen taitoaan puhua ranskaa. Yleistasokuvaukset vastaavat Euroopan neuvoston laatiman EVK:n kuutta taitotasoa A1–C2 ja tutkimukseen osallistujan tuli valita kuvauksista vain yksi.

Lomakkeen toisessa osassa on 42 yksityiskohtaista itsearviointiväittämää, joissa arvioidaan tarkemmin osallistujan ranskan kielen puhumisen taitoa erilaisissa puhumiseen liittyvissä kielenkäyttötilanteissa. Tähän lomakkeen osaan tutkimukseen osallistujan tuli vastata rastittamalla jokainen 42 väittämää joko ”kyllä” tai ”ei”-vastauksilla. ”Kyllä”-vastaus tuli valita, jos väittäjä sopi oppijan ranskan puhumisen taitoon, ja ”ei”-vastaus, jos väittäjä ei sopinut osallistujan puhumisen taitoon.

Taulukko 3 sisältää lomakkeen 42 itsearviointiväittämää jaoteltuina kolmeen ryhmään sen mukaan, millaisiin ja minkätasoisiin kielenkäyttötilanteisiin ne liittyvät. Suurin osa (23 kpl) itsearviointilomakkeen väittämistä liittyy jollakin tavalla kahden kielenkäyttäjän väliseen interaktioon. Kielen tuottamiseen liittyviä väittämiä taas on lomakkeessa vähiten (8 kpl). Eri taitotasolla on väittämiä 6–9 kappaletta valtaosan painottuessa asteikon alapäähän (A1 7 kpl, A2 8 kpl ja B1 9 kpl). Ylimmälle kolmelle taitotasolle (B2, C1 ja C2) sijoittuu kullekin 6 väittämää.

TAULUKKO 3 Puhumisen itsearviointilomakkeen väittämien luokittelu kielenkäyttötilanteen ja taitotason mukaan.

Kielenkäyttötilanne	Taitotaso							Yht.
	A1	A2	B1	B2	C1	C2		
Yleinen	1	-	3	1	2	4	11	
Interaktio	6	4	6	4	2	1	23	
Tuottaminen	-	4	-	1	2	1	8	
Yhteensä	7	8	9	6	6	6	42	

Lomakkeen kolmas osa sisältää mahdollisuuden antaa vapaita kommentteja. Tämän osan tavoitteena oli saada mahdollisimman paljon palautetta lomakkeen ja sen väittämien toimivuudesta. Osan täyttäminen oli vapaaehtoista, ja läheskään kaikki osallistajat eivät kommentteja antaneet. Osan aluksi oli vastaamisen helpottamiseksi kirjoitettu muutama kysymys, joissa kysyttiin muun muassa sitä, oliko oppijalla mitään erityisiä vaikeuksia vastata lomakkeen kysymyksiin (= väittämiin) sekä pyydettiin tarkentamaan, millaisia vaikeudet olivat, jos niitä oli. Lisäksi lomakkeessa kysyttiin, olivatko jotkut väittämät ehkä hankalampia kuin jotkut toiset; jos osallistuja oli tätä mieltä, häntä kehoitettiin merkitsemään kyseisen väittämän numero sekä selitys, miksi väittäjä tuntui hankalalta.

#### 4.2.6.3 Osallistujan vastauslomake

Osallistujan vastauslomakkeita (liitteet 5 ja 6) on yhteensä kaksi samansisältöistä eli yksi kutakin tehtävää varten. Osallistujan vastauslomake on 2-sivuinen lomake, jonka ensimmäisellä sivulla annetaan yksityiskohtaiset ohjeet siitä, mitä osallistujan odotetaan kielistudiossa tekevän. Lomaketta käytettiin kielistudioistunnon aikana, kun tutkimukseen osallistuva oppija vertasi omaa suoritustaan joko tehtävässä 1 tai tehtävässä 2 jokaiseen kyseistä tehtävää koskevaan näytteeseen. Osallistujan päätehtävänä oli vastata rastittamalla lomakkeeseen ”kyllä” tai ”ei” sen mukaan, kokiko hän suoriutuneensa tehtävästä paremmin kuin nauhalta kuultu puhuja. Lomakkeen vastausvaihtoehtojen määrä rajoitettiin kahteen (kyllä tai ei) tilastollisten analyysien tarpeiden vuoksi. On huomattava, että ”kyllä”-vastaus tarkoitti aina sitä, että oppija koki suoriutuneensa tehtävästä paremmin kuin näytteen puhuja. Sen sijaan ”ei”-vastauksella voitiin tarkoittaa kahta asiaa:

1. Oppija ei kokenut suoriutuvansa tehtävästä paremmin kuin kuunneltavan näytteen puhuja tai
2. Oppija koki, että hän suoriutui tehtävästä korkeintaan yhtä hyvin kuin näytteen puhuja ja koki siten olevansa suunnilleen samaa tasoa tämän kanssa.

Lomakkeen toisen sivun vasemmassa laidassa on kuunneltavien puhenäytteiden numerot, minkä jälkeen seuraavat valintaruudut, joihin oppija merkitsi itsearviointinsa (”kyllä” tai ”ei”) verrattuaan omaa suoritustaan nauhan kuhunkin näytteeseen. Valintaruutujen jälkeen lomakkeessa on vapaita kommentteja varten viiva, johon oppijoita kannustettiin kirjoittamaan, miksi he valitsivat joko ”kyllä”- tai ”ei”-vaihtoehdon. Valintaa tehdessään osallistujia kehoitettiin miettimään, mihin piirteisiin he kiinnittivät huomiota kuunneltavassa näytteessä verratessaan omaa suoritustaan näytteeseen. Kuten edellä esitellyssä lomakkeessa, myös tässä lomakkeessa vapaiden kommenttien antaminen oli vapaaehtoista.

Lisäksi lomakkeen 2. sivun oikeassa reunassa on ”vaikea päättää” -sarake, johon oppijat voivat merkitä rastin siinä tapauksessa, että he näytteiden kuuntelun jälkeen epäröivät ”kyllä-” ja ”ei”-vaihtoehtojen välillä. Kielistudiossa oppijoita muistutettiin kuitenkin siitä, että vaikka he merkitsisivät ”vaikea päättää” -sarakeeseen rastin, heidän päätehtävänsä tutkimuksen kannalta oli kuitenkin valita jokaisen kuunneltavan näytteen kohdalla joko ”kyllä”- tai ”ei” -vaihtoehto. Valinta ”kyllä” tai ”ei” -vaihtoehtojen väliltä oli siis pakollista kaikkien näytteiden osalta, ja kommenttien antaminen tai ”vaikea päättää” -sarakeen rastittaminen olivat molemmat suositeltavia mutta eivät pakollisia. Tämä suositeltavuus perustuu siihen, että tutkimukseen osallistuneet oppijat joutuivat keskittymään intensiivisesti työhön kielistudiossa kahden tunnin ajan, eikä heitä siten voitu velvoittaa täyttämään muita lomakkeen kohtia kuin kvantitatiivisten analyysien kannalta välttämättömät ”kyllä” tai ”ei” -valinnat. On

lisäksi huomattava, että esimerkiksi Eurooppalaisen kielisalkun<sup>14</sup> itsearviointiväittämiin vastataan niin ikään dikotomisilla kyllä/ei-vaihtoehdoilla. Vapaiden kommenttien antamiseen kuitenkin kannustettiin, koska niiden avulla toivottiin saatavan hyödyllistä tietoa niistä syistä, joiden perusteella oppijat vastasivat ”kyllä” tai ”ei”.

Osallistujan vastauslomakkeiden rakenteesta puhuttaessa on huomattava, että niiden ”kyllä” ja ”ei”-vaihtoehtoihin perustuva rakenne pohjautuu aineiston keruun kanssa samaan aikaan ajoittuneeseen jo edellä mainitun DIA-LANG-järjestelmän kehittämistyöhön, joka osaltaan siis on vaikuttanut niihin valintoihin, joita tässä tutkimuksessa on tehty.

#### 4.2.7 Arviointiasteikoista

Kielten arvioinnissa käytettäviä taitotasoasteikkoja tarvitaan silloin, kun arvioidaan kykyä selviytyä todellisen elämän vaatimuksista kielen avulla (EVK 2004, 251). Tässä tutkimuksessa tehtiin puhe-suoritusten arviointia kahdessa eri tehtävässä ja kahdessa eri vaiheessa: ensimmäisessä vaiheessa arvioitiin kielistudiossa kuunneltaviksi tarkoitettujen Yki-korpuksen suoritusten tasot ja toisessa vaiheessa tutkimukseen osallistuneiden oppijoiden puhe-suoritusten tasot. Suoritusten arvioinnissa käytettiin Eurooppalaiseen viitekehukseen pohjautuvia arviointiasteikkoja, joita muokattiin tutkimuksessa käytettyjen tehtävien arviointiin soveltuviksi (liitteet 7 ja 8).

Eurooppalaisen viitekehysten arviointiasteikkojen kehittämistyön pohjana on vuonna 1991 Rüşchlikonissa järjestetyn hallitusten välisen symposiumin päätös, jonka jälkeen toteutettiin vuosina 1993–1996 sveitsiläinen *Swiss National Science Research Councilin* asiaan liittyvä hanke. Hankkeen tarkoituksena oli kehittää osiopankki, joka sisältäisi kalibroituja ja läpinäkyviä kuvaimia kielen kommunikatiivisten taitojen hallinnasta EVK:ssa esitettävää kuvainjärjestelmää ja mahdollisesti myös Eurooppalaisen kielisalkun kehittämistyötä varten (EVK 2004, 293; North & Schneider 1998, 224). Ensimmäiset arviointiasteikot kehitettiin Yhdysvalloissa, (Fulcher 2003, 9) ja ennen EVK:n asteikkoja hyvin pitkään käytössä olivatkin lähestulkoon yksinomaan joko suoraan tai epäsuorasti amerikkalaiseen 1950-luvulla kehitettyyn FSI<sup>15</sup>-asteikkoon pohjautuvat arviointiasteikot (North & Schneider 1998, 217). Euroopassa on kuitenkin varsinkin 1990-luvulta lähtien tehty paljon kehitystyötä omien arviointiasteikkojen luomiseksi (North & Schneider 1998, 21).

Viitekehysten kuvaimet ja arviointiasteikot on toteutettu IRT:n (*Item Response Theory*) eli latenttien piirteiden analyysin avulla, jossa IRT on pohjana tiettyille mittaus- ja asteikon laatimismalleille. Ns. Raschin malli, joka perustuu todennäköisyysteoriaan ja jota käytetään pääasiassa yksittäisten koeosoiden vaikeusasteen määrittämiseen, edustaa voimakkainta näistä malleista. Vaikeustason määrittämisen idea on pelkistettynä seuraava: jos henkilön kielitaito on korkealla tasolla, hänen voidaan olettaa selvyvän hyvin alimmille tasoille

<sup>14</sup> [http://www.coe.int/t/dg4/education/elp/ELP-REG/Default\\_EN.asp](http://www.coe.int/t/dg4/education/elp/ELP-REG/Default_EN.asp)

<sup>15</sup> Foreign Service Institute

sijoittuvista osioista ja päinvastoin jos henkilön kielitaito on vielä alimmilla taitotasolla, hänen mahdollisuutensa selviytyä ylimpien taitotasojen osioista ovat huonot. Raschin malli mahdollistaa sen, että kielitaito ja tehtävien vaikeustaso voidaan ilmaista samalla asteikolla. (EVK 2004, 287–288, 293.) Raschin mallin käytöllä kuvainten analysoinnissa on lukuisia etuja, joista voidaan mainita mm. se, että kuvaimet on laadittu alun perin relevanteiksi oppimistapahtumaa ajatellen ja ne ovat itsenäisesti toimivia objektiivisesti asteikoille asetettuja kuvaimia. Lisäksi tällä tavalla analysoidut kuvaimet liittyvät mihin tahansa todellisen elämän kielenkäyttötilanteeseen eivätkä siis esimerkiksi johonkin yksittäiseen tehtävään. Kuvainten sisältö, ymmärrettävyys ja yhteinen tulkintatapa on lisäksi varmistettu käyttämällä informantteja, joiden kommenttien pohjalta huonosti toimineet, ei-relevantit tai kokonaisuuteen huonosti sopineet kuvaimet on voitu poistaa tai muokata uudelleen mahdollisimman aikaisessa vaiheessa. Raschin mallin yhtenä tärkeänä etuna verrattuna perinteiseen osioanalyysiin on se, että mallin mukaisten kuvainten voidaan sanoa olevan riippumattomia välineestä (*instrumental free*) tai henkilöstä, joka on esimerkiksi laatinut alkuperäiset kuvaimet; nämä molemmat piirteet lisäävät kuvainten ja kuvaimia sisältävien arviointiasteikkojen vakautta ja objektiivisuutta. (North & Schneider, 1998, 222–223; Gyagenda & Engelhard, 2009, 240.)

Viitekehyksen asteikkojen käyttö tässä tutkimuksessa on perusteltua, koska niitä on ollut yhteistyössä luomassa suuri joukko kielen opetuksen, oppimisen ja arvioinnin asiantuntijoita monista eri maista ja opetuskulttuureista ja siten niiden voidaan todeta olevan monipuolisen ja pitkäjänteisen yhteisen kehittämistyön tulos. Yhteiseen ja jaettuun näkemykseen perustuvien arviointiasteikkojen käyttö takaa sen, että oppijoiden puhesuoritusten arvioinnin voidaan odottaa olevan vakaalla pohjalla. Viitekehyksen asteikot ovat lisäksi empiirisesti validoituja arviointiasteikkoja, joiden tasoille on sijoitettu kielitaidon kuvauksia eri menetelmiä yhdistäen (intuitiivinen, laadullinen ja määrällinen vaihe sekä tulkintavaihe) (EVK 2004, 282, 294–296). Eurooppalaisen viitekehyksen asteikkojen käyttöä puoltaa myös se, että tämä tutkimus liittyy läheisesti Viitekehykseen tutkimuksessa käytetyn itsearviointilomakkeen kautta. Northin (2002, 100) mukaan DIALANG-hankkeessa validoitiin Viitekehyksestä lainattuja lukemisen, kirjoittamisen ja kuuntelemisen kuvaimia käyttämällä niitä itsearviointikuvaimina. Suomessa toteutetun kuvainten validointitutkimuksen (Kafandjieva & Takala 2002) ja sveitsiläisen validointitutkimuksen välinen korrelaatio oli korkea (.89). Puhumisen kuvaimia onkin käytetty Viitekehyksessä esimerkiksi kirjoittamisen ja tekstin ymmärtämisen kuvainten yhtenäistämiseen.

Arviointiasteikkoja on luokiteltu mm. niiden käyttötarkoitusten mukaan. Viitekehyksessä arviointiasteikkotyyppejä esitellään Aldersonin kehittämän kolmiosaisen kategorian mukaan (EVK 2004, 65–69):

1. testien käyttäjille suunnatut useimmiten holistiset ja myönteiseen sävyyn muotoillut yleistä kielen hallintaa kuvaavat asteikot (mitä oppija osaa tehdä)

2. testien arvioijille suunnatut yksityiskohtaiset suoritukseen laatuun keskittyvät asteikot, jotka saattavat olla negatiivisesti muotoiltuja (kuinka hyvin oppija suoriutuu tehtävästä)
3. testien laatijoille suunnatut kommunikatiivisiin tehtäviin liittyvät arviointiasteikot (mitä oppija osaa tehdä) (EVK 2004, 65–68).

Viitekehyksessä mainitaan yllä olevan arviointiasteikkokategorian lisäksi myös diagnostisesti suunnatut suoritusten laadullisia piirteitä painottavat arviointiasteikot, jotka on tarkoitettu erityisesti kielten opettajien käyttöön (EVK 2004, 68). North ja Schneider (1998, 218) ovat vertailleet edellä esitettyjä arviointiasteikkoja Bachmanin (1990, 325–330) arviointiasteikkokategorioihin. He toteavat, että Aldersonin kategoriassa testien arvioijille suunnatut arviointiasteikot vastaavat sisällöltään Bachmanin interaktiivista kielenhallintaa (interactive-ability) kuvaavia asteikkoja, koska tämäntyyppisten asteikkojen käytön tarkoituksena on tuoda yhteneväisyyttä ja johdonmukaisuutta suoritusten arviointityöhön. Toisaalta hyvin suuri osa kielitaidon arviointiasteikoista on Bachmanin (1990) mukaan todelliseen elämään liittyviä ja siinä tarvittavaa kielenkäyttöä ja käyttäytymistä kuvaavia arviointiasteikkoja.

Todettakoon, että tutkimuksen arviointiasteikoissa on käytetty elementtejä Viitekehysten monista eri arviointiasteikoista, joita on yhdistetty soveltuviksi kokonaisuuksiksi kahden tutkimuksessa käytetyn tehtävän arvioinnin kannalta. Asteikkojen laadinnan voidaan siten todeta olleen analyttistä siinä mielessä, että useiden olemassa olevien asteikkojen sisällöllisiä elementtejä on yhdistetty yhdeksi asteikoksi. Tämän tutkimuksen käyttöön laaditut asteikot ovat välimuotoja arvioijakeskeisistä ja suorittajakeskeisistä asteikoista. Arvioijakeskeisille asteikoille on tyypillistä se, että ne on tarkoitettu ohjaamaan arviointiprosessia keskittyen odotetun suoritusten laadulliseen puoleen. Suorittajakeskeisille asteikoille tyypillistä taas on se, että niiden avulla raportoidaan kielenkäyttäjän tyypillistä osaamista tietyllä tasolla. Suorittajakeskeiset asteikot ovat holistisia, ja niissä käytetään sanamuotoja, jotka ovat jokseenkin positiivisia, koska kielten oppijat käyttävät niitä oman edistymisensä todentamiseen. (EVK 2004, 65–68.) Koska tutkimuksessa laadittuja asteikkoja on käytetty sekä suoritusten arviointiin että palautteen antamiseen suorittajille, on selvää, että asteikkojen sanamuodot on pyritty pitämään niin positiivisina kuin mahdollista.

#### 4.2.7.1 Asteikkojen tasot

Eurooppalaisessa viitekehyksessä (EVK 2004, 39) muistutetaan, että tasojen lukumäärä ja niiden etäisyys toisistaan riippuvat paljolti siitä, millainen koulutusjärjestelmä on (silloin kun kyse on kielten opetuksesta) ja mihin tarkoituksiin asteikot laaditaan. Luoman (2004, 80) mukaan asteikkojen tasojen määrää päättäessä on otettava huomioon se, että jos on tarkoitus antaa palautetta oppijoille, palaute on sitä tarkempaa, mitä enemmän tasoa on käytössä. Toisaalta pitää huomioida, että tasoa pitää olla riittävän vähän, jotta arvioijat pystyvät erottelemaan niitä johdonmukaisesti. Asteikkojen tasojen määrä onkin yleisesti 4–6 tason välillä.

Vaikka Viitekehys suosittaakin usein sitä, että arviointiasteikot voidaan jakaa niin moniin alatasoihin kuin kulloinkin on tarve, tässä tutkimuksessa käytetään kuitenkin Viitekehysten kuutta perustaitotasoa (A1–C2). Tämä valinta on tehty siitä syystä, että kuudelle kehyksessä esitellylle perustasolle sijoittuvasta osaamisen tasosta vallitsee laaja konsensus. Perustasojen (A, B ja C-tasot) jakaminen kahteen eri tasoon yhden tason sisällä (esim. A1 ja A2) katsottiin sopivan hyvin tämänkaltaisen tutkimuksen tarpeisiin. Kuusi tasoa sisältävän asteikon käyttö on tässä tutkimuksessa linjassa myös tutkimuksessa käytetyn 6-tasoisien puhumisen itsearviointiasteikon kanssa. Myös puhumisen itsearviointilomakkeessa esitettyjen 42 yksityiskohtaisen itsearviointiväittämän taitotasot on asetettu kuudelle Viitekehysten perustasolle.

#### 4.2.7.2 Asteikkojen sisältö ja kuvainten muotoilu

Lukuisten tutkimusten mukaan ei ole olemassa konsensusta siitä, kuinka monta kriteeriä tulisi sisällyttää tietyn asteikon eri tasoille. Yleensä virallisten kielitutkintojärjestelmien tutkintojen asteikoissa käytettyjen kriteerien määrä vaihtelee kolmen ja kuuden kriteerin välillä. Kriteerien määrää ajateltaessa kyse on tietenkin siitä, että seikkoja, joihin arvioijien tulee kiinnittää huomiota arviointityön aikana, ei voi olla liian monta ilman, että jokin niistä saattaa jäädä vähemmälle huomiolle. Kyse on myös siitä, että arvioijien kognitiivinen suorituskyky saattaa häiriintyä, jos kriteereitä tai kategorioita on enemmän kuin neljä tai viisi (EVK 2004, 263).

Jotkut tutkijat ovat todenneet, että arviointiasteikkojen tulisi arvioida ainoastaan kullekin tasolle tyypillisimpiä kielitaidon komponentteja. Siten esimerkiksi ylimpien tasojen testisuorittajien ääntämisen arvioiminen saattaa tutkijoiden mukaan olla turhaa (Pollitt & Murray 1996).

Asteikoissa käytettyjen kuvainten muotoilu vaatii aina harkintaa. Viitekehyksessä todetaankin, että vaikka sen asteikot ovat syntyneet laajan kansainvälisen yhteistyön ja kompromissin tuloksena, asteikot ovat aina silti vain suositusluontoisia eivätkä käyttäjiä velvoittavia määräyksiä (EVK 2004, 13). Viitekehyksessä muistutetaan myös, että kehyksen käyttäjien tulisi pitää teosta eräänlaisena perusdokumenttina, jota voidaan tai jota jopa pitää muokata eri kontekstien tarpeisiin soveltuvaksi. Viitekehysten mukaan asteikoissa käytettyjen kuvainten tulisi olla myönteisiä, täsmällisiä, selkeitä, lyhyitä, läpinäkyviä, erillisinä toimivia, oppijan osaamista painottavia, tarkalleen määrättyä asiaa kuvaavia sekä toisistaan riippumattomia (EVK 2004, 55, 278–281).

On huomattava, että tämän tutkimuksen käyttöön laaditut asteikot ja niiden sanamuodot pohjautuvat vuonna 1998 laadittuun englanninkieliseen viitekehysten silloiseen versioon (luonnos 2), josta julkaistiin samana vuonna suomenkielinen versio (Huttunen & Huttunen 1998). Kyseisen luonnoksen asteikkojen muotoiluissa ja sanamuodoissa on eroa, jos niitä vertaa vuonna 2001 julkaistuun viralliseen englannin- ja ranskankieliseen Viitekehykseen tai WSOY:n vuonna 2004 julkaisemaan suomenkieliseen käännökseen. Tästä johtuen eri taitotasoille sijoittuvien kuvainten muotoiluissa on syytä tutustua edellä mainittuun luonnos 2:een vuonna 2001 ja 2004 julkaistun Eurooppalaisen viitekehys-



sen painetun version sijasta, jos haluaa verrata tutkimuksessa käytettyjen arviointiasteikkojen ja Viitekehyksen arviointiasteikkojen muotoiluja ja kuvainten sanamuotoja. Tutkimuksen aineiston laatimisesta on kulunut sen verran aikaa, että kehyksen asteikkoja on voitu uudelleen muotoilla käyttäjiltä saadun palautteen mukaisesti vuonna 2001 julkaistuu Viitekehyksen versioon. Onkin mielenkiintoista verrata tutkimuksen käyttöön laadituissa arviointiasteikoissa tehtyjä muutoksia uusimman Viitekehyksen asteikkojen muotoiluihin. Joissakin tapauksissa voidaan näet havaita, että tietyt kohdat, jotka korjattiin tutkimuksen käytössä olleisiin asteikkoihin, on muutettu myöhemmin samansuuntaiseksi Viitekehyksen viimeiseen vuodelta 2004 olevaan versioon.

Tässä tutkimuksessa käytettyihin arviointiasteikkoihin tehtiin joitakin muutoksia, joissa otettiin huomioon tehtävien 1 ja 2 vaatimukset. Käytettyjen asteikkojen muotoilussa yritettiin huomioida se, että asteikkojen kuvaimia käyttivät tai lukivat sekä kielitaidon arvioinnin asiantuntijat arvioidessaan osallistujien puhe-suorituksia että tutkimukseen osallistuvat oppijat lukiessaan suoritusistaan saamaansa palautetta. Muutoksia on tehty eniten asteikon alatasoille eli tasoille A1, A2 ja B1. Muutokset koostuvat pääosin alkuperäisten kuvainten lyhentämisestä ja liian hankaliksi määriteltyjen rakenteiden yksinkertaistamisesta. Kuvainten lyhentäminen on voitu tehdä joko siten, että liian pitkänä pidetty kuvain on pilkottu useaan lyhyempään osaan, tai siten, että kuvaimen jokin osa on jätetty pois – samanlaista kuvaimien muokkausta tehtiin myös tutkimuksen kuluessa meneillään olleessa DIALANG-hankkeessa. Kuvainten lyhentäminen olivat välttämättömiä myös käytännön näkökulmasta, koska tutkimuksessa käytetyt asteikot sisältävät elementtejä viidestä eri asteikosta. Kun kuvainten lyhentämisiä toteutettiin, piti siten tehdä kompromisseja sen välillä, mikä oli välttämätöntä jättää arvioitavan tehtävän kannalta ja mikä puolestaan vähemmän tarpeellista.

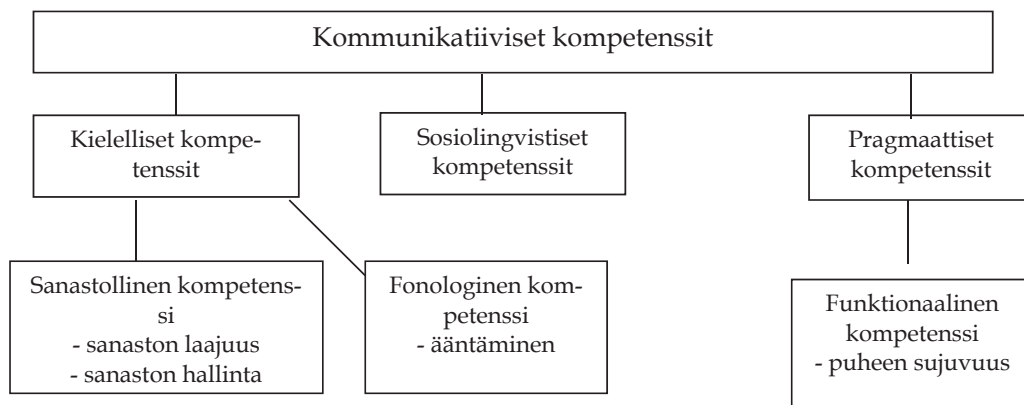
Joissakin tapauksissa kuvainten negatiivinen muoto muutettiin positiivisemmaksi. EVK:ssa suositellaan kuvainten myönteistä muotoilua ja todetaan, että useimpien olemassa olevien arviointiasteikkojen heikkoutena on se, että varsinkin niiden alimmat tasot sisältävät negatiivissävytteisiä ilmauksia. (EVK 2004, 247, 278). Samoin alkuperäisistä Viitekehyksen asteikoista jätettiin pois elementtejä, joita ei pidetty relevantteina käytettyjen tehtävien kannalta (esimerkiksi erilaiset viittaukset kommunikaatiotilanteeseen tehtävässä 1 ja/tai viittaukset konkreettisiin jokapäiväisiin tarpeisiin). Kuten tunnettua, Viitekehyksen asteikkojen kaikki taitotasot eivät aina välttämättä sisällä jotain tiettyjä kriteeriä koskevia kuvaimia. Siten esimerkiksi tehtävän 1 arviointiasteikon ylimmillä tasoilla on kuvaimia Viitekehyksen sanaston hallinta -nimisestä asteikosta, kun taas alimmalla tasolla A1 kuvaimia ei ole. Esittelen alla lyhyesti sitä, mistä Viitekehyksen luonnos 2:n asteikoista otettiin elementtejä tutkimuksen arviointiasteikkoihin. Lisäksi perustelen pääpiirteittäin niitä valintoja, joihin perustuen luonnoksen 2 kuvaimiin on tehty tiettyjä muutoksia.

#### 4.2.8 Tehtävän 1 arviointiasteikon esittely

Tehtävän 1 arviointiasteikossa (liite 7) on elementtejä seuraavista Viitekehyksen (luonnos 2) arviointiasteikoista:

- puheen sujuvuus
- sanaston laajuus
- sanaston hallinta
- fonologinen hallinta
- pitkä monologi: kokemuksen kuvaaminen.

Kaikki muut yllä mainitut asteikot paitsi monologi sijoittuvat kommunikatiivisen kompetenssin alle. Viitekehyksen mukaan kommunikatiiviseen kompetenssiin kuuluvat seuraavat kolme kompetenssia: kielelliset, sosiolingvistiset ja pragmaattiset kompetenssit. Nämä kolme kompetenssia jakaantuvat vielä alakompetensseihin kuvion 2 mukaisesti (EVK 2004, 155):



KUVIO 2 Tehtävä 1: arviointiasteikossa käytetyt kommunikatiiviset kompetenssit.

Puheen sujuvuus sijoittuu Eurooppalaisen viitekehyksen luokittelun mukaan pragmaattisen kompetenssin ja vielä tarkemmin sen alakompetenssin funktionaalisen kompetenssin alle. Sanaston laajuutta ja hallintaa sekä kielen ääntämistä pidetään kielellisinä kompetensseina. Kaksi ensimmäistä näistä kuuluvat sanastolliseen kompetenssiin ja viimeinen eli ääntäminen sijoittuu fonologisen kompetenssin alle. Yleisesti ottaen on muistettava, että mainittujen kolmen kommunikatiivisen kompetenssin pääkategorian (kielellinen, sosiolingvistinen ja pragmaattinen) alle sijoittuu Viitekehyksessä paljon muitakin alakategorioita kuin ne, joita tässä tutkimuksessa on käytetty ja jotka on esitetty edellä olevassa kuviossa. Siten kielelliset kompetenssit sisältää yhteensä 6, sosiolingvistiset 5 ja pragmaattiset kompetenssit 2 kompetenssia. Tutkimuksessa käytetty Pitkä monologi: kokemuksen kuvaaminen -niminen asteikko (uusimmassa Viitekehyksessä nimellä Puhuminen: kokemuksen kuvaaminen) ei kuulu lainkaan kom-

munikatiiviseen kompetenssiin, vaan se on sijoitettu kielen tuottamistoimintojen ja strategioiden alle (EVK 2004, 92).

#### *Puheen sujuvuus*

Kuvainten lyhentämisen lisäksi arviointiasteikon puheen sujuvuutta koskevia kuvauksia muotoiltiin uudelleen. Tasolla B1 EVK:n luonnoksessa 2 on kuvain, jossa puhutaan *ilmausten rakentelun umpikujista*. Umpikuja-sana on kuitenkin sanana niin vahva, että siitä tulee mielikuva kielenkäyttäjistä, joka ei saa puheutuostaan vietyä loppuun, vaan joka joutuu umpikujaan yrittäessään sanallisesti ilmaista asiaansa. Kuitenkin monilla kielenkäyttäjillä voi olla puhetuotoksessa hyvinkin monenlaisia ongelmia ilman, että heidän suullinen tuotoksensa päättyisi missään vaiheessa umpikujaan. Tutkimuksen käyttöön laaditussa arviointiasteikossa negatiivissävytteinen umpikuja-sana korvattiin siten *ilmausten rakentelun hankaluudet* -sanamuodolla. Vuoden 2004 Viitekehyksessä tästä samaisesta kohdasta puhutaan *puheen muotoiluun liittyvinä ongelmina*. B1-tasolle tehtiin toinenkin muutos: *pystyy jatkamaan puhettaan ja ilmaisemaan ajatuksensa ilman apua* on muutettu muotoon *pystyy jatkamaan puhettaan ja ilmaisemaan ajatuksensa kohtalaisen sujuvasti*. Asteikon muutettu muotoilu sopii paremmin tehtävään 1, jossa tutkimukseen osallistuvalla oppijalla ei ole puhekkumppania, joka häntä voisi auttaa. Lisäksi asteikosta on poistettu kaikki muutkin maininnat, jotka liittyvät jonkinlaiseen kommunikaatioon kahden henkilön välillä.

#### *Sanaston laajuus*

Viitekehyksen taitotasolla B2 mainitaan kiertoilmausten yhteydessä epäröinnit, jotka johtuvat aukoista sanavarastossa. Kiertoilmausten mainitseminen on jätetty tarkoituksellisesti pois tämän tutkimuksen asteikon tasolta B2, koska pidän kykyä kiertoilmausten käyttöön myönteisenä asiana siinä tapauksessa, että kommunikaatio tai tuotos ilman kiertoilmausten käyttöä muuten häiriintyisi. Näkemykseni mukaan kiertoilmausten käyttö on tapa saada sanoma perille ja saattaa tilanne loppuun, ja siksi kiertoilmausten käyttöön ei haluttu yhdistää sellaisia kielteisiä sanoja kuten puute tai epäröinti. Tasolla A2 olleet viittaukset rutiininomaisesta arkipäivän tiedonvälityksestä muutettiin, koska tämänkaltaiset kuvaukset liittyvät täysin erilaisiin tilanteisiin ja sopivat siten huonosti kuvasta kertomistehtävän arvioimiseen. Tilalle otettiin kuvain, jossa todetaan, että kielenkäyttäjällä on tällä tasolla riittävä sanavarasto tehtävästä selviytyäkseen.

#### *Sanaston hallinta*

Sanaston hallintaa koskevat muutokset arviointiasteikossa koskevat erityisesti tasoja B2, B1 ja A2. Tason B2 kuvainta lyhennettiin siten, että *hyvä ja täsmällinen sanaston hallinta* on jätetty pois, koska kuvain *vain joitakin sekaannuksia ja vääriä sanavalintoja esiintyy* pitää sisällään ajatuksen siitä, että tällä tasolla sanaston hallinta on yleisellä tasolla tarkkaa. Tasolla B1 oleva kuvain on esimerkki siitä, kuinka negatiivissävytteinen virhe-sana korvattiin tutkimuksen asteikossa : ...,

*vaikkakin suurempia virheitä esiintyy kompleksisimpia ajatuksia ilmaistaessa... korvattiin muotoilulla ...vaikkakin sekaannuksia voi esiintyä kompleksisimpia ajatuksia ilmaistaessa.*

#### *Fonologinen hallinta*

Muutoksia on tehty tasojen C2, B1, A2 ja A1 kuvaimiin. On huomattava, että EVK:n luonnos 2:ssa on ääntämistä koskevan asteikon yhteydessä huomautus, jonka mukaan kolme alimman tason kuvausta on väliaikaisesti sijoitettu luonnos 2:n kyseiseen versioon. Huomautuksessa todetaan lisäksi, että vuoden 1994 tutkimuksessa alemman tason kuvaimet oli muotoiltu negatiivisesti ja ne hylättiin analyysissä, koska niiden tulkinta oli epä johdonmukaista (EVK luonnos 2 1998, 44).

Viitekehityksen luonnoksessa 2, kuten myös vuoden 2001/2004 EVK:ssa tasojen C1 ja C2 kuvain on sama. Olen kuitenkin tutkimuksessa halunnut tehdä eron näiden tasojen kuvainten välillä lisäämällä taitavasti-sanana asteikon ääntämistä koskevaan kohtaan tasolle C2 (*Osoa taitavasti vaihdella intonaatiota ja asetta lausepainon oikein ilmaistakseen hienompia merkitysvaihteita*).

Tasolla B1 ollut kuvaimen osa, ... ja virheellistä ääntämistä esiintyy silloin tällöin korvattiin tutkimuksessa ilmaisulla *ei-kohdekielenomaista ääntämistä esiintyy silloin tällöin*. Arvottava virhe-sana on siten jälleen muutettu vähemmän tuomitsemaan muotoon tutkimuksen arviointiasteikossa.

A2-tason kuvain on esimerkki liian pitkästä kuvaimesta, joka paloiteltiin kahteen osaan tutkimuksessa siten, että puolipistettä edeltävästä osasta tuli ensimmäinen kuvain ja puolipisteen jälkeisestä osasta toinen kuvain (*Ääntäminen on yleensä niin selkeää, että se on ymmärrettävää huolimatta selvästi havaittavasta vierasperäisestä korostuksesta; keskustelukumppanit joutuvat kuitenkin pyytämään toistoa aika ajoin*). Samassa yhteydessä *keskustelukumppanit*-sana on muutettu *kuulijoiksi* (*Kuulija saattaa joutua kuitenkin pyytämään toistoa aika ajoin*). Toiston pyytäminen toteutuu tehtävässä 1 siten, että puhe suorituksen arvioijat joutuvat palaamaan suorituksessa taaksepäin halutessaan varmistaa jotakin tiettyä ääntämiseen liittyvää kohtaa puhe suorituksessa.

Tehtävää 1 koskevia muutoksia tehtiin eniten ääntämisen osalta tasolle A1. EVK:n Luonnos 2:n kuvain, jossa viitataan syntyperäisiin kielen puhujiin ja heidän suhteeseensa hänen (puhujan) kieliryhmänsä edustajien kanssa, muutettiin suoraan tutkimuksessa toteutuvaan muotoon eli tutkimuksen asteikossa viitataan toisaalta syntyperäisiin ranskan kielen puhujiin ja toisaalta ranskan kieltä käyttäviin suomalaisiin. Lisäksi tutkimuksen asteikosta poistettiin kokonaan alkuperäisessä kuvaimessa ollut viittaus hyvin rajallisiin sanoihin ja ilmauksiin. Sen sijaan tasolle A1 jätettiin alkuperäisten kuvainten viittaukset siihen, että tällä taitotasolla olevan puhujan sanasto ja ilmaukset ovat suppeita ja edeltäkäs opittuja sekä siihen, että kuulijalta vaaditaan jonkinlaista ponnistelua, jotta hän ymmärtäisi puhujaa. Kaikkien näiden viittausten muotoilu muutettiin kuitenkin vähemmän negatiiviseksi. Lisäksi tutkimuksen asteikkoon on lisätty A1-tason puhujalle tyypillinen ilmausten etsiminen.

Lopuksi esitetään muutama kommentti koskien luonnoksessa 2 mainittuja kolmen alimman tason ääntämisen kuvaimia, joihin viitattiin edellä. Kun verrataan luonnoksessa 2 olevien kolmen alimman tason eli tasojen A1, A2 ja B1 kuvaimia ja vuoden 2001/2004 Viitekehyksessä olevia vastaavia kuvaimia, voidaan todeta, ettei viimeksi mainitussa versiossa ole mielestäni kovinkaan paljon muutettu kuvaimia, joiden piti luonnos 2:n mukaan olla väliaikaisia niiden negatiivisten muotoilujen vuoksi. Kuvaimet ovat molemmissa Viitekehyksen versioissa hyvin samantyyllisiä; erona tuoreempaan kehykseen näyttää olevan se, että ääntämistä koskien on tehty joitakin pieniä ”lievennyksiä”. Lievennykset toteutuvat usein joidenkin sanojen poisjättämisenä kuten esimerkiksi luonnos 2:ssa mainittu viittaus *hyvin rajallisiin sanoihin ja ilmauksiin*, jota ei enää lainkaan mainita vuoden 2001/2004 Viitekehyksen ääntämistä koskevassa asteikossa. Samaten A2-tason kuvainta on muutettu lievemmäksi jättämällä pois ilmaus *selvästi havaittava* vierasperäiseen korostukseen viitattaessa. Muuten kolmen alimman tason ääntämistä koskevat kuvaimet ovat hyvin samanlaisia molemmissa Viitekehyksen versioissa.

*Pitkä monologi: kokemuksen kuvaaminen*

Monologia koskeva asteikko sijoittuu EVK:n luonnos 2:ssa Tuottaminen/produktio-osuuden alle (Pitkä monologi: kokemuksen kuvaaminen) ja vuoden 2001 / 2004 EVK:ssa Kielen tuottamistoiminnot ja -strategiat -osuuden alle (Puhuminen: kokemuksen kuvaaminen). Tutkimuksen asteikkoon otettiin kokemuksen kuvaaminen -asteikosta elementtejä vain tasoille C2 ja B1. Muilla tasoilla ei ole viittauksia tähän asteikkoon, koska erityisesti asteikon alatasoilla kuvaimet keskittyvät puhujan itsensä, hänen perheensä ja elinolosuhteidensa tai muuhun puhujan välittömän jokapäiväisen elinympäristön kuvaamiseen. Tehtävässä 1 ei kuitenkaan millään tavalla pyydetä kuvailemaan puhujan omaan elämään liittyviä asioita, ja tästä syystä asteikosta onkin otettu aineksia mukaan tutkimuksen asteikkoon hyvin vähän.

#### **4.2.9 Tehtävän 2 arviointiasteikon esittely**

Tehtävän 2 arviointiasteikko (liite 8) on koottu Viitekehyksen luonnos 2:ssa olevista kommunikatiiviseen kielikompetenssiin liittyvistä asteikoista. Tehtävän 2 arviointiasteikossa on elementtejä seuraavista EVK:n asteikoista:

- puheen sujuvuus
- sanaston laajuus
- sanaston hallinta
- fonologinen hallinta.

Edellä lueteltujen kriteerien lisäksi tutkimuksen asteikon kaikille tasoille on lisätty seuraavat kriteerit: kohteliaisuusmuotojen hallinta, tilanteen vaatiman muodollisuusasteen hallinta ja tilanteessa reagointi. Näitä tarvitaan tehtävän 2 arvioinnissa. Koska tehtävässä 2 esitellään tärkeä ranskalainen vieras puhujan

kollegalle, on puhe-suorituksen arvioinnissa kiinnitettävä muiden puheen tuottamiseen liittyvien piirteiden lisäksi huomiota siihen, että esittely tapahtuu riittävää kohteliaisuutta ja professionaalisen tilanteen edellyttämää muodollisuutta noudattaen. Tilanteessa reagointi liittyy siihen, kuinka puhuja reagoi ja selviää tehtävän edellyttämästä puhetilanteesta (reagoinnin luontevuus / kömpelyys).

Todettakoon lisäksi, että koska monet tehtävä 2:n asteikon kriteereistä ovat samoja kuin tehtävässä 1 (puheen sujuvuus, sanaston laajuus ja hallinta, fonologinen hallinta), tehtävässä 1 selostetut asteikkojen muotoilu koskevat kommentit ja kriteerit koskevat myös tehtävää 2.

#### *Puheen sujuvuus*

Tasojen C2, C1 ja B1 kuvaimet ovat täysin samat kuin tehtävässä 1. Tasolla B2 olevaa EVK:n luonnos 2:n kuvainta, jossa viitataan syntyperäisiin puhujiin, muutettiin siten, että asteikossa käytetään syntyperäisen puhujan tilalla *toista osapuolta*. Tason A2 alkuperäinen kuvaus, jossa viitataan tuttuja aiheita koskevien lauseiden helppoon rakentelemiseen, lyhyistä keskusteluista selviytymiseen, epäröintiin ja väärin aloituksiin, muutettiin muotoon, jossa korostuu erityisesti se, että vaikka puhujan lauseet ovatkin vielä lyhyitä ja epäröiviä, puhuja selviytyy kohtalaisesti tehtävästä (*puhuja pystyy rakentelemaan yksinkertaisia lauseita niin, että selviytyy vuorovaikutustilanteesta kohtalaisesti huolimatta selvästä epäröinnistä*).

#### *Sanaston laajuus*

Sanaston laajuutta koskevien kuvainten muotoilu on täysin sama kuin tehtävässä 1 tasoilla C2, C1, B2 ja A1. Muutoksia tämän kriteerin osalta tehtiin tasoille B1 ja A2. Tasolle B1 lisättiin viittaus professionaaliseen kontekstiin. Tason A2 kuvainta muutettiin EVK:n luonnos 2:n kuvaimen nähden siten, että tehtävässä 2 korostettiin sitä, että puhetapahtuma liittyy kommunikaatio-tilanteesta selviytymiseen eikä perustason kommunikaatiotarpeiden tyydyttämiseen (*omaa riittävään sanavaraston tehtävän edellyttämästä kommunikaatiotilanteesta selviytyäkseen. Sanasto on kuitenkin kapeaa*).

#### *Sanaston hallinta*

Viitekehyksen luonnos 2 ei sisällä A1-tasolla sanaston hallintaan viittaavaa kuvainta. Eri tasojen kuvaimet ovat hyvin pitkälle luonnos 2:n tai Viitekehyksen (2004) mukaisia. Tasolta A2 poistettiin kuitenkin viittaukset konkreettisiin ja jokapäiväisiin tarpeisiin liittyvän sanavaraston hallinnasta.

#### *Fonologinen hallinta*

Ääntämiseen liittyvät muutokset on selostettu tehtävän 1 yhteydessä. Tasolle A2 tehtiin lisäys, jossa viitataan kommunikaatiotilanteeseen ja alkuperäisessä asteikossa oleva keskustelukumppanit-sana muutettiin tutkimuksen arvioin-

tiasteikossa yksikköön (*Todellisessa kommunikaatiotilanteessa keskustelukumppani saattaisi kuitenkin joutua pyytämään toistoa aika ajoin*).

*Tilanteessa reagointi sekä kohteliaisuusmuotojen ja tilanteen edellyttämän muodollisuusasteen hallinta*

Tutkimuksen arviointiasteikkojen laatimisaikaan EVK:n luonnos 2:ssa ei ollut sosiolingvistisiä taitoja koskevia kuvaimia, ja tästä syystä jouduin laatimaan ne itse tehtävän 2 arviointia varten. Näiden kriteereiden tarkoituksena on selvittää sitä, oliko tutkimukseen osallistuvan oppijan puheasuoritus sosiolingvistisiltä piirteiltään sellainen, että se voitaisiin hyväksyä tilanteessa, jossa ranskan kielen natiivi ja ranskaa vieraana kielenä käyttävä henkilö ovat vuorovaikutuksessa tehtävä 2:n kaltaisessa todellisessa kommunikaatiotilanteessa. On huomattava, että ei ole sattumaa, että Viitekehyksen ensimmäiset versiot eivät sisältäneet sosiolingvistiseen kompetenssiin liittyviä erillisiä asteikkoja, koska sosiolingvististen kompetenssien elementtien järjestäminen asteikoksi osittautui ongelmalliseksi. Nykyisen EVK:n mukaan tietyt kategoriat menestyivät kuvaimien kehittämisvaiheessa heikoimmin opettajien arvioinneissa, ja ne poistettiin analyysistä. Poistettavien kategorioiden ryhmään kuuluivat myös sosiokulttuurisia kykyjä ja taitoja ilmaisevat kategoriat, joiden kuvaimet käsittelivät juuri sosiokulttuurista ja sosiolingvististä kykyä ja valmiutta. Tässä yhteydessä mainitaan myös, että monia elementtejä jätettiin pois nykyisestä *sosiolingvistinen tarkoituksenmukaisuus* -asteikosta ja ainoastaan edellä mainitusta asteikosta löytyvät elementit onnistuttiin sijoittamaan asteikkoon. (EVK 2004, 170-171; 296.) EVK:ssa muistutetaan myös, että sosiolingvististen taitojen todellinen hallinta alkaa vasta tasolta B2; sitä alemmilla tasoilla kyse on ainoastaan sosiaalisiin suhteisiin ja kohteliaisuussääntöihin liittyvien kielellisten piirteiden hallinnasta (EVK 2004, 171). Tutkimuksen arviointiasteikossa luonteva tilanteessa reagointi ja kohteliaisuusmuotojen ja tilanteen edellyttämän muodollisuusasteen hallinta onkin sijoitettu tasolta B2 ylöspäin.

Yhteenvedona todettakoon vielä, että tutkimuksessa käytettyjen tehtävien 1 ja 2 arviointiasteikon muutoksista valtaosa on kuvaimien lyhentämisiä tai negatiivisten kuvainten muotoilua myönteisempään muotoon.

#### 4.2.10 Yki-korpuksen käsittely

Tutkimuksessa käytetyn Yki-korpuksen suoritusten taitotason selvittäminen oli yksi tärkeimmistä tutkimuksen laatuun liittyvistä tehtävistä. Jotta saataisiin riittävän luotettava kuva Yki-korpuksen suoritusten tasoista, päätettiin arviointityöhön käyttää viittä koulutettua kielitaidon arvioinnin asiantuntijaa. Viidestä arvioijasta kaksi oli ranskan kielen natiiveja (nainen ja mies), ja loput kolme arvioijaa olivat suomalaisia naisia. Tutkimuksen tekijä toimi itse myös yhtenä arvioijana. Kaikki viisi käytettyä arvioijaa on koulutettu tehtäväänsä ja suurimmalla osalla heistä oli kielitaidon arviointiin liittyvää työkokemusta Yki-järjestelmästä. Toinen ranskan kielen natiiviarvioijista koulutettiin kuitenkin erikseen arviointitehtävään. Tähän perehdyttämiskoulutukseen kuului mm.

Euroopan neuvoston Viitekehyksen ideologiaan ja erityisesti sen arviointias-teikkojen käyttöön tutustuminen, Yki-tutkintoon tutustuminen yleisellä tasolla sekä tämän tutkimuksen käyttöön laadittuihin arviointias-teikkoihin perehtyminen. Lisäksi on mainittava, että kaikilla viidellä arvioijalla oli opetuskokemusta ranskan kielen opetuksesta suomalaisessa yliopistokontekstissa. Osalla arvioi-jista oli lisäksi kokemusta ranskan kielen opettamisesta lukiossa.

Kuten aiemmin on jo selostettu, tutkimuksessa käytetyn Yki-korpuksen puhe-suoritukset olivat virallisia Yki-tutkintosuorituksia. Siten ne kaikki oli jo kertaalleen arvioitu tutkintopäivien jälkeen ja niiden taitotasot oli määritelty tuolloin. Valittujen suoritusten taitotasot piti kuitenkin arvioida uudelleen tätä tutkimusta varten siitä syystä, että tutkintojen suorittamisen aikaan Yki-järjestelmässä oli käytössä 9-portainen arviointias-teikko, joka ei vielä tällöin ollut yhteismitallinen Euroopan neuvoston 6-portaisen asteikon kanssa. Tästä syystä tutkimukseen mukaan valitut Yki-suoritukset tuli arvioida uudelleen käyttäen tätä tutkimusta varten luotuja arviointias-teikkoja, jotka pohjautuvat Viitekehyksen 6-portaisiin asteikkoihin.

Jokainen viidestä arvioijasta suoritti arviointityön itsenäisesti. Jokaisella arvioitavalla näytteellä oli oma koodinsa, joten arvioijilla ei ollut käytettävissä minkäänlaisia suorittajia koskevia tietoja. Kaikilla arvioijilla oli arvioitava-na kaksi kasettia (tehtävä 1:n kasetti ja tehtävä 2:n kasetti), joista molemmat sisälsivät 50 puhe-suoritusta. Lisäksi arvioijilla oli paperilla molemmat tehtävät, käytettävät arviointias-teikot ja lomakkeet, joihin oli merkitty näytteiden numerot ja taitotasot (A1–C2). Suoritusten arviointi tapahtui merkitsemällä kunkin suorituksen kohdalle lomakkeeseen rasti sen tason kohdalle, jolle arvioija sijoitti kunkin näytteen.

Yki-korpuksen suorituksia koskevan arviointityön päättymisen jälkeen kaikille arviointityössä mukana olleille arvioijille annettiin kirjallista palautetta heidän arviointiensa tasosta suhteessa neljän muun arvioijan antamaan taito-tasoon samoissa suorituksissa. Palautteenanto tehtiin nimettömänä, ja kaikki arvioijat saivat siten oman koodin, jonka perusteella he saattoivat verrata omia arviointejaan muiden koodillisten arvioijien suorittamiin arviointeihin. Palautteen antamista pidettiin tärkeänä, koska vapaaehtoisten asiantuntevien arvioiji-en työpanos tutkimukselle oli ensiarvoisen tärkeää. Arviointeja koskevaa palau-tetta voidaan yleisemminkin pitää arviointiin liittyvän ammattitaidon jatkuvan ylläpitämisen kannalta välttämättömänä.

Kun palautteessa verrattiin eri arvioijien antamia taitotasoja toisten arvioi-jien vastaaviin tasoihin, tarkoituksena oli tehdä arvioijat tietoisiksi omasta ar-vioijaprofiilistaan eli esimerkiksi siitä, onko jollakulla taipumusta olla pikem-minkin lempeä kuin ankara arvioinneissaan. Tunnettuahan on, että arvioijilla on arviointityötä tehdessään tietynlaisia tendenssejä tai käyttäytymismalleja, jotka ovat yhteydessä heidän taustaansa ja yksilölliseen arviointityyliinsä ja jot-ka yleensä ovat melko muuttumattomia (Knoch, 2010, 181). Tietyn tendenssin toistaminen ei kuitenkaan välttämättä ole ongelma, jos arvioija on tietoinen omasta taipumuksestaan tietynlaiseen arviointilinjaan ja jos hän niin sanotusti pitää linjansa ja pyrkii aina arvioidessaan noudattamaan samansuuntaista ten-denssiä. Sen sijaan epäjohdonmukaisuus arviointityössä tietyn arvioijan koh-



dalla on aina arviointityön luotettavuuden kannalta huono asia. Arvioijille annettussa palautteessa kerrottiin myös, että arvioijat saavuttivat suuremman yksimielisyyden tehtävän 1 kuin tehtävän 2 puhe-suoritusten arvioinneissa.

#### **4.2.11 Aineiston ja menettelytapojen toimivuuskokeilu**

Tutkimukseen osallistuvien oppijoiden käyttöön tarkoitettun kielistudioaineiston toimivuus kokeiltiin etukäteen. Toimivuuskokeilun tarkoituksena oli varmistaa tehtävien, kuunneltavien kasettien, täytettävien lomakkeiden ja annettujen ohjeiden ymmärrettävyys ja toimivuus. Lisäksi huomiota kiinnitettiin vastaamisajan riittävyteen ja kielistudio-osuuden kokonaisajan pituuteen. Kaksi ranskan kielen opiskelijaa pyydettiin kielistudioon, jossa he osallistuivat tutkimukseen samalla tavalla kuin oli tarkoitus tehdä tutkimukseen mukaan ilmoittautuneiden oppijoiden kanssa.

Toimivuuskokeilu koettiin erittäin hyödylliseksi. Sen ansiosta huomattiin muun muassa, että erityisesti kielistudiossa suullisesti annettaviin lisäohjeisiin tarvittiin joitakin tarkennuksia. Kokeilu mahdollisti myös sen, että kielistudioistunto voitiin toteuttaa kaikkien osallistuvien ryhmien osalta noudattaen täsmälleen samoja käytäntöjä, jotka oli jo edeltäkäs in hyviksi havaittu.

#### **4.2.12 Tutkimukseen ilmoittautuminen ja kielistudiot**

Tutkimukseen osallistumisesta kiinnostuneille oppijoille esiteltiin listat, joihin oli merkitty mahdolliset tutkimukseen osallistumisajat. Tämän jälkeen kukin oppija valitsi listasta itselleen parhaiten sopivan ryhmän ja ilmoitti valinnasta tutkimuksen tekijälle sähköpostitse. Sen jälkeen kaikki tiettyyn ryhmään ilmoittautuneet oppijat kokoontuivat kielistudiossa sovittuun aikaan. Oppijoilta hankittavan aineiston kerääminen kesti kuukausia, koska studiorhymiä tuli järjestää monta studiopaikkojen vähäisyyden vuoksi. Lisäksi tutkimuksessa oli osallistujia monista eri oppilaitoksista eri puolilta Suomea, mikä tarkoitti sitä, että riittävän laajan aineiston kerääminen edellytti matkustamista.

Koska kielistudioistunnot järjestettiin yleensä iltaisin, kielistudioiden varaaminen ei aiheuttanut suurempia ongelmia. Sen sijaan studion varaamiseen liittyvät yhteydenotot ja neuvottelut eri yliopistojen vastuuhenkilöiden kanssa veivät aikaa. Jotta kielistudioistunnot varmasti sujuisivat hyvin, edeltä käsin oli selvitettävä, millaiset laitteet missäkin studiossa oli ja miten ne toimivat. Tämän asian selvittäminen vaati yleensä henkilökohtaista tapaamista yliopistojen kielistudioista vastaavien henkilöiden kanssa. Alusta alkaen pidettiin ensiarvoisen tärkeänä, että kielistudiot varmasti toimisivat moitteettomasti silloin, kun oppijat olisi saatu koottua studioon. Tutkimusaineiston kerääminen olisi voinut hidastua huomattavasti, jos tekniset ongelmat olisivat estäneet ryhmien osallistumisen, koska samoja oppijoita ei välttämättä olisi saatu tulemaan paikalle toista kertaa. Teknisen avun saaminen paikalle iltaisin mahdollisessa kielistudiossa ilmenneessä ongelmatilanteessa oli tietenkin myös asia, joka tuli varmistaa etukäteen – onneksi tässä tutkimuksessa ei ilmennyt varsinaisia teknisiä ongelmia huolellisten etukäteisvalmistelujen ansiosta. Todettakoon vielä lopuk-

si, että tämänkaltaisen kielistudioaineiston kerääminen yliopistojen nykyisten kustannustehokkuuteen perustuvien laskentamallien mukaan voisi koitua tämänkaltaisen tutkimuksen tekijälle yllättävän kalliiksi. Tämän tutkimuksen kielistudioaineiston keräämisen aikaan kielistudioiden vuokraaminen oli ilmaista, mitä se yliopistoissa harvemmin enää tänä päivänä on.

### 4.3 Tutkimusaineiston keruu

#### 4.3.1 Osallistujien rekrytointi ja tietoa osallistujista

Tutkimuksen osallistujia koskevat tiedot kerättiin osallistujan taustatiedot -lomakkeen avulla kielistudioistuntojen alussa. Valtaosa tutkimukseen osallistuneista oppijoista (78,2 %) edustaa neljää suomalaista yliopistoa. Lisäksi mukana tutkimuksessa oli oppilaita kahdesta lukiosta (21,4 %). Tarkemmat tiedot osallistuneiden oppilaitoksista ja lukumääristä on koottu taulukkoon 4.

TAULUKKO 4 Tutkimuksen osallistujien oppilaitokset, osallistujamäärät ja prosenttiosuudet / oppilaitos.

Oppilaitos	Osallistujamäärä	Osallistuneiden %-osuus
Jyväskylän yliopisto	81	38,6
Helsingin yliopisto	44	21,0
Tampereen yliopisto	8	3,8
Lappeenrannan teknillinen yliopisto	31	14,8
Lyseon lukio, Jyväskylä	34	16,2
Normaalikoulun lukio, Jyväskylä	11	5,2
Yhteensä	209	99,5
Puuttuu	1	,5
Yhteensä	210	100,0

Suurin osa tutkimukseen osallistujista oli yliopisto-opiskelijoita. He muodostavat varsin heterogeenisen ryhmän, sillä heidän opiskeluprofiilinsa olivat hyvin erilaisia. Jotkut osallistujista olivat ranskan kielen opiskelijoita yliopistojen ranskan kielten laitoksilla (ranska pää- tai sivuaineena), kun taas toiset olivat muiden aineiden kuin ranskan kielen opiskelijoita (esim. Lappeenrannan teknillisen yliopiston 31 opiskelijaa), jotka olivat opiskelleet ranskaa tai opiskelivat sitä parhaillaan esimerkiksi kielikeskusten järjestämällä vieraan kielen kursseilla.

Tutkimukseen haettiin osallistujia monen eri kanavan avulla. Tutkimuksesta kertovia tiedotteita jaettiin eri puolille yliopistoa. Ranskan kielen opettajille ja heidän opiskelijoilleen esiteltiin tutkimusta ja sen tavoitteita. Opettajille

tarjottiin heidän niin halutessaan mahdollisuus siihen, että heidän tunneillaan olevat ranskan opiskelijat pystyivät osallistumaan tutkimukseen ranskan tunti-en aikana. Opiskelijoihin otettiin suoraan yhteyttä myös käyttämällä yliopistojen sähköpostilistoja. Joissakin tapauksissa tutkimuksen tekijä meni paikalle etukäteen sovittuun tilaisuuteen kertomaan tutkimukseen osallistumismahdollisuudesta (mm. ranskan kielen opiskelijoiden yhdistysten tilaisuudet). Suurin osa tutkimukseen osallistujista rekrytoitiin kuitenkin eri puolilla yliopistoa esillä olleiden tiedotteiden avulla.

Tutkimukseen osallistujien valinta oli sattumanvaraista siinä mielessä, että pääsääntöisesti jokainen ranskan kielen oppija, joka oli kiinnostunut saamaan arvion omasta ranskan kielen suullisesta taidostaan, sai ilmoittautua tutkimukseen mukaan. Oppijoiden tuli kuitenkin täyttää kaksi kriteeriä: heidän tuli olla suomalaisia (äidinkieleltään suomen- tai ruotsinkielisiä) ja aikuisiässä olevia. Tässä tutkimuksessa mukana olevat kahden lukioluokan oppilaat lasketaan siis kuuluvaksi aikuisväestöön.

Osallistujien taustatiedot -lomakkeella kerättiin yhteystietojen lisäksi seuraavia tietoja tutkimukseen osallistujilta:

- ikä
- sukupuoli
- äidinkieli
- aiempi itsearviointikokemus.
- aiemmat ranskan kielen opinnot
- ranskan kielen opintojen aloittamisikä
- kontaktit ranskan kieleen tutkimushetkellä
- ranskan kielen käytön frekvenssi

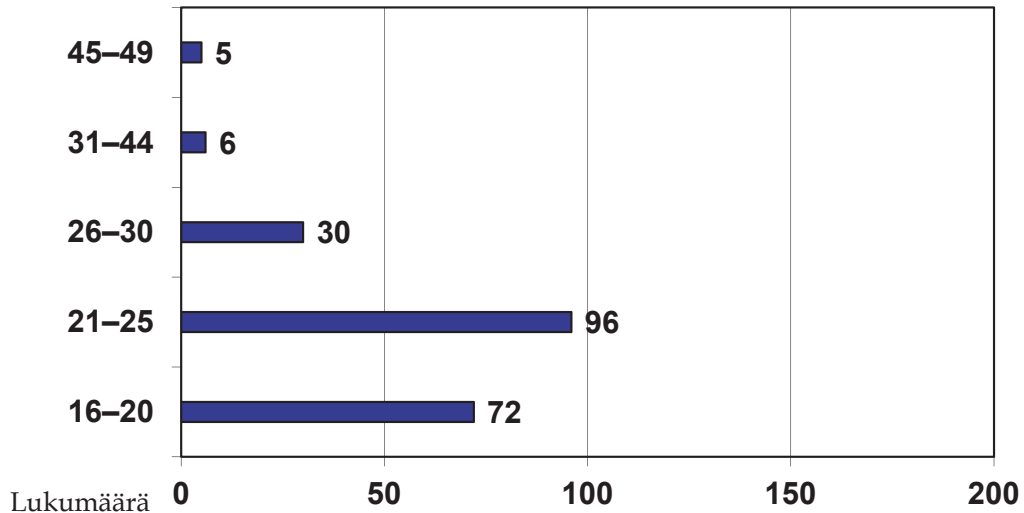
Kutakin kohtaa käsitellään seuraavassa kolmessa eri kohdassa.

#### **4.3.1.1 Osallistujien perustiedot**

##### *Ikä*

Tutkimukseen osallistujat jaettiin ikänsä perusteella viiteen ikäryhmään (kuvio 3). Kuten kuviosta näkyy, osallistujien suurimman ikäryhmän muodostivat 21–25-vuotiaat opiskelijat (N = 96). Seuraavaksi suurin osallistujaryhmä edusti edellistä nuorempia oppijoita, eli tähän ryhmään sijoittuivat kaikki kahdelta lukioluokalta tutkimuksessa mukana olleet opiskelijat, joiden ikäjakauma oli 16 ja 20 vuoden välillä (N = 72). Kolmanneksi suurin ikäryhmä (N = 30) edusti oppijoita, joiden ikäjakauma oli 26 ja 30 vuoden välillä. Yli 30 vuoden ikäisiä osallistujia oli tutkimuksessa minimaalisen vähän (N = 11).

## Ikäryhmät



KUVIO 3 Tutkimukseen osallistuneiden oppijoiden ikäjakauma.

*Sukupuoli*

Tutkimuksen 209 osallistujasta oli naisia 174 ja miehiä 35. Miesten vähäinen lukumäärä heijastelee hyvin sitä tosiasiaa, että vieraiden kielten opiskelijoiden enemmistö lukion jälkeisissä kieliopinnoissa Suomessa koostuu naisista. Tämän tutkimuksen osallistujien sukupuolijakauma on siis odotuksenmukainen.

*Äidinkieli*

Kuten aiemmin jo mainittiin, tässä tutkimuksessa tarkasteltiin äidinkieleltään joko suomen- tai ruotsinkielisten suomalaisten itsearviointia. Rajaus äidinkielen mukaan tehtiin siitä lähtökohdasta, että oli oletettavaa, että suomalainen koulu- ja koulutuslaitos edustaa yhtenäistä oppimiskulttuuria ja että tällä seikalla olisi ainakin jossain määrin yhtenäistävä tekijä tutkimustulosten yleistettävyyden kannalta. Tutkimukseen osallistuneista vain muutama oli äidinkieleltään ruotsinkielinen, joten suomenkieliset ovat tässä tutkimuksessa valtaenemmistö.

On huomattavaa, että tutkimusaineiston keräämisen aikaan 2000-luvun alussa Suomessa oli yleiselläkin tasolla vielä melko vähän maahanmuuttajia ja siten heidän osuutensa yliopistoissa ja lukioissa oli verraten pieni. Tästä syystä vain äidinkieleltään suomen- ja ruotsinkielisten oppijoiden osallistuminen tutkimukseen oli jokseenkin automaattista. Huomattakoon kuitenkin, että toisessa kahdesta lukiolaisryhmästä oli maahanmuuttajaoppilas, jonka äidinkieli ei ollut kumpikaan Suomen virallisista kielistä. Kyseistä oppilasta kohdeltiin kuitenkin tutkimukseen osallistumisen osalta täysin samalla tavalla kuin hänen luokkatovereitaan, eli oppilas suoritti molemmat tehtävät kielistudiossa ja täytti tutki-

mukseen käytettävät lomakkeet, minkä jälkeen hän sai suorituksistaan palautetta siinä kuin luokkatoverinsakin. Tämän oppilaan suorituksia ei kuitenkaan sisällytetty tutkimusaineistoon.

#### *Aiempi itsearviointikokemus*

Aiempaa itsearviointikokemusta kysyttiin taustatietolomakkeessa ”kyllä”- tai ”ei”-vastausten avulla. Jos osallistuja rastitti ”kyllä”-vaihtoehdon, hänen tuli myös kertoa, mihin tehty itsearviointi oli liittynyt (esimerkiksi jonkun kielen arviointiin tai omiin työskentelytapoihin).

Enemmistöllä tutkimukseen osallistujista (58,1 %) ei ollut aiempaa itsearviointikokemusta. Sen sijaan itsearviointi oli tuttua 41,6 % :lle tutkimukseen osallistujista.

Suurin osa itsearviointiaktiiviteeteista oli tehty eri aineiden tunneilla (43 vastausta). Myös omiin työskentelytapoihin ja -taitoihin liittyvä itsearviointi oli suhteellisen yleistä (29 vastausta). Suoraan kieliin liittyvää itsearviointia oli harjoittanut vain 17 tutkimuksen osallistujaa. Varsinaisiin oppiaineisiin liittyvän itsearvioinnin lisäksi osallistujat olivat harjoittaneet itsearviointia seuraavissa yhteyksissä: henkilökohtaisten ominaisuuksien, ryhmätyö- ja kommunikaatio- taitojen, motivaation, auton ajotaidon, fyysisen kunnon ja ajanhallinnan arviointi.

Osallistujien tekemien itsearviointien voidaan täten todeta edustavan hyvin monia erilaisia itsearviointiaktiiviteetteja. Sen sijaan vieraiden kielten taidon itsearviointiin liittyvää kokemusta tutkimukseen osallistuneilla oli melko vähän.

#### **4.3.1.2 Ranskan kielen opinnot**

Tiedot tutkimukseen osallistuneiden aiemmista ranskan kielen opinnoista kerättiin antamalla kouluasteet (peruskoulu, lukio, yliopisto, muu) ja vuosien määrä (1–6 tai enemmän) sekä pyytämällä osallistujia ympyröimään, kuinka monta vuotta kullakin kouluasteella heillä oli ranskan kielen opintoja. Jos osallistuja valitsi ”muu”-vaihtoehdon luetelluista kouluasteista, häntä pyydettiin tarkentamaan, missä edeltäviä opintoja oli ollut. Vähän yli puolet (51,9 %) tutkimukseen osallistujista oli opiskellut ranskan kieltä kaksi vuotta peruskoulussa. Toisaalta on huomattava, että niiden osallistujien määrä, jotka eivät olleet lainkaan opiskelleet ranskan kieltä peruskoulussa, oli melko suuri (38,6 %). Noin 10 prosentilla oppijoista oli ranskan kielen opintoja peruskoulussa 3 vuotta tai enemmän.

#### *Opinnot lukiossa*

Hieman yli puolet (53,8 %) osallistujista oli opiskellut ranskaa lukiossa kolme vuotta. Lähes 20 % (19,5 %) osallistujista ei kuitenkaan ollut opiskellut ranskan kieltä lainkaan lukiossa. Tutkimukseen osallistuneista liki 15 % (14,3 %) ilmoitti opiskelleensa ranskan kieltä lukiossa vain yhden vuoden ja 11,9 % kahden vuoden ajan.

### Opinnot yliopistossa

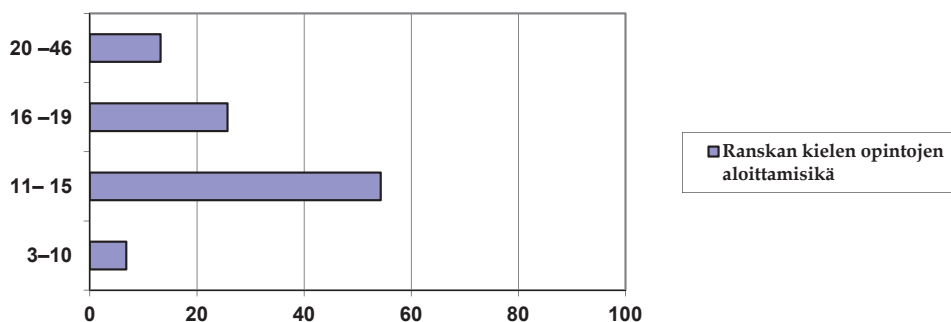
Tutkimukseen osallistuneista melko harva oli opiskellut ranskaa yliopistossa. 31,9 % ei ollut lainkaan opiskellut ranskan kieltä yliopistossa, ja yhtä moni (31,9 %) ilmoitti opiskelleensa kyseistä kieltä vain yhden vuoden yliopistossa. Kaksi vuotta (13,8 %) ja kolme vuotta (10 %) kieltä opiskelleiden määrä oli yhteensä noin 24 %. Ranskan kieltä pisimpään yliopistossa opiskelleiden prosenttiosuudet olivat seuraavat: 5 vuotta 2,9 %, 6 vuotta 2,4 % ja 7 vuotta 1,4 % eli yhteensä pitkään ranskan kieltä yliopistossa opiskelleiden osuus oli 6,7 % kaikista tutkimukseen osallistuneista.

### Opinnot muualla

Yli 70 %: lla (73,8 %) tutkimukseen osallistuneista ei ollut lainkaan ranskan kielen opintoja muualla kuin peruskoulussa, lukiossa tai yliopistossa. Toinen huomionarvoinen ryhmä on vuoden muualla ranskan kieltä opiskelleet (16,2 %). Tarkasteltaessa missä muualla opiskelleet ovat ranskan kieltä opiskelleet tai oppineet kärkipaikalle nousevat ulkomailla suoritetut opinnot (29 vastausta). Suurin osa näistä henkilöistä (19 henkilöä) oli ollut vuoden opiskelijavaihdossa Ranskassa. 7 tutkimukseen osallistujaa oli ollut suorittamassa opintoja myös muissa ranskankielisissä maissa tai provinseissa kuten Belgiassa, Sveitsissä tai Quebecissä. 3 osallistujaa oli suorittanut ranskan kielen opintoja ulkomaisessa yliopistossa, ja samaten 3 osallistujaa oli oppinut ranskan kieltä oleskelemalla jossain ranskankielisessä maassa suorittamatta kuitenkaan kyseisen kielen opintoja. Muina opiskelupaikkoina osallistujat mainitsivat työväenopiston, avoimen yliopiston, Ranskan kulttuurikeskuksen, kauppa- ja sihteeriopiston sekä aikuiskoulutuskeskuksen.

### Ranskan kielen opintojen aloittamisikä

Osallistujat on jaettu kuviossa 4 neljään ikäryhmään sen mukaan, minkäikäisinä he ovat aloittaneet ranskan kielen opiskelun.



KUVIO 4 Tutkimukseen osallistuneiden oppijoiden ranskan kielen opintojen aloittamisikä.

Näyttäisi siltä, että valtaosa (54,3 %) suurimpaan ryhmään kuuluvista (aloi- tusikä 11–15 vuotta) osallistujista on aloittanut ranskan kielen opintonsa valin- naisena B2-kielenä 7. ja 9. luokan välillä tai valinnaisena A2-kielenä 5. luokalla. Toiseksi suurimman ryhmän muodostavat ne osallistujat, jotka ovat aloittaneet ranskan kielen opintonsa 16–19 vuoden iässä (25,7 %). Tämä ryhmä edustaa siis niitä oppijoita, jotka aloittavat lukiossa ranskan kielen valinnaisena B3-kielenä. Kaksi muuta ikäryhmää edustaa huomattavasti pienempiä ryhmiä: ensimmäi- nen edustaa oppijoita, jotka ovat aloittaneet opintonsa hyvin varhain eli 3. ja 10. ikävuoden välillä (2,4 %), ja toinen aikuisena ranskan opintonsa aloittaneita (20–46-vuotiaat, 13,2 %).

#### 4.3.1.3 Ranskan kielen käyttö

##### *Kontaktit ranskan kieleen tutkimushetkellä*

Osallistujan taustatiedot -lomakkeessa annettiin viisi eri vaihtoehtoa, joista osal- listujan tuli rastittaa ne toiminnot, joilla hän oli tutkimukseen osallistumisaika- na tekemisissä ranskan kielen kanssa. Lomakkeen kysymys oli ”kuinka olet ny- kyisin tekemisissä ranskan kielen kanssa?” Valittavat vastausvaihtoehdot sel- viävät taulukosta 5:

TAULUKKO 5 Tutkimuksen osallistujien kontaktit ranskan kieleen tutkimushet- kellä.

	%
En mitenkään	17,6
Puhun ranskaa	48,1
Kuuntelen ranskaa (esim. televisiosta)	71,4
Kirjoitan ranskaksi (esim. kirjeitä)	46,2
Luen ranskaksi	53,8

Ymmärtämistaitojen, lukemisen ja kuuntelemisen, suuret prosenttiosuudet oli- vat odotettavia, kun tiedossa oli, että monet tutkimukseen osallistuneista olivat ranskan kielen opinnoissaan melko alkuvaiheessa. Suurin osa osallistujista ker- toi kuuntelevansa ranskaa (71,4 %). Lukeminen ranskaksi oli seuraavaksi suosi- tuin aktiviteetti (53,8 %). Produktiivisten taitojen, kirjoittamisen ja puhumisen, prosenttiosuudet olivat keskenään hyvin samansuuntaisia ja verraten hyviä, kun otetaan huomioon, että vieraan kielen tuottamista yleisesti pidetään suh- teellisen vaativana: 46,2 % oppijoista kertoi kirjoittavansa ja 48,1 % puhuvansa ranskaa. Todettakoon vielä, että tähän tutkimukseen osallistuneet oppijat vai- kuttavat selvästikin olleen kiinnostuneita ranskan kielestä, sillä yli 80 % (81,9 %) heistä oli käyttänyt jollakin tavalla ranskan kieltä tutkimukseen osallistumisen aikoihin.

*Ranskan kielen käytön frekvenssi*

Kun osallistujilta kysyttiin, kuinka usein he käyttivät ranskan kieltä esimerkiksi työssään, opinnoissaan tai vapaa-aikanaan, heidän tuli valita yksi taulukossa 6 näkyvästä neljästä vaihtoehdosta:

TAULUKKO 6 Tutkimuksen osallistujien ranskan kielen käytön frekvenssi tutkimushetkellä.

	Määrä	%-osuus
(Melkein) päivittäin	46	21,9
Muutaman kerran viikossa	113	53,8
Kerran kahdessa viikossa	18	8,6
Kerran kuussa tai harvemmin	32	15,2
Yhteensä	209	99,5
Puuttuu	1	,5

Yli puolet tutkimukseen osallistuneista oppijoista (53,8 %) kertoi käyttävänsä ranskaa muutaman kerran viikossa. Kun edellä mainittu prosenttiosuus (53,8 %) lisätään melkein päivittäin ranskaa käyttäneiden prosenttiosuuteen (21,9 %), voidaan todeta, että tähän tutkimukseen osallistuneista noin 75 % käytti ranskan kieltä muodossa tai toisessa tutkimukseen osallistumisen aikoihin säännöllisesti joko päivittäin tai ainakin muutaman kerran viikossa. Näinkin huomattava prosenttiosuus kertoo siitä, että tähän tutkimukseen osallistuneet oppijat olivat selvästikin kukin omalla tavallaan hyvin motivoituneita ranskan kielen käyttämisen ja oppimisen suhteen.

#### 4.3.2 Työn tarkoituksen esitleminen ja toimintaohjeiden antaminen

Ennen jokaisen ryhmän saapumista tarkistettiin kielistudion tekninen toimivuus. Aineiston keruuprosessi oli kaikille tutkimukseen osallistujille samanlainen, mikä tarkoitti mm. sitä, että tutkimuksessa käytetyt lomakkeet jaettiin kaikille täsmälleen samassa järjestyksessä.

Tutkimuksen tekijä oli itse paikalla jokaisen kielistudioryhmän saapuessa studioon ja johti kielistudio-osuutta koko studiossa oloajan. Tällä tavalla voitiin varmistaa käytänteiden ja ohjeiden yhdenmukaisuus eri ryhmien välillä.

Kielistudio-osuus aloitettiin tutkimuksen tavoitteiden lyhyellä esittelyllä, minkä jälkeen selostettiin tehtävien suoritusjärjestys. Osallistujille kerrottiin, että tutkimuksen päätarkoituksena oli tarkastella sitä, miten heidän ranskan kielen puhumisen itsearviointinsa sujuu. Selostuksessa mainittiin myös tutkimuksessa käytössä olleet kaksi erityyppistä itsearviointimenetelmää, eli pape-



rilla tehtävä itsearviointi ja kuunneltaviin puhenäytteisiin perustuva itsearviointi. Tässä yhteydessä mainittiin myös, että lomakkeissa pyydetyt kommentit (mm. itsearviointiväittämien toimivuudesta sekä niistä piirteistä, joiden pohjalta osallistuja valitsi, suoriutuuko hän omasta mielestään paremmin vai huonommin tehtävästä kuin kuunneltavan näytteen puhuja) olivat tutkimuksen kannalta tärkeitä, mutta eivät pakollisia. Studio-osuuden ohjeistuksen lopuksi muistutettiin vielä, että tutkimukseen osallistuneiden kaikkia tietoja ja suorituksia käsiteltäisiin luottamuksellisesti ja nimettöminä tutkimustarkoituksiin.

Kukin kielistudiossa paikalla ollut ryhmä eteni samanaikaisesti tutkimukseen kuuluvien tehtävien suorittamisessa, joten kielistudio-osuuteen ei olisi voinut ottaa mukaan mahdollisia myöhässä saapuneita oppijoita. Oppijat perehtyivät yksityiskohtaisesti jokaiseen tutkimuksessa käytettyyn lomakkeeseen tutkimuksen tekijän johdolla ennen lomakkeen käyttöönottoa. Studiossa annettujen suullisten ohjeiden lisäksi samat ohjeet löytyivät oppijoiden edessä olevista lomakkeista kirjallisina. Oppijoita kannustettiin kysymään, jos jokin kohta tehtävässä ei tuntunut selvältä. Yksityiskohtaisen lomakkeiden läpikäynnin tarkoituksena oli se, että kielistudiotilanteeseen usein liittyvää jännitystä voitaisiin ehkäistä ja että jokainen oppija varmasti ymmärtäisi, mitä hänen odotettiin tekävän kielistudiossa. Tällä seikalla oli ensiarvoisen tärkeä merkitys, jotta kaikki lomakkeet voitiin kerätä tutkimuksen tarpeisiin oikein ja täydellisesti täytettyinä. Tutkimuksen tekijän vastuulla oli kielistudion tekninen yhteiskäyttö, joten tutkimukseen osallistujien ei tarvinnut keskittyä esimerkiksi oman suorituksensa nauhoittamiseen.

#### **4.3.2.1 Osallistujan taustatiedot -lomakkeen täyttäminen**

Kielistudio-osuus aloitettiin aina taustatietolomakkeen täyttämällä, mikä kesti noin viisi minuuttia. Kysymyksiä tästä lomakkeesta tuli ainoastaan ranskan kielen opintoihin liittyen.

#### **4.3.2.2 Puhumisen itsearviointi -lomakkeen täyttäminen**

Kyseessä oli paperilla täytettävä 3-sivuinen itsearviointilomake, jonka ensimmäisellä sivulla osallistuja arvioi omaa ranskan kielen puhumisen taitoaan merkitsemällä rastilla sen taitotason (A1–C2), joka parhaiten vastasi hänen omaa puhumisen taitoaan ranskan kielessä. Tämän jälkeen osallistuja vastasi joko ”kyllä” tai ”ei” 42: een lomakkeen sisältämään yksityiskohtaiseen puhumisen itsearviointiväittämään. Lomakkeen täyttäminen kesti noin 15 minuuttia ja herätti jonkin verran kysymyksiä ja kommentteja osallistujien täyttäessä sitä. Lomakkeeseen ja sen toimimiseen sekä väittämien sisältöön liittyviä oppijoiden kommentteja käsitellään tarkemmin jäljempänä työssä.

### 4.3.2.3 Tehtävän 1 suorittaminen ja suorituksen nauhoittaminen

Osallistujille jaettiin aluksi tyhjä C-kasetti, johon jokaista pyydettiin merkitsemään nimensä ja studiopaikkansa numero. Tässä vaiheessa varmistettiin vielä kerran kielistudion tekniikan toimivuus, jotta varmasti saataisiin kaikkien paikalla olleiden osallistujien puhe-suoritus nauhoitettua asianmukaisesti. Jokainen osallistuja sanoi nauhalle nimensä, minkä jälkeen he kuuntelivat, että oma nimi oli varmasti tallentunut kasetille.

Kun oli varmistettu, että kaikkien nimet kuuluivat nauhoilta, jaettiin tehtävä 1 paperilla. Puhumistehtävä esiteltiin ja osallistujille annettiin 1,5 minuuttia valmistautumisaikaa ennen puhumistehtävän suorittamista ja suorituksen nauhoittamista. Valmistautumisaikana oppijoiden oli mahdollista halutessaan kirjoittaa paperille sanoja ja / tai ilmauksia. Tehtävä 1:n suoritusaika oli 1,5 minuuttia, minkä aikana oppijat kuvailivat tehtävän 1 piirroksessa olevaa mieshenkilöä. Tehtäväpaperiin oli kuvailun helpottamiseksi kirjoitettu kahdeksan apukysymystä (esim. minkä tyyppisestä ihmisestä on kyse), joiden tarkoituksena oli antaa vinkkejä, mistä osallistuja esimerkiksi voisi puhua tehtävää suorittaessaan. Ennen tehtävän suorittamista korostettiin kuitenkin, että apukysymykset olivat paperilla vain sitä varten, että niiden pohjalta olisi helpompi tuottaa tekstiä ja että niitä ei ollut pakko käsitellä puhumistuotoksen aikana. Oppijoita kehoitettiin lisäksi käyttämään mielikuvitustaan tehtävän suorittamisen aikana. Puhe-suorituksen tyyli jätettiin tarkoituksellisesti vapaaksi eikä oppijoiden suorituksilta vaadittu siten mitään tietynlaista tyyliä. Heille kerrottiin, että tarkoituksena oli puhua kuvan pohjalta mahdollisimman paljon yhtäjaksoisesti suoritusaajan puitteissa, mutta että toisaalta suorituksen pituutta ei ollut syytä keinotekoisesti pitkittää.

### 4.3.2.4 Oman suorituksen kuuntelu

Suoritettuaan tehtävän 1 jokainen osallistuja kuunteli oman suorituksensa kasetilta yhden kerran. Ennen oman suorituksen kuuntelua osallistujille selvitettiin, että oma suoritus kuunneltiin erikseen heti tehtävän suorittamisen jälkeen siksi, että pystyttäisiin keskittymään suorituksen analysointiin rauhassa täysin erillisenä toimintona itse puhetehtävän suorittamisesta. Oman suorituksen kuunteleminen oli välttämätöntä myös, jotta voitiin edetä tutkimukseen liittyvään toiseen itsearviointiosioon, jossa omaa suoritusta verrattiin nauhalla oleviin puhenäytteisiin yksi kerrallaan.

### 4.3.2.5 Osallistujan vastauslomakkeen täyttäminen (kuunnellut näytteet, tehtävät 1 ja 2)

Tehtävää 1 koskevan oman suorituksen kuuntelemisen jälkeen vuorossa oli tutkimuksen toisen itsearviointitavan suorittaminen eli oman puhe-suorituksen vertaaminen tehtävän 1 kuunneltaviin puhenäytteisiin. Osallistujat kuuntelivat siten 26 puhenäytettä ja vertasivat, selviytyivätkö he tehtävästä omasta mieles-

tään paremmin kuin nauhan näytteen puhuja. Tämä itsearviointiosuus kesti 32 minuuttia, ja se oli kielistudiosession pisin osuus.

Kun tehtävää 1 koskevat näytteet oli kuunneltu, tehtiin samat toimenpiteet täsmälleen samassa järjestyksessä tehtävälle 2 eli ensin valmistauduttiin tehtävän suorittamiseen (tehtävä 2:n valmistautumisaika oli 40 sekuntia), sen jälkeen nauhoitettiin tehtävä 2 (suoritus-aika 30 sekuntia), sitten kuunneltiin oma suoritus ja lopuksi tehtiin tehtävä 2:n näytteitä koskeva itsearviointi vertaamalla omaa suoritusta nauhalta kuunneltaviin 23 puhenäytteeseen. Tehtävä 2 on lyhyt tilanteessa reagoitetehtävä, ja siten koko tehtävää 2 koskevan prosessin suorittaminen kesti yhteensä vain 15 minuuttia. Kielistudio-osuuden lopuksi kaikki käytössä olleet lomakkeet ja kasetit kerättiin osallistujilta pois.

## 4.4 Tutkimusaineiston käsittely

### 4.4.1 Maamerkkikasetin laatiminen

Kielistudiossa kerättyjen osallistujien puhe-suoritusten arviointia varten koottiin molemmista tehtävistä maamerkkikasetti. Maamerkkikasettien käyttämistä suositellaan mm. niissä tilanteissa, joissa on tarpeen verrata kielitaidon tasoja toisiinsa. Viitekehysten mukaan maamerkkien tehtävä on tukea sitä, että asetteikkojen käyttäjät rakentavat yhteneväistä tulkintaa keskustellen koesuoritus-esimerkeistä (tässä tapauksessa puhe-suorituksista). Koesuoritus-esimerkkejä verrataan standardisoi-tuihin taitotasomäärittelyihin ja esimerkkeihin (tässä tapauksessa puhumisen maamerkkeihin) (EVK 2004, 249). Maamerkkikasetin näytesuorituksiksi valittiin näytteitä, joiden tasosta vallitsi yksimielisyys viiden arvioijan välillä Yki-korpuksen suoritusten arvioinnin aikana. 19 minuutin pituisen kasetin tarkoituksena oli helpottaa ja yhdenmukaistaa oppijoiden puhe-suoritusten arviointityötä. Kasetille kerättiin tyypillisiä näytteitä jokaiselta taito-tasolta (A1–C2) tehtävästä 1 ja tehtävästä 2; näytteitä oli kultakin tasolta kaksi lukuun ottamatta tasoja C1 ja C2, joista oli vain yksi näyte kustakin. Kaikilla tasoilla ei kuitenkaan ollut täydellistä yksimielisyyttä viiden arvioijan välillä, ja näissä tapauksissa maamerkeiksi valittiin näytteet, joiden tasosta vallitsi mahdollisimman suuri yksimielisyys arvioijien kesken (esim. neljä arvioijaa viidestä oli samaa mieltä puhe-suorituksen tasosta).

Maamerkkikasetti oli tarkoitettu arvioijien kuunneltavaksi ennen tutkimukseen osallistuneiden oppijoiden puhe-suoritusten arvioinnin aloittamista, jotta tietylle taitotasolle tyypilliset piirteet voitiin palauttaa mieliin.

### 4.4.2 Ohjeistuksen antaminen arvioijille

Yhdenmukaisen arvioinnin varmistamiseksi arvioijille annettiin myös kirjallisia ohjeita ennen oppijoiden puhe-suoritusten arviointityön aloittamista. Ohjeissa

otettiin huomioon ne kommentit ja huomautukset, joita oli saatu viideltä arvioijalta Yki-korpuksen suoritusten arviointityön aikana. Arvioijille jaettiin maa-merkkikasetit, kirjalliset ohjeet sekä arvioinnissa tarvittavat arviointiasteikot ja lomakkeet suoritusten arviointia varten.

Yki-korpuksen suoritusten arvioinnin jälkeen havaittiin, että arvioijien välisessä yksimielisyydessä oli suuria eroja tehtävän 1 ja tehtävän 2 välillä siten, että tehtävä 2:n arviointi koettiin haasteelliseksi mm. arvioitavan tuotoksen lyhyiden vuoksi. Kirjalliset ohjeet koskivat siten pääasiassa tehtävää 2.

Ensimmäinen ohje arvioijille annettiin niitä tuotoksia koskien, joissa puhe-suoritus katkeaa yllättäen. Kyseessä saattoi olla esimerkiksi tapaus, jossa oppija aluksi puhui ja jäi sitten empimään tuottamatta kuitenkaan enää lisää puhetta. Tällaisessa tapauksessa suoritus katkaistiin ajan säästämiseksi siihen kohtaan, missä varsinainen puhe-suoritus loppui. Toinen puhe-suorituksen katkeamiseen liittyvä tilanne saattoi olla sellainen, jossa henkilö jatkoi nauhalle puhumista, vaikka tehtävän suoritus-aika oli jo tullut täyteen. Myös tällaisessa tapauksessa suoritus katkaistiin arviointityöhön liittyvän ajan säästämiseksi suoritusajan loppuessa. Arvioijille kerrottiin, ettei suorituksen katkeaminen ollut arvioinnin kannalta tärkeää ja että suoritus tuli arvioida katkaisemisesta huolimatta normaalisti.

Toinen ohje koski tehtävässä 2 mainitun Julien Fabre-Perrault'n sukupuolta. Monet oppijat näet olivat erehtyneet luulemaan tärkeää ranskalaista miesvierasta naiseksi. Nähtävästi monille tutkimukseen osallistuneille ei ollut selvää, että tehtävässä mainittu Julien-nimi viittaa mieheen eikä naiseen. Tämänkaltaisen kömmähdys ei ole mitenkään tärkeä kielitaidon arvioinnin kannalta, koska todellisessa kommunikaatiotilanteessahan kielenkäyttäjä mitä todennäköisimmin ei erehtyisi keskustelukumppaninsa sukupuolesta. Arvioijia kehoitettiin arvioidessaan osallistujien puhe-suorituksia olemaan kiinnittämättä mitään huomiota mahdolliseen sekaannukseen tehtävässä esintyvän ranskalaisvieraan sukupuolen suhteen.

Kolmas arvioijille annettu ohje koski professionaalisen tilanteen vaatiman kielenkäytön hallintaa. Tehtävän osalta edellytettiin, että tehtävän suorittajien tuli puhua tilanteeseen sopivalla tavalla. Tehtävän oikeanlainen suorittaminen vaati siten tietynlaista sosiokulttuurista osaamista ja sosiaalisiin suhteisiin liittyvien käytänteiden hallintaa eli erityisesti kohteliaisuutta ja teitittelyä. Arviointiohjeissa korostettiin, että mm. suomalaisessa kulttuurissa laajalle levinnyt sinuttelu ei tulisi kysymykseen tehtävän 2 kaltaisessa ammattikontekstiin liittyvässä tilanteessa, jossa esitellään tärkeä ranskalainen vieras suomalaiselle kollegalle. Sen sijaan suomalaiselle kollegalle puhuttaessa sinuttelu on tietenkin täysin hyväksyttävää. Arvioijia muistutettiin myös siitä, että tehtävää 2 koskevassa arviointiasteikossa oli erikseen mainittu kohteliaisuusmuotojen ja kommunikaatiotilanteen vaatiman muodollisuusasteen hallinta. Näiden piirteiden hallinta osoittautui ymmärrettävästi monille osallistujille erittäin haastavaksi. Perusohje arviointityöhön oli se, että esittelyn tuli olla ymmärrettävä ja mahdollinen samankaltaisessa todellisessa tilanteessa. Jos osallistuja käytti sinuttelumuotoa ranskalaiselle tärkeälle vieraalle puhuessaan, hänen taitotasoaan kehoitettiin kuitenkin laskemaan vain yhdellä tasolla.

Neljäs arvioijille annettu ohje liittyi tehtävänannon täyttymiseen. Jotkut tutkimukseen osallistuneet oppijat olivat näet joko ymmärtäneet tehtävänannon väärin tai lukeneet sen huolimattomasti. Tehtävä 2:n tehtävänannossa edellytettiin kahta asiaa:

1. Esitellään toisilleen puhujan työtoveri Kristiina Saukkonen- Riipinen ja tärkeä ranskalainen vieras Julien Fabre-Perrault.
2. Kerrotaan ranskalaiselle vieraalle, että suomalainen työtoveri puhuu hyvin ranskaa.

Osa suorittajista kertoi joko epähuomiossa tai studiotilanteen aiheuttamasta paineesta johtuen, että ranskalainen vieras Fabre-Perrault puhuu hyvin ranskaa sen sijaan, että heidän olisi tietenkin pitänyt kertoa ranskalaiselle vieraalle esittelijänä toimivan puhujan suomalaisen työtoverin Saukkonen-Riipisen puhuvan hyvin ranskaa. Arvioijia kehoitettiin jättämään huomiotta tämänkaltainen huolimattomuudesta johtuva sekaannus, jolla ei kuitenkaan ole sen suurempaa merkitystä tutkimukseen osallistumisen kaltaisessa epävirallisessa tilanteessa.

#### 4.4.3 Osallistujien puhe-suoritusten arviointi

Osallistujien puhe-suoritusten arviointiin osallistui yhteensä kolme arvioijaa, joista kaksi arvioi kaikki suoritukset molemmista tehtävistä. Nämä kaksi arvioijaa olivat toimineet arvioijina myös tutkimuksen Yki-korpuksen suorituksia tasoille asetettaessa. Arvioitavan aineiston suuruuden vuoksi edellä mainittujen kahden arvioijan lisäksi päätettiin käyttää nk. vaihtuvaa arvioijaa kolmantena arvioijana. Koska arvioitavia suorituksia oli niinkin paljon kuin 210 kpl/tehtävä ja koska arvioijat toimivat tehtävässään vapaaehtoisuuden pohjalta, päätettiin kolmannen arvioijan työtaakkaa helpottaa jakamalla arvioitavat puhe-suoritukset arvioijien kesken suunnilleen samankokoisiin ryhmiin. Täten jokainen arvioija, joka kuului kolmannen arvioijan ryhmään, sai arvioitavakseen noin 40 – 50 suoritusta kummastakin tehtävästä.

#### 4.4.4 Aineiston koodaus

Jotta tutkimuksen aikana eri lomakkeilla kerättyjä tietoja voitiin analysoida tilastollisesti, kaikki lomakkeiden sisältämät tiedot tuli muuttaa numeeriseen muotoon. Muuttuja 1 edusti itsearviointilomakkeissa käytössä olleita ”kyllä”-vastauksia ja vastaavasti muuttuja 0 edusti lomakkeiden ”ei”-vastauksia. Samalla tavoin arvioijien antamat taitotasot muutettiin numeeriseen muotoon siten, että esimerkiksi taitotaso A1 vastasi muuttujaa 1, taitotaso A2 muuttujaa 2 jne.

Osallistujan taustatiedot -lomakkeessa oli monia kysymyksiä, joissa oli monta eri vastauskategoriaa (esimerkiksi kuinka osallistujat käyttivät ranskan kieltä tutkimuksen aikaan). Jokainen lomakkeen kohta muutettiin numeeriseen muotoon siten, että jokaiselle tiedolle määriteltiin oma sarake, johon voitiin merkitä tietyn koodin saanut tieto (esim. sarakkeeseen 1 tuli suorittajanumero

kolmella numerolla ilmaistuna eli suorittaja numero 1 koodattiin koodilla 001, suorittaja numero 2 002 jne.).

Numeerisessa muodossa olevien koodikirjojen tietojen pohjalta tehtiin tutkimuksen tilastolliset analyysit. Koodikirjojen olemassaolo helpotti tutkimuksen edistymistä monessa eri vaiheessa mahdollistaen muun muassa sen, että milloin tahansa voitiin tarvittaessa tarkastaa jonkin tutkimukseen liittyvän tiedon oikeellisuus.

#### 4.4.5 Palautteen antaminen osallistujille

Tutkimukseen osallistuneille oppijoille annettiin erikseen palautetta (liite 9), joka koski heidän tutkimukseen osallistumistaan. Oppijoille selvitettiin palautteessa, että heidän puhe-suorituksiaan arvioi kolme kielitaidon arviointiin erikoistunutta arvioijaa, joista yksi oli ranskan kielen natiivi ja kaksi suomalaisia. Palautteessa kerrottiin myös, että jokainen kolmesta arvioijasta oli kuunnellut jokaisen suorittajan molemmat puhe-suoritukset (tehtävät 1 ja 2) ja antaneet siten itsenäisesti taitotasoarvion niistä käyttäen työssään Euroopan neuvoston kielitaidon arviointiasteikkoihin perustuvia 6-portaisia asteikoita. Oppijoiden saamien tasoarvioiden kerrottiin siten olevan kolmen arvioijan antama keskiarvo heidän puhe-suorituksistaan molemmissa suoritetuissa tehtävissä.

Oppijoille annettiin aluksi ranskan kielen puhumisen taidosta heidän oma arvionsa, jonka he olivat siis itse merkinneet tutkimukseen osallistuessaan puhumisen itsearviointilomakkeen etusivulle. Tämän jälkeen palaute sisälsi erillisen arvion sekä tehtävästä 1 että tehtävästä 2. Tasoarvioiden jälkeen palautteessa oli ymmärtämisen helpottamiseksi osallistujan saamaa tasoarviota / saamia tasoarvioita kuvaavat sanalliset määritelmät, joissa kerrottiin, mitä kyseisellä taitotasolla oleva kielenoppija tyypillisesti osaa tehdä. Palautteeseen sisältyi myös lyhyt selostus Euroopan neuvoston arviointiasteikkojen ideologiasta sekä kielitaidon arviointiin liittyvistä seikoista, joilla saattoi olla merkitystä oppijoiden saamien tasoarvioiden kannalta. Tässä palautteen osassa oppijoita muistutettiin siitä, että kyseiset tasoarviot oli annettu *tietyistä tehtävistä tietyinä ajankohdina* (alleviivattu palautteessa). Oppijoille kerrottiin, että suullinen kielitaito vaihtelee jokaisella jonkin verran tilanteesta toiseen ja että heidän saamaansa tasoarvioon ovat voineet vaikuttaa alentavasti tietyt seikat, kuten esimerkiksi studiotilanteen ”jännittävyys”, rajoitettu puheaika ja todelliseen kommunikatiotilanteeseen liittyvä mahdollisuus käyttää esimerkiksi eleitä tai käsimerkkejä. Palautteessa todettiin lisäksi, että olisi todennäköistä, että suurin osa tutkimukseen osallistujista suoriutuisi hieman paremmin vastaavasta todellisesta kielenkäyttötilanteesta kuin tutkimukseen osallistumistilanteesta. Palautteen lopuksi oppijoille annettiin yhteystiedot Euroopan neuvoston laatiman EVK:n sekä DIALANG-itsearviointityökalun verkkosivuille. Tutkimuksen tekijä antoi luonnollisesti osallistujille myös omat yhteystietonsa siltä varalta, että heillä olisi jotain kysyttävää tai kommentoitavaa tutkimukseen tai esimerkiksi saamiinsa tasoarvioihin liittyen.

## 5 AINEISTON ANALYYSIT JA TULOKSET

Tässä työn osassa tarkastellaan tutkimuksen aikana tehtyjä kvantitatiivisia ja kvalitatiivisia analyysejä sekä niiden tuloksia.

Luvussa 5.1 keskitytään tarkastelemaan tutkimuksessa käytettyjen itsearviointimittareiden eli tutkimusaineiston keräämisessä käytettyjen lomakkeiden luotettavuutta. Itsearviointilomakkeiden luotettavuudesta puhuttaessa lomakkeista käytetään itsearviointimittari-termiä. Kun puhutaan lomakkeiden täyttämiseen tai muuhun konkreettiseen toimintaan liittyvästä seikasta, käytetään ymmärrettävyyden vuoksi usein itsearviointilomake-termiä. On kuitenkin syytä huomioida, että joissakin kohdin valinta näiden kahden termin käytön välillä riippuu kontekstista. Siksi on mahdollista, että termejä käytetään joissakin kohdin synonyymeinä. Mittareiden luotettavuutta selvittäessä tarkastellaan mittarien toimivuutta kokonaisuutena sekä väittämien tai puhenäytteiden laatua ja vaikeutta niin klassisen (perinteinen osioanalyysi) kuin modernin testi-teorian välineitä (Rasch-analyysi) hyödyntäen.

Mittareiden luotettavuuden selvittämisen jälkeen tarkastellaan ulkoisten arvioijien arviointien luotettavuutta luvussa 5.2. Luvussa 5.3 tarkastellaan itsearviointien ja ulkoisten arviointien välisiä korrelaatioita. Luvussa 5.4 analysoidaan tutkimukseen osallistujilta kerättyjä laadullisia kommentteja, jotka liittyvät heidän itsearvioinnissa käyttämiinsä lomakkeisiin ja puhenäytteisiin. Luvussa 5.4.1 esitellään itsearviointi paperilla -mittaria koskevien kommenttien luokittelu. Kommentit on luokiteltu eri ryhmiin niiden sisältämien ominaispiirteiden mukaisesti. Luku 5.4.2 ja vielä tarkemmin sen alaluvut 5.4.2.1 (tehtävä 1) ja 5.4.2.2 (tehtävä 2) sisältävät kyseisiä tehtäviä koskevat kommenttianalyytit. Aluksi esitellään taulukkomuodossa eniten ja vähiten ”vaikea päättää” -rasteja saaneet näytteet. Sen jälkeen tarkastellaan osallistujien kommentteja, jotka koskevat kolmea eniten ja vähiten ”vaikea päättää” -rasteja kerännyttä näytettä. Vapaat kommentit on ryhmitelty kahdeksaan eri luokkaan. Luvussa pyritään myös luomaan kuuden yksityiskohtaisesti analysoidun puhenäytteen sisällön avulla yleiskuva siitä, mitkä kielen piirteet saattavat vaikuttaa siihen, että jokin kuunneltava näyte tuntuu oppijasta verrattavuuden kannalta joko harvinaisen

”selkeältä” (= vähän ”vaikea päättää” -rasteja) tai päinvastoin erityisen ”hankalalta” (= paljon ”vaikea päättää” -rasteja). Luku 5.4.3 sisältää yhteenvedon molempien tehtävien osalta eniten tai vähiten ”vaikea päättää” -rasteja keränneiden näytteiden ominaisuuksista.

On huomattava, että erityisesti kirjallisten itsearviointiväittämien analyysi painottuu enemmän huonosti toimineisiin väittämiin. Tähän on syynä se, että työssä on haluttu keskittyä etsimään mahdollisia yhteisiä tekijöitä tai piirteitä, jotka voisivat selittää kyseisten väittämien toimimattomuuden. Kuitenkin myös hyvin toimineita väittämiä tarkastellaan luvussa lyhyesti.

Tutkimuksen kvantitatiiviset analyysit tehtiin SPSS-ohjelmalla sekä erityisesti mittareiden ja arviointien ominaisuuksien selvittämiseen kehitettyjen Winsteps- ja Facets-ohjelmien avulla (Linacre 2011). Käytettyjen mittareiden luotettavuutta selvitettiin mm. korjattujen osio-testi-korrelaatioiden ja Cronbachin alfan avulla. Arvioijien luotettavuutta (yhdenmukaisuutta ja systemaattisuutta) selvitettiin mm. korrelaatioiden ja myös Facets-analyysin avulla.

Aineisto kvalitatiivisiin analyyseihin kerättiin aineistonkeruuprosessin aikana tutkimukseen osallistujilta keräämällä tietoa mm. itsearviointilomakkeissa käytettyjen itsearviointiväittämien ja kuunneltujen puhenäytteiden toimivuudesta ja ymmärrettävyydestä. Osallistujilta kerättiin myös itsearviointia koskevia vapaita kommentteja.

## 5.1 Itsearviointimittareiden luotettavuus

Luvussa 5.1 analysoidaan tutkimusaineiston keräämisessä käytettyjen kolmen itsearviointimittarin luotettavuutta yleisen toimivuuden, väittämien tai puhenäytteiden laadun tai sen kautta, kuinka vaikeaksi oppijat ovat väittämät tai puhenäytteet kokeneet. Luotettavuustarkastelu perustuu tilastotieteellisten analyysien tuloksiin, osallistujien mahdollisiin kommentteihin, viiden arvioijan arvioimiin näytteiden taitotasoihin sekä tutkimuksen tekijän omiin tulkintoihin.

### 5.1.1 Paperilla tehty itsearviointi

Itsearviointilomakkeen kaikki 42 yksityiskohtaista itsearviointiväittämää analysoitiin tilastollisia tunnuslukuja käyttäen. Näitä olivat sekä klassiseen testiteoriaan kuuluvat tunnusluvut (osio-testi-korrelaatiot ja Cronbachin alfa) että moderniin testiteoriaan kuuluvat erottelykykyä ja mallin sopivuutta (fit) kuvaavat tunnusluvut. Myös tutkimuksen aikana kielistudiossa kuunnellut tehtävää 1 ja tehtävää 2 koskevat näytteet analysoitiin edellä mainittuja tunnuslukuja käyttäen. Luotettavuuteen liittyvien analyysien tarkoituksena oli selvittää, ovatko tutkimuksen aikana käytetyt itsearviointiväittämät ja kuunnellut näytteet selviä ja ymmärrettäviä sekä näyttävätkö ne toimivan siten kuin niiden odotetaan toimivan. Jotta voitaisiin saada yleiskuva paperilla tehdyn itsearvioinnin toimivuudesta, lomakkeessa käytettyjen väittämien toimivuutta tarkastellaan alla yksityiskohtaisesti tunnusluku kerrallaan.



### 5.1.1.1 Mittarin toimivuus kokonaisuutena

Sekä perinteinen osioanalyysi (SPSS:n reliabiliteettianalyysi) että moderniin testiteoriaan perustuva Rasch-analyysi Winstepillä osoittivat, että tutkimuksessa käytetty, väittämiin perustuva itsearviointiväline toimi hyvin. Cronbachin alfa tälle 42-osioiselle mittarille oli .93. Mittarin erottelukyky oppijoille oli 3.97 ja osioille 10.38 Winstep-analyysin mukaan (ks. liitteet 10 ja 11). Tämä tarkoittaa, että mittari pystyy erottelamaan oppijat luotettavasti noin neljään ryhmään heidän itsearvioidun kielitaitonsa perusteella, kun taas mittarin osiot voidaan aineiston perusteella jakaa ainakin yhteentoista eri luokkaan (taitotasoon).

Kummatkin analyysit osoittivat kuitenkin, että yksi väittämistä, numero 29, käyttäytyi eri tavalla kuin muut. Sen osio-testi-korrelaatio oli negatiivinen (-.51) ja Winstepin Infit mean square -arvo poikkeavan suuri (3.32, kun ihanearvo on 1.0 ja epäilyttävän osion raja on noin 1.3 tärkeissä tutkinnoissa ja selvästi ongelmallisen osion raja-arvo on noin 2.0; ks. Linacre 2011). Lisäksi yksi väittämistä (numero 20) oli niin helppo, että kaikki tutkittavat vastasivat siihen ”kyllä”, joten koska vastauksissa ei esiintynyt mitään vaihtelua, se jouduttiin jättämään pois osioanalyysistä. Yksityiskohtaisempi analyysi itsearviointiväittämistä seuraa tuonnempana, jolloin käsitellään osioiden korrelaatioita koko mittariin.

Kun osio 29 poistettiin ja analyysit suoritettiin uudestaan, mittarin kokonaislaatua kuvaavat tunnusluvut nousivat. Cronbachin alfa nousi .94:ään, Winstepin oppijoiden erottelukykyarvo 4.43:een ja osioiden erottelukyky 11.25:een (ks. liite 10).

### 5.1.1.2 Väittämien laadun analyysi

Seuraavaksi tarkastellaan itsearviointiväittämien ominaisuuksia ja laatua sekä perinteiseen testiteoriaan kuuluvien osio-testi-korrelaatioiden että moderniin testiteoriaan kuuluvan Rasch-analyysin tulosten avulla. Samalla analysoidaan väittämien sisältöä laadullisesti. Sekä osio-testi-korrelaatioanalyysin että Rasch-analyysin tarkoituksena on osoittaa, näyttääkö jokin tietty osio (tässä tapauksessa väittäjä) toimivan samansuuntaisesti kuin samaan kokonaisuuteen kuuluvat muut väittämät eli mittaavan samaa ominaisuutta kuin muut. Väittämien laadullisen / sisällön analyysin avulla pyritään sitten osoittamaan, löytyykö väittämien väliltä jokin tai joitakin yhteisiä elementtejä, jotka selittäisivät niiden toimivuuden tai toimimattomuuden. Jos väittäjä toimii hyvin, voidaan olettaa, että esimerkiksi pitkälle kieliopinnoissaan edenneet oppijat vastaisivat ”kyllä” niihin väittämiin, joiden kielenkäyttötilanne vaatii kyseisen kielen hyvää hallintaa.

Paperilla tehdyn itsearviointilomakkeen väittämät ryhmiteltiin taulukkoihin 7 ja 8 taitotasoittain (A1–C2) siten, että taulukon ensimmäisessä sarakkeessa on väittämien taitotasot alkaen taitotasosta A1. Toisessa sarakkeessa on huonosti toimineiden väittämien numerot sekä sulkeissa niiden vastausjakaumat (”kyllä”-vastausten (= 1. luku) ja ”ei”-vastausten (= 2. luku) lukumääriä kyseisessä väittämässä). Kolmannessa sarakkeessa on tarkasteltavien tunnuslukujen arvot

(perinteisessä osioanalyysissä osio-testi- korrelaatiot ja Rasch-analyysissä Infit mean square ja Infit Zstd -arvot). Taulukon oikeanpuoleisessa reunassa esitellään hyvin toimineet väittämät taitotasoin ja niiden tunnuslukujen arvot samalla tavalla kuin huonosti toimineiden väittämien osalta.

Väittämän hyvyttä tarkasteltiin ensin perinteisen osioanalyysin näkökulmasta osio-testi-korrelaatioiden avulla. Ne saatiin sekä SPSS:n reliabiliteettianalyysistä että Winstepin osio-testi-korrelaatioanalyysistä. Jos osio-testi-korrelaation arvo oli yli .30, väittämän voidaan todeta arvon suuruudesta riippuen toimineen joko kohtalaisesti tai hyvin (vrt. Bachman 2004, 138). Kaikki arvon .30 alle jäävät arvot puolestaan kertovat, että väittämän toimivuudessa on mahdollisesti ongelmia – se on ehkä epäselvästi muotoiltu tai mittaa jotakin muuta ominaisuutta kuin muut väittämät.

Perinteisen osioanalyysin jälkeen väittämiä tarkasteltiin moderniin testi-teoriaan perustuvan osioanalyysin (Rasch-analyysi) näkökulmasta, jolloin tunnuslukuina käytettiin Infit mean square ja Infit Zstd -arvoja, joita voidaan pitää johdonmukaisuuteen liittyvinä indekseinä. Infit mean square -arvon suositusrajoissa on esiintynyt pientä variaatiota aikojen kuluessa. Muun muassa Linacre (2002), Facets-ohjelmiston kehittäjä, suositti aiemmin 2000-luvun alussa, että kyseisen arvon tulisi sijoittua välille .5 ja 1.5, jotta sitä voitaisiin pitää mittaamisen kannalta sopivana arvona. Tässä tutkimuksessa, jossa tarkastelu kohdistuu mittareiden yksittäisiin osioihin, haluttiin kuitenkin käyttää high stakes -testeissä noudatettuja edellistä tiukempia rajoja. Siten Rasch-analyysin tunnuslukujen lähtökohtana käytettiin sitä, että väittämien Infit mean square -arvon tulee sijaita .7 ja 1.3 välillä ja Infit Zstd -arvon rajojen +2 ja -2 sisällä, jotta niitä voidaan pitää hyvin toimivina mittarin osioina (Linacre 2011). Infit mean square arvojen tarkoitus on osoittaa, missä määrin jokin osio sopii Rasch-analyysin malliin. Kyseisen arvon odotusarvo on 1.0. Jos arvo on pienempi kuin .70, se tarkoittaa, että tarkasteltava osio on liian ennustettava ja tietyllä tavalla hallitsematon. Osio saattaa toimia epäodotuksenmukaisesti hyvin, ja se sisältää siten vähemmän sattumanvaraisuutta, kuin voisi odottaa. Jos taas Infit mean square -arvo on suurempi kuin 1.3, kyseessä on osio, joka ei sovi malliin ja toimii hyvin epäodotuksenmukaisella tavalla (Linacre 2002, 878). Yleisesti ottaen outfit-arvot ottavat aineistoa analysoitaessa huomioon myös ääriarvot, kun taas aineistossa esiintyvät ääriarvot vaikuttavat vähemmän infit-arvoihin (Bond & Fox 2007).

On huomattava, että Rasch-analyysin avulla tarkasteltavia tunnuslukuja oli siis 2 (Infit mean square ja Infit Zstd), kun taas perinteistä osioanalyysiä käytettäessä tarkasteltavia tunnuslukuja oli vain 1 (osio-testi-korrelaatio). Tästä seuraa, että kun väittämiä lajiteltiin alla olevassa taulukossa huonosti ja hyvin toimineisiin Rasch-analyysin mukaan, väittämiä saattoi joutua lajitelluksi huonosti toimivaksi väittämäksi helpommin Rasch-analyysissä kuin perinteisessä osioanalyysissä, koska jos jompikumpi Rasch-analyysissä tarkasteltavista tunnusluvuista ei ollut suositusten mukainen, väittämiä sijoitettiin taulukon huonosti toimivien väittämien joukkoon. Tämä yksityiskohta on syytä ottaa huomioon perinteisen osioanalyysin ja Rasch-analyysin avulla huonosti toimivien väittämien kokonaisuutta tarkasteltaessa. Toisaalta on huomattava, että vaikka Rasch-analyysiin liittyen ilmoitetaan sekä Infit mean square että Infit

Zstd -arvo, näistä Infit mean square on arvo, joka on tärkeämpi, koska se osoittaa selkeämmin mahdollisen ongelman olemassaolon. Niinpä myös Raschanalyysiin perustuviissa tutkimusartikkeleissa raportointi keskittyy yleensä Infit mean square -arvoihin. Tästä syystä painotan näistä kahdesta tilastollisesta arvosta Infit mean square -arvoa. Kaikista käsitellyistä väittämistä mainitaan myös väittämän vaikeustaso eli se, pysyvätkö väittämät niille määrätyillä taitotasoilla vai ovatko tutkimukseen osallistujat kokeneet ne helpoimmiksi tai vaikeammiksi kuin niiden määritelty vaikeustaso on.

Taulukkojen 7 ja 8 jälkeen esitellään ensin molempien testiteorioiden näkökulmasta joitakin hyvin toimineita väittämiä, minkä jälkeen tarkasteluvuorossa ovat instrumentissa huonosti toimineet väittämät molempien osioanalyysien näkökulmasta. Tarkennettakoon tässä yhteydessä, että termejä "hyvin" ja "huonosti" toimineet osiot käytetään alla asian käsittelyn helpottamiseksi, ei absoluuttisina määreinä. Erilaisten osioanalyysien tuottamat tulokset kertovat vain sen, mitkä osiot (väittämät) eroavat toisista osioista ja kuinka paljon. Esitetyt raja-arvot ovat myös viitteellisiä, eikä poikkeavan arvon saanut osio ole automaattisesti huono, vaan lopullinen arvio siitä, mistä syystä osio poikkeaa muista täytyy tehdä tarkastelemalla sen sisältöä. Vasta sen jälkeen voidaan päättää onko osio niin huono, että se pitäisi esimerkiksi poistaa kokonaan mittarista.

TAULUKKO 7 Kirjallisiin itsearviointiväittämiin perustuva itsearviointi: perinteisen osioanalyysin mukaiset hyvin ja huonosti toimineet väittämät.

Väittämän taitotaso	Huonosti toimineet väittämät	Osio-testi-korrelaatio	Hyvin toimineet väittämät	Osio-testi-korrelaatio
<b>A1 (7)</b>				
A1	41 (207-2)	.11	32	.49
A1	19 (206-2)	.12	12	.40
A1	9 (202-7)	.18	30	.34
A1	27 (203-6)	.19		
<b>A2 (7)</b>				
A2	29 (130-79)	-.50	31	.58
A2	1 (203-6)	.22	11	.57
A2			14	.45
A2			10	.39
A2			5	.34
<b>B1 (9)</b>				
B1			40	.70
B1			25	.67
B1			7	.68
B1			22	.69
B1			17	.63
B1			26	.62
B1			39	.59
B1			4	.53
B1			23	.39
<b>B2 (6)</b>				
B2			37	.73
B2			2	.69
B2			42	.68
B2			35	.68
B2			8	.66
B2			36	.52
<b>C1 (6)</b>				
C1			33	.64
C1			24	.63
C1			3	.56
C1			16	.53
C1			13	.51
C1			34	.31
<b>C2 (6)</b>				
C2				
C2	21 (3-206)	.21		
C2			38	.54
C2			18	.52
C2			6	.51
C2			15	.49
			28	.42

TAULUKKO 8 Kirjallisiin itsearviointiväittämiin perustuva itsearviointi: modernin osioanalyysin (Rasch-analyysi) mukaiset hyvin ja huonosti toimineet väittämät.

Väittämän taitotaso	Huonosti toimineet väittämät	Infit mean square	Infit Zstd	Hyvin toimineet väittämät	Infit mean square	Infit Zstd
<b>A1 (7)</b>						
A1	19	1.30	.7	12	.85	-.6
A1				32	.87	-.8
A1				30	1.01	.1
A1				27	1.12	.5
				9	1.19	.7
				41	1.20	.5
<b>A2 (7)</b>						
A2	29	3.32	9.9	14	.89	-.5
A2	5	1.68	5.5	10	.89	-.4
A2	31	.75	-2.0	1	.94	-.1
A2				11	.95	-.4
A2						
<b>B1 (9)</b>						
B1	23	1.40	3.7	40	.82	-1.7
B1	25	.77	-2.4	7	.83	-1.8
B1	22	.78	-2.4	26	.98	-.1
B1				17	1.01	.2
B1				39	1.03	.3
B1				4	1.11	1.0
B1						
B1						
<b>B2 (6)</b>						
B2	36	1.33	2.6	35	.83	-1.5
B2	37	.75	-2.6	2	.88	-1.2
B2				42	.91	-.8
B2				8	1.01	.1
B2						
B2						
<b>C1 (6)</b>						
C1	13	.70	-1.8	24	.78	-1.8
C1	33	.78	-2.0	3	.88	-.8
C1				34	.93	-.2
C1				16	.94	-.4
C1						
C1						
<b>C2 (6)</b>						
C2				15	.72	-1.6
C2				6	.87	-.7
C2				28	.90	-.4
C2				21	.93	.0
C2				18	1.06	.5
C2				38	1.13	1.1

### 5.1.1.3 Perinteisen osioanalyysin ja Rasch-analyysin mukaiset hyvin toimineet väittämät

Kuten aiemmin jo todettiin, paperilla tehdyn itsearviointilomakkeen 42 yksityiskohtaista itsearviointiväittämää toimivat kokonaisuudessaan hyvin. 42 väittämästä vain seitsemän korrelaatiokerroin oli matala (käsitellään jäljempänä). Tasoilta B1, B2 tai C1 ei löydy yhtään väittämää, joka olisi toiminut huonosti, vaan näiden tasojen väittämät näyttävät toimineen hyvin ja niiden korrelaatioarvot ovatkin pääosin korkeita.

Myös Rasch-analyysin mukaan paperilla tehdyn itsearvioinnin väittämät toimivat pääosin hyvin. 42 väittämästä neljässä Infit mean square tai seitsemässä Infit Zstd -arvo oli hyväksyttävien rajojen ulkopuolella. Taitotason C2 kaikki väittämät toimivat hyvin Rasch-analyysin pohjalta.

Aluksi tarkastellaan yksityiskohtaisesti joitakin eritasoisia väittämiä, jotka ovat toimineet hyvin sekä perinteisen että Rasch-analyysin näkökulmasta. Tarkastelun tarkoituksena on yrittää selvittää, sisältävätkö kyseiset väittämät mahdollisesti sellaisia elementtejä, jotka selittäisivät niiden hyvän toimivuuden. Tarkastelun kohteena ovat väittämät 40, 2 ja 42.

#### Väittämä 40

(taso B1; korrelaatiokerroin .70; Infit mean square .82; Infit Zstd -1.7; vastausjakauma 77–131)

Osaan kertoa melko yksityiskohtaisesti kokemuksistani ja kuvata tunteita ja reaktioita.

#### Väittämä 2

(taso B2; korrelaatiokerroin .69; Infit mean square .88; Infit Zstd -1.2; vastausjakauma 101–108)

Osaan puhua ranskaa luontevasti, sujuvasti ja tehokkaasti tutuissa tilanteissa.

#### Väittämä 42

(taso B2; korrelaatiokerroin .68; Infit mean square .91; Infit Zstd -.8; vastausjakauma 106–103)

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

Kun tarkastellaan edellä esiteltyjä sattumanvaraisesti valittuja hyvin toimineita väittämiä, voidaan todeta, että ne kaikki sisältävät viittauksia jokapäiväiseen kontekstiin tai joitakin persoonaan liittyviä elementtejä (kuten omat kokemukset ja tunteet, tutut tilanteet ja konkreettiset asiat). Nämä kolme esimerkkipäätämää ovat kaikki myös suhteellisen lyhyitä eivätkä ne sisällä konjunktiolla toisiinsa liitettyjä virkkeitä. Lisäksi väittämässä 42 on sulkeissa selvennetty esi-

merkein, mitä väittämässä tarkoitetaan konkreettisilla asioilla. Väittämien vaikeustasosta voidaan todeta sen verran, että tutkimukseen osallistujat ovat pitäneet väittämiä 2 (B2) ja 42 (B2) juuri sen tasoisina väittäminä kuin mikä niiden määritelty taitotaso on. Sen sijaan väittäjä 40 (B1), jonka taitotaso on B1, sijoittuu huomattavasti todellista tasoaan ylemmäksi B2–C1-tasoisien osioiden joukkoon. Oppijat ovat siten pitäneet väittämää 40 vaikeampana kuin sen määritelty taitotaso on (kuvio 5). Väittäjä 40:n sisältöä tarkastellaan lisää osioiden vaikeustasoa käsiteltäessä jäljempänä.

On syytä huomata, että perinteisen osioanalyysin ja Rasch-analyysin tuottamat tulokset eivät välttämättä pidä samoja väittämiä hyvin toimivina. Kuitenkin on tärkeää tiedostaa, että perinteinen osioanalyysi ja Rasch-analyysi yleisellä tasolla antavat samansuuntaista tietoa osioista, mikä tarkoittaa sitä, että suurin osa lomakkeen osioista nousee esille sekä perinteisessä että Rasch-analyysissä, mikäli osiossa voidaan epäillä olevan jotain ongelmaa. Esimerkkeinä hyvin toimivasta väittämästä perinteisen osioanalyysin mukaan voidaan ottaa väittäjä 25:

#### **Väittäjä 25**

(taso B1; korrelaatiokerroin .67; Infit mean square .77; Infit Zstd -2.4; vastausjakauma 145–64 )

Osaan keskustella tutuista asioista silloinkin kun minulla ei ole ollut mahdollisuutta valmistautua puhumaan.

Kuten yllä olevista arvoista näkyy, väittäjän 25 osio-testi-korrelaatio on .67, ja sen Infit mean square pysyy sallituissa rajoissa (.70–1.30), mutta sen sijaan sen Infit Zstd menee hyväksyttävien raja-arvojen +2 ja -2 yli. Väittäjä 25 pysyy sille määrätyllä taitotasolla (B1) (kuvio 5). Väittäjä 25 oli mukana Euroopan komission tekemässä Eurooppalaisessa kielitaitoindikaattoritutkimuksessa, jonka osana selvitettiin 54 000 eurooppalaisen koululaisen taitoja oppilaiden ensimmäisen ja toisen vieraan kielen luetun ja kuullun ymmärtämisen sekä kirjoittamisen osatiedoissa. Vaikka puhuminen ei kuulunut tutkimuksessa mukana olleisiin varsinaisiin osataitoihin, tutkimuksessa selvitettiin myös puhumisen itsearviointiväittämien vaikeusastetta oppilaiden näkökulmasta. Puhumisen itsearviointiväittämät tasolla B1 (väittäjä 25) ja B2 sijoittuivat heti tutkimuksessa vaikeusasteeltaan ylimmäisenä olevien kirjoittamisen itsearviointiväittämien alle (First European Survey on Language Competences. Final Report 2012, 29).

Samalla tavalla voidaan ottaa toinen esimerkki väittämästä (väittäjä 30), joka Rasch-analyysin mukaan toimii erinomaisesti, mutta joka ei taas perinteisen osioanalyysin näkökulmasta kuuluisi erinomaisesti toimineisiin:

**Väittäjä 30**

(taso A1; korrelaatiokerroin .34; Infit mean square 1.01; Infit Zstd .1; vastausjakauma 195–14 )

Osaan pyytää ihmisiltä esineitä ja asioita ja vastaavasti antaa pyydettyjä esineitä ja asioita.

Väittäjän 30 osio-testi-korrelaatio jää perinteisessä osioanalyysissä melko alhaiseksi (.34), kun taas sen Rasch-analyysiä edustavat arvot Infit mean square (1.01) ja Infit Zstd -arvot (.1) ovat erittäin hyviä. Suurin syy matalaan osio-testi-korrelaatioon lienee se, että valtaosa oppijoista arvioi pystyvänsä suoriutumaan ranskaksi tällaisesta tilanteesta, jolloin vähäinen vaihtelu vastauksissa ei edes mahdollista kovin korkeaa korrelaatiota. Väittäjä 30 pysyy hyvin sille määrättyllä taitotasolla (ks. kuvio 5 s. 154).

*Yleistä huonosti toimineista väittäjistä*

Kaikki seitsemän perinteisen osioanalyysin mukaan huonosti toiminutta väittäjää sijoittuvat taitotasoasteikon ylä- tai alapäähän: tasolle A1 (4 kpl), tasolle A2 (2 kpl) ja tasolle C2 (1 kpl). Todettakoon samalla, että näiden kolmen tason hyvin toimineiden väittäjien korrelaatiot ovat jonkin verran alhaisempia kuin niiden tasojen, joilla huonosti toimineita väittäjiä ei ollut lainkaan.

Rasch-analyysissä väittäjiä tarkasteltiin siis Infit mean square ja Infit Zstd -tunnuslukujen avulla. Myös sen tarkastelun pohjalta huonosti toimivia väittäjiä on paperilla tehdyssä itsearviointilomakkeessa yhteensä 7 kappaletta. (Jos aineistosta poistettu toimimaton väittäjä 29 jätetään pois laskuista, huonosti toimineita väittäjiä on Rasch-analyysin pohjalta paperilla tehdyssä itsearviointilomakkeessa vain 6 kappaletta.). Jos tarkastellaan väittäjien huonosti toimivuutta siten, että molempien tarkasteltavien arvojen tulisi olla suositeltavien raja-arvojen ulkopuolella tietyssä väittäjässä, niin paperilla tehdyssä itsearviointilomakkeessa on vain 4 huonosti toimivaa väittäjää. Useimmissa tapauksissa kyse on siis siitä, että jompikumpi tarkasteltavista arvoista, yleensä tämän tutkimuksen tarkastelun kannalta vähemmän tärkeä Infit Zstd -arvo, on suositeltavien rajojen ulkopuolella. Tämän vuoksi huonosti toimivuutta on Rasch-analyysissä mahdollista tulkita monella tapaa; tärkeintä on pitää mielessä, että jokainen poikkeava arvo voi olla mahdollisen ongelman osoittaja. Kaikkia kolmea arvoa eli osio-testi-korrelaatiota, Infit mean square ja Infit Zstd -arvoa olisikin paras tarkastella toisiaan täydentävinä arvoina. Huonosti toimivat väittäjät jakautuvat melko tasaisesti eri tasoille; vain C2-tasolla ei ole huonosti toimivia väittäjiä. Rasch-analyysin mukaan kolme itsearviointimittarin väittämistä näyttäisi olevan erityisen ongelmallisia. Näissä kaikissa kolmessa väittäjässä kumpikaan tarkasteltavista edellä mainituista tunnusluvuista ei pysy hyväksyttävissä rajoissa. Huonoiten toimineet väittäjät ovat väittäjät 5 (A2), 23 (B1) ja 36 (B2). Tarkastelen niitä erikseen perinteisen osioanalyysin huonosti toimineiden väittäjien jälkeen. Seuraavaksi kuvataan hiukan tarkemmin huonosti toimineiden väittäjien yleisiä ominaisuuksia.



#### 5.1.1.4 Perinteisen osioanalyysin mukaiset huonosti toimineet väittämät

Käsittelen nyt jokaisen huonosti toimineen väittämän ja pohdin samalla mahdollisia syitä väittämän toimimattomuuteen. Käytän tarkastelussa apuna myös mahdollisia osallistujien kommentteja kyseisestä väittämästä<sup>16</sup>. Tarkastelu aloitetaan perinteisen osioanalyysin mukaisilla huonosti toimivilla väittämillä (7 kpl) ja sen jälkeen tarkastelua jatketaan vastaavilla Rasch-analyysissä esille nousseilla väittämillä. Myös perinteisen osioanalyysin osoittamiin huonosti toimiviin väittämiin on merkitty vertailun helpottamiseksi Rasch-analyysin Infit mean square ja Infit Zstd -arvot. On yleisesti todettava, että vaikka perinteisen osioanalyysin antama korrelaatiokerroin olisikin alhainen, niin saman osion Infit mean square ja Infit Zstd -arvot pysyvät hyväksyttävien raja-arvojen sisäpuolella, ei matala korrelaatiokerroin yksistään välttämättä tarkoita sitä, että osion toimivuudessa olisi ongelmaa.

##### Väittämä 1

(taso A2; korrelaatiokerroin .22; Infit mean square .94; Infit Zstd -.1; vastausjakauma 203 –6)

Osaan keskustella jokapäiväisistä käytännön asioista yksinkertaisesti, jos keskustelukumppani puhuu selkeästi ja hitaasti.

Väittämä 1:n korrelaatiokerroin on heikohko (.22), mutta toisaalta kaksi muuta tarkasteltavaa arvoa pysyvät hyvin suositeltavien rajojen sisäpuolella. Kyseessä on itsearviointimittarin ensimmäinen väittämä, ja tästä johtuen onkin mahdollista, että väittämän paikalla on vaikutusta saatuihin vastauksiin. Itsearviointilomakkeiden täyttämisen tuntuu olevan vaikeutena se, että ohjeistuksesta huolimatta kukin osallistuja tulkitsee lomakkeen väittämiä omalla tavallaan. Väittämien riittävän yksiselitteinen muotoilu onkin haasteellista, mutta paraskaan muotoilu ei poista persoonakohtaista väittämien tulkintaa. On huomattava, että oppijat ovat pitäneet tätä väittämää helpompana kuin sille määritelty taitotaso on (A2). Siten tämä väittämä sijoittuu A1-tasoisten väittämien joukkoon väittämien vaikeustason kuvaavassa kuviossa (kuvio 5). Tässä aloittelevan ranskan kielen oppijan kommentti väittämän 1 sisällöstä ja väittämien tulkinnasta yleisestikin:

*"Jokapäiväiset asiat on aika laaja käsite, joten tulkitsin tämän niin, etten vielä osaa. Monet muutkin lomakkeen kohdat on ehkä parempi tulkita 'luulen osaavani' eikä 'osaan'" (1).*

<sup>16</sup> Numero suoran lainauksen jäljessä viittaa siihen taitotasoon, jolle osallistuja on oman puhumisen taitonsa arvioinut.

**Väittämä 9**

(taso A1; korrelaatiokerroin .18; Infit mean square 1.19.; Infit Zstd .7; vastausjakauma 202–7)

Osaan vastata haastattelussa esitettyihin, yksinkertaisiin ja suoriin henkilökohtaisia asioita koskeviin kysymyksiin, kun ne esitetään hitaasti ja selkeästi.

Tässä väittämässä on kyse samanlaisesta tilanteesta kuin äsken eli perinteisen osionalyysin mukainen osio-testi-korrelaatio on huono (.18), mutta sen sijaan modernin testiteorian mukaiset Infit mean square ja Infit Zstd -arvot ovat kunnossa. Väittämä pysyy vaikeustasoltaan sille määrättyllä taitotasolla (A1) (kuvio 5). Kyseessä on alimman tason A1 väittämä, joka nähtävästi ei ole riittävän yksiselitteinen käsitteiltään, mikä näkyy joidenkin osallistujien kommentissa.

*”Kysymys 9: riippuu millaisesta haastattelusta on kyse” (2).*

Haastattelun tyyppi voisikin olla selvästi määritelty väittämässä (esim. virallinen, ei-virallinen, yksityiselämään tai työkontekstiin liittyvä), mikä varmasti vähentäisi väittämän sisällön monitulkintaisuutta. Toisessa väittämää 9 koskevassa kommentissa tulee esille yleisempi ongelma, joka liittyy tutkimuksessa käytettyjen lomakkeiden täyttämiseen eli siihen, mikä on se ajallinen konteksti, jonka perusteella väittämiin vastataan<sup>17</sup>.

*”9 on vaikea siksi, etten tiedä tämänhetkistä kielitaitoni tasoa. Asuessani Ranskassa puoli vuotta vuonna 1996 selvisin kysytynlaisista tilanteista, mutta siitä on kauan...”(2).*

Yllä oleva kommentti liittyy DIALANG-järjestelmän ”tietoa itsearviointista” -osioon. Osiossa listataan mahdollisia syitä eroavaisuuksiin oppijan oman itsearviointin ja DIALANG-järjestelmän antaman taitotason välillä. Edellistä kommenttia koskevat selostukset siitä, että oppija saattaa käyttää kieltä epä-säännöllisesti tai että hän ei ole käyttänyt kieltä vähään aikaan. Nämä molemmat syyt voivat vaikeuttaa oman arvion muodostamista oppijalla tutkimukseen osallistumishetkellä olevasta kielitaidon tasosta.

**Väittämä 19**

(taso A1; korrelaatiokerroin .12; Infit mean square 1.30; Infit Zstd .7; vastausjakauma 206–2)

Osaan ilmaista yksinkertaisia numeroita, määriä, hintoja ja aikoja.

<sup>17</sup> Tutkimukseen osallistujille annettiin ohje vastata väittämiin tämänhetkisen tilanteen pohjalta.

On huomattava, että väittämä 19 on ainoa väittämä, jonka sekä perinteisen osioanalyysin että Rasch-analyysin tulokset osoittivat sisältävän mahdollisia ongelmia. Korrelaationkerroin on huono (.12) ja samanaikaisesti Infit mean square -arvo on ylärajalla (1.30). Onkin hyvin todennäköistä, että tämä väittämä sisältää elementtejä, jotka saattavat selittää sen huonosti toimimisen. Vaikka kyseessä on alimman tason väittämä, tutkimukseen osallistujilla näyttää olleen vaikeuksia väittämän kanssa. Yhtenä selityksenä vaikeuksille voidaan varmasti pitää sitä, että ranskan ja suomen kielen tavat numeroiden muodostamiseen eroavat suuresti toisistaan. Ranskan kielessä käytetään tiettyjen kymmenlukujen muodostamiseen kerto- ja yhteenlaskun yhdistelmiä (esimerkiksi  $90 = 4 \times 20 + 10$ ), jotka ovat numeroita muodostettaessa tuntemattomia suomalaisille oppijoille. On siten todettava, että vaikka tutkimuksessa käytettyjen väittämien tasot perustuvat laajaan kansainväliseen asiantuntijayhteistyöhön, näyttäisi siltä, että joissakin kielissä jotkut asiat ovat kielenoppijalle hankalampia kuin toisissa. Tämän vuoksi voidaan olettaa, että suomalaisten oppijoiden mielestä ranskan kielessä käytössä oleva kymmenlukujen muodostamistapa on verrattain vaikea hallita, koska se on niin erilainen kuin suomen kielen tapa muodostaa kymmenlukuja. Suhteellisen heikkojen tunnuslukujen vuoksi olisi houkuttelevaa tehdä johtopäätös, jonka mukaan ainakin osa tutkimukseen osallistuneista oppijoista mitä luultavimmin sijoittaisi väittämän 19 vaikeustason ylemmäksi taitotasoteikossa kuin taitotasolle A1. Tämä tulkinta ei kuitenkaan saa tukea, kun tarkastellaan tämän väittämän osalta osioiden vaikeustasoa kuvaavaa kuviota (kuvio 5), sillä kuvion mukaan väittämä 19 pysyy hyvin sille määritellyllä A1-tasolla, mikä tarkoittaa sitä, että oppijat ovat pitäneet keskimäärin kyseistä väittämää A1-tasoisena vaikeusasteeltaan.

#### **Väittämä 20**

(taso A2; -)

Osaan tervehtiä, kysyä kuulumisia ja vastata yksinkertaisiin asioihin, joita minulle kerrotaan.

Väittämä 20 on esimerkki väittämästä, joka ei ole toiminut lainkaan. Kaikki tutkimukseen osallistujat ovat vastanneet väittämään ”kyllä”, ja sen vuoksi väittämän osalta ei ole olemassa minkäänlaista varianssia. Väittämä 20 on siis tutkimuksen kannalta turha, sillä sen avulla ei saada minkäänlaista tietoa, ja siten se on poistettu analyyseistä.

#### **Väittämä 21**

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

Puhun ranskan kieltä melkein yhtä hyvin kuin äidinkieltäni.

Väittämän 21 korrelaatiokerroin on jälleen heikko, vaikka samalla kaksi muuta moderniin testiteoriaan kuuluvaa arvoa ovat hyviä. Väittämän vaikeustasoa tutkittaessa voidaan todeta sen pysyvän tasolla C2; väittämää on pidetty kaikista vaikeimpana lomakkeen väittämistä (kuvio 5). Tarkasteltaessa väittämää ensimmäinen mieleen tuleva kysymys on, kokevatko tämänkaltaiseen tutkimukseen osallistuvat oppijat tai vieraan kielen oppijat yleisestikään hallitsevansa omaa äidinkieltään erityisen hyvin. Varsinkin vieraita kieliä opiskelevat henkilöt huomaavat usein jossain vaiheessa opintoja, että oman äidinkielen erinomainen hallinta ei ole suinkaan itsestäänselvyys. Toinen kysymys, josta ihmiset väittelevät helposti, on se, voiko vierasta kieltä koskaan hallita samalla tasolla kuin omaa äidinkieltään. Tästä esimerkkinä pidemmälle ranskan kielen oppinnoissaan edenneen oppijan kommentti:

*"Vaikea määritellä 'melkein yhtä hyvin kuin äidinkieltäni'. Koen, että ranskani on vahva, mutten usko, että voisi oppia jonkun kielen kuin äidinkieltäni ellei asu ko. kieltä puhuvassa maassa ja käytä ko. kieltä jatkuvasti. Vastaisin mieluummin niin päin 'etten puhu ihan yhtä hyvin kuin äidinkieltäni'" (5).*

Eräs osallistuja lähestyy väittämää 21 koskevassa kommentissaan asiaa siitä näkökulmasta, että tutkimuksen väittämissä esiintyvien kielenkäyttötilanteiden hallinta on yleiselläkin tasolla haasteellista ilman, että haastavuudella olisi jokin tekemistä äidinkielen tai vieraan kielen kanssa:

*"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 vakuutusterminologiaa pystyäkseni puhumaan aiheesta asiantuntevasti)" (3).*

#### **Väittämä 27**

(taso A1; korrelaatiokerroin .19; Infit mean square 1.12; Infit Zstd .5; vastausjakauma 203–6)

Osaan tehdä ostoksia tilanteissa, joissa voin osoittamalla tai muilla eleillä tukea sanotavaani.

Korrelaatiokerroin on alhainen (.19) väittämässä 27. Vaikeustasoltaan tämä väittämä pysyy omalla tasollaan eli A1-tasolla (kuvio 5). Osallistujat eivät valitettavasti ole kommentoineet väittämää 27, joten tulkinat niistä syistä, jotka voisivat selittää väittämän toimimattomuutta, jäävät tutkimuksen tekijän tulkinnoiksi. Kyseessä on alimmalle taitotasolle A1 sijoittuva väittämä, johon vain kuusi osallistujaa on vastannut "ei". Kun otetaan huomioon, että väittämässä esitellyn kielenkäyttötilanteen hallintaan riittää hyvin alkeellinen kielitaito, voitaisiin olettaa, että kaikkien näiden kuuden "ei"-vastauksen antaneen henkilön kielitaidon taso sijoittuisi tehtävissä 1 ja 2 tasolle A1. Kuuden osallistujan taitotasot ovat olleet seuraavat:

Kaksi osallistujaa tasolla A1, yksi tasolla A2, kaksi tasolla A2 ja B1 ja yksi tasolla C1. On selvää, että viimeisen osallistujan "ei"-merkintä osoittaa sen, että kyseinen henkilö on epähuomiossa valinnut väärän vaihtoehdon.

Kun tarkastellaan väittämän esittämän tilanteen yksinkertaisuutta, on yllättävää, että taitotasolla A2 ja B1 olevat oppijat pitivät kyseistä ostosten tekemiseen liittyvää kielenkäyttötilannetta itselleen liian vaativana, vaikka väittäjä antaa puheen lisäksi mahdollisuuden näyttää sormella tai muulla tavoin, mitä haluaa tai tarkoittaa.

#### **Väittäjä 29**

(taso A2; korrelaatiokerroin  $-0.50$ ; Infit mean square 3.32; Infit Zstd 9.9; vastausjakauma 130–79)

Selviän lyhyistä sosiaalisista tilanteista, mutta pystyn harvoin pitämään keskustelua yllä.

Väittäjä 29 on tutkimuksen ainoa väittäjä, jonka korrelaatiokerroin on negatiivinen ja jonka kaksi muutakin tilastollista arvoa ovat yli suositeltujen rajojen. Väittäjä ei täten ole toiminut itsearviointilomakkeen osana. Väittäjä 29 on tyyppiesimerkki väittäjästä, joka koostuu kahdesta osasta, jotka lisäksi tuntuivat sulkevan toisensa pois. Negatiivinen korrelaatiokerroin kertoo siitä, että tämän väittämän ja muiden itsearviointilomakkeen väittämien variaatiot eivät ole samansuuntaisia ja että väittäjä 29 jostain syystä toimii eri tavalla kuin lomakkeen muut väittämät. Väittämän voidaankin todeta mitanneen jotain muuta kuin mitä tarkoitus oli ja jotain muuta kuin mitä lomakkeen muut väittämät mittasivat.

Tämän väittämän esittämät kaksi lausetta, joita erottaa toisistaan muttasana, voidaan todeta olevan sekä ristiriitaisia että vaikeatulkintaisia ja osallistujien kommenttien mukaan myös toisensa poissulkevia. Osallistujien vastaukset ja kommentit kertovat, että väittämän sisällön tulkinta on ollut mahdoton tehtävä. Jo oppijoiden valitsemia vastausvaihtoehtoja tarkasteltaessa huomataan väittämän toimineen huonosti, kun voidaan havaita, että suuri joukko pidemmälle kieliopinnoissaan edenneistä oppijoista on vastannut tähän A2-tasoiseen väittämään "ei", kun taas kieliopintojensa alkutaipaleella olevat oppijat ovat vastanneet väittämään "kyllä". Osaselityksenä oppijoiden ristiriitaisilta tuntuviin vastausvalintoihin voi olla se, että eri tasoilla olevat oppijat osallistuvat tyyppillisesti vaativuudeltaan jonkin verran erilaisiin sosiaalsiin tilanteisiin. Sitä eri kielitaidon tasoilla olevat oppijat miettivät tiettyjä omalle taitotasolleen tyyppillisiä sosiaalisia tilanteita ja valitsevat sen mukaan vastauksensa. Mitä ylempällä tasolla oppijan kielitaito on, sitä vaativampaa yleensä on samantasoisten sosiaalisten tilanteiden hallinta. Lisäksi ne oppijat, joiden kielitaito sijoittuu ylimmille tasoille, saattavat olla hyvinkin kriittisiä oman kielitaitonsa puutteiden suhteen. Yksi kommentteja antaneesta oppijoista totesi myös, että väittämän ilmaisemassa tilanteessa kyse ei ehkä olekaan pelkästään ranskan kielen hallinnasta, vaan pikemminkin sosiokulttuurisesta osaamisesta yleensä-

kin. Tämä väittämä sai osakseen osallistujilta eniten kommentteja (25), joista tässä esimerkkinä kolme:

*"Väittämä 29 oli sikäli ongelmallinen, että oli vaikea tietää, mitä pitäisi vastata jos sekä selviää yksinkertaisista tilanteista että pystyy pitämään keskustelua yllä – oletin, että siinäkin tapauksessa tulisi vastata 'kyllä'"(3).*

*"29: tähän oli vaikeampi vastata, koska vastaus 1. osaan on 'kyllä', mutta toiseen 'ei'" (5).*

*Hiukan vaikea kysymys (2 lausetta, joista toinen pitää paikkansa ja toinen ei) – laittaa-ko kyllä vai ei? Hiukan epäselvä kysymys. (3)*

Väittämässä 29 sekä perinteisen osioanalyysin että Rasch-analyysin arvot ja osallistujien valintojen ristiriitaisuus tuovat esille sen, että väittämä 29 on epäselvä ja sen vuoksi ei selvästikään kelvollinen mittainstrumentin osa. Todettakoon kuitenkin, että vaikka väittämä 29 ei toiminut itsearviointilomakkeen osana, kyseessä on vain yksi väittämä kokonaisuudessa, joka yleisellä tasolla toimi hyvin.

**Väittämä 41** (taso A1; korrelaatiokerroin .11; Infit mean square 1.20 ; Infit Zstd .5 vastausjakauma 207–2)

Osaan esittäytyä ja esitellä muita ihmisiä ja käyttää tavanomaisia tervehdys- ja hyvästelyilmauksia.

Väittämän 31 korrelaatiokerroin on hyvin matala (.11), mutta sen sijaan sen Infit mean square ja Infit Zstd -arvot sijoittuvat suositellulle välille. Lisäksi väittämä pysyy sille määritellyllä A1-tasolla alimmaisena (kuvio 5).

Myös tämä väittämä sisältää monia eri aktiviteetteja:

1. itsensä esittely
2. toisten esittely
3. tavanomaisten tervehdys- ja hyvästelyilmausten käyttö.

Väittämään vastaaminen dikotomisesti "kyllä" tai "ei" voi siten olla vaikeaa, jos osaa esimerkiksi esitellä itsensä, mutta tuntee, ettei välttämättä osaa esitellä muita. On lisäksi muistettava, että tutkimukseen osallistui kahden lukioluokan oppilaita, joilla ei nuoresta iästään johtuen välttämättä ollut kokemusta siitä, kuinka esitellä ihmisiä toisilleen. Kokemuksen puute väittämässä kuvatussa tilanteesta on voinut vaikeuttaa erityisesti nuorten vastaamista tähän väittämään. Muiden esittely vieraalla kielellä työkontekstiin liittyvässä pikemminkin virallisessa tilanteessa onkin toki jonkin verran kokemusta ja/ tai taitoa vaativa tehtävä. Oppijoilta tulleet kommentit keskittyivät siihen, että väittämän kuvaama tilanne ei ollut riittävän selvä. Tässä yksi esimerkki kommentista:

*"En kyllä kaikenlaisissa tilanteissa, mutta epävirallisissa tilanteissa, ystävällisten ihmisten parissa" (1).*

Väittämää 41 koskevat osittain samat kommentit, joita käsiteltiin väittämässä 29. Toisin sanoen tässäkin väittämässä on monia erilaisia elementtejä liitetty yhteen, mikä on tehnyt väittämään vastaamisesta vaikeaa.

### 5.1.1.5 Rasch-analyysin mukaiset huonosti toimineet väittämät

Siirryn nyt tarkastelemaan moderniin testiteoriaan pohjautuvan Rasch-analyysin mukaisia mahdollisesti ongelmallisia itsearviointilomakkeen väittämiä. Tarkastelussa keskityn erityisesti jo edellä mainittujen arvojensa puolesta selkeästi ongelmallisten väittämien analysointiin (väittämät 5, 23 ja 36).

Aloitan tarkastelun väittämästä 5:

#### **Väittämä 5**

(taso A2; korrelaatio .34; Infit Mnsq 1.68; Infit ZStd 5.5; vastausjakauma 146–63)

Osaan muotoilla kutsuja ja anteeksipyyntöjä ja vastata niihin.

A2-tasoinen väittämä 5 näyttää selvästikin olevan toiminnaltaan ongelmallinen. Myös tämä väittämä kuuluu niihin väittämiin, joita perinteinen osioanalyysi ei nosta esille, sillä se ylittää hiuksenhienosti yleisen hyväksyttävyyssrajan .30. Sen sijaan sen Infit MnSq (1.68) ja ZStd (5.5) -arvot osoittavat, että väittämässä on jotain ongelmaa. Molemmat edellä mainituista arvoista ovat liian korkealla. Osioiden vaikeustasoa kuvaavassa kuviossa (kuvio 5) väittämä 5 sijoittuu taitotasojen A2–B1 välimaastoon eli oppijat ovat pitäneet sitä hieman sille määrättyä taitotasoa A2 vaikeampana.

Väittämä 5 on saanut kommentteja viideltä oppijalta. Väittämän sisällön tarkka merkitys näyttää jälleen aiheuttaneen päänvaivaa oppijoille. Tässä eräs kommentti aloittelevalta ranskan kielen oppijalta, joka jopa tekee ehdotuksen siitä, miten väittämän saisi muokattua selkeämmäksi:

*"Kysymys 5: muotoilla kutsuja? Eikö löytyisi konkreettisempaa? (esimerkiksi Osaan kutsua ihmisiä kotiini)?" (1).*

Väittämien tarkempaa muotoilua toivovien oppijoiden kommentit olivat tämänkaltaisia väittämässä 5:

*"En täysin ymmärrä, mitä tarkoittaa 'osaan muotoilla kutsuja'; tarkoitetaanko virallisia kutsuja?" (4).*

*"Kysymykseen 5 olisin voinut vastata että 'jonkunlaisia'" (3).*

Eräs oppijoista tuntui olevan sitä mieltä, että väittämä ei kuuluisi puhumista koskevaan itsearviointilomakkeeseen:

*Väittämä 5 käsittelee pikemminkin kirjoitettua kuin puhuttua kieltä? (4)*

Voidaan siis todeta, että väittämä 5 kuuluu niihin lomakkeen väittämiin, joiden sisällön muotoilu olisi syytä tarkentaa.

Seuraavaksi tarkasteluun otetaan B1-tasoinen väittämä 23:

**Väittämä 23**

(taso B1; korrelaatio .39; Infit Mnsq 1.40, Infit ZStd 3.7; vastausjakauma 133–76)

Osaan ylläpitää keskustelua, mutta minua voi joskus olla vaikea ymmärtää, kun yritän sanoa tarkasti, mitä haluan.

Väittämää 23 kommentoitiin lähes yhtä monta kertaa (18) kuin edellistä väittämää. On myös syytä mainita, että tämän väittämän kommenttien yhteydessä viitattiin usein myös edellä käsiteltyyn väittämään 36.

Väittämän saamia tilastollisia tunnuslukuja tarkasteltaessa voidaan todeta, että kyseessä on väittämä, joka ei erotu perinteisessä osioanalyysissä siitä johtuen, että se ylittää niukasti hyväksyttävyyssrajan .30. Sen sijaan väittämän saamat Infit Mnsq ja ZStd -arvot ovat molemmat suositeltujen raja-arvojen ulkopuolella: Infit Mnsq on 1.40 (raja-arvo 1.30) ja ZStd puolestaan 3.7 (hyväksyttävyyssraja-arvot +2 ja -2:n välillä). Väittämien vaikeustaso kuvaavassa kuviossa (kuvio 5) sitä vastoin kyseinen väittämä pysyy hyvin sille määritellyllä taitotasolla B1. Väittämä 23 on siten tyypillinen esimerkki osiosta, jonka mahdollinen toimimattomuus olisi jäänyt huomaamatta, jos väittämistä ei olisi tehty Raschanalyysijä ja pyydetty oppijoiden kommentteja. Tätä väittämää koskevissa kommentteissa näkyi selvästi osallistujien epävarmuus siitä, olivatko he ymmärtäneet oikein väittämän sisällön tarkoituksen.

*Numero 23 on ristiriitainen. En siis osaa ylläpitää keskustelua ja kyllä: minua on hankala ymmärtää (2).*

*23: kumpaan kysymykseen tulisi vastata, jos toiseen olisi vastaus kyllä ja toiseen ei?(4).*

*Jäin miettimään, miten käsite "keskustelun ylläpitäminen" pitää ymmärtää. Miten paljon puhe on ylläpitoa, mikä ei vielä ole (3).*

Väittämä 23 koostuu kahdesta eri osasta, joten siihen vastaaminen voi olla vaikeaa, jos osallistuja tuntee, että väittämän toiseen puoleen vastaus olisi myönteinen ja toiseen kielteinen. Tarkastellaanpa väittämän vastaamista kuvitteellisen kielitaidoltaan taitotasosteikon alapäähän (tasot A1 ja A2) sijoittuvan henkilön silmin: kumpi vastausvaihtoehdoista tulisi valita, jos henkilö kokee, ettei hän osaa ylläpitää keskustelua (*osaan ylläpitää keskustelua* = ei), ja toisaalta, että häntä on vaikea ymmärtää (= kyllä)? Väittämän 23 sisällön epäselvä muotoilu tekee siitä monissa tapauksissa mahdollottoman vastattavan.

Lisäksi voidaan pohtia sitä, pystyvätkö kielenkäyttäjät aina tulkitsemaan keskustelukumppaninsa käyttäytymisestä esimerkiksi sen, että nämä pitävät ei-



natiivin puhetta vaikeana ymmärtää. Olisiko väittämässä paikallaan keskittyä ilmaisemaan vain väittämiin vastaavan henkilön näkökulma eikä yrittää tulkita keskustelukumppanin käyttäytymistä? Jos keskustelukumppanin käyttäytymisen halutaan tuoda esille väittämässä, se voitaisiin tehdä esimerkiksi siten, että väittämässä ilmaistaisiin selkeästi, että kyseessä on väittämään vastaajan tulkinta keskustelukumppanin käyttäytymisestä.

Seuraavaksi tarkastellaan väittämää 36:

#### **Väittämä 36**

(taso B2; korrelaatio .52; Infit Mnsq 1.33, Infit ZStd 2.6; vastausjakauma 61–148)

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

Väittämän korrelaatio on perinteisen osioanalyysin näkökulmasta katsoen täysin kunnossa (.52), mutta molemmat modernia testiteoriaa edustavat arvot ovat kohonneita: Infit mean square 1.33 ja Infit Zstd 2.6. Sen sijaan väittämän vaikeustaso on oppijoiden mielestä ollut täysin kohdallaan eli taitotasolla B2 (kuvio 5). Väittämä 36 keräsi yhteensä 19 kommenttia oppijoilta. Kommenteista kävi selkeästi ilmi, että oppijat eivät olleet ymmärtäneet väittämän sisältöä eivätkä siten olleet lainkaan varmoja omasta vastausvalinnastaan. Kyseisen väittämän keräämät kommentit voidaan jakaa kahteen ryhmään sen mukaan, oliko kommentoija sijoittanut oman itsearviointinsa tason taitotasoasteikon ala- vai yläpäähän. Taitotasoasteikon yläpäähän itsearviointinsa sijoittaneille tyypillistä oli se, että he viittasivat kommentissaan todelliseen kielenkäyttötilanteeseen, jossa he olivat olleet ranskankielisten henkilöiden kanssa. Kommenteista löytyy mielenkiintoisia tulkintoja ja asenteita ranskan kielen natiivipuhujien käyttäytymisestä sellaista ei-natiivia 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).*

*Aika usein olen törmännyt sellaisiin tilanteisiin, joista henkilökohtaisesti päätellen, ettei tuollaista voi olla olemassa, vaan aina natiivi joutuu jotenkin muuttamaan käyttäytymistään non-natiivin kanssa (4).*

Edellä esiteltyjen kommenttien vastaisiakin kommentteja toki löytyi. Niistä esimerkkinä on seuraava ranskan kielen natiivien suhtautumista ulkomaalaiseen kielenpuhujaan pohtiva kommentti:

*Väittämä 36 on minusta jotenkin vaikea, koska syntyperäiset ovat yleensä niin fiksuja, etteivät näytä tunteitaan (ainakaan ärsytystä). Paitsi se pariisilainen tarjoilija, joka ker-  
ran vähän simputti... (4).*

Eräs tutkimukseen osallistuneista oppijoista puolestaan pohti keskustelukump-  
panin käyttäytymistä seuraavasta 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 vaikea-  
selkoisuutta 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, tuntu-  
vat saavan paljon palautetta erityisesti ylimmillä taitotasolla olevilta henkilöil-  
tä, joilla on jo kokemusta natiivien kanssa keskustelemisestä. Suurin osa tähän  
väittämään liittyvistä kommentteista saatiinkin tasolle 4 oman kielitaitonsa arvi-  
oineilta oppijoilta. Nämä henkilöt ovat varmasti oman ranskan kielen taitonsa  
kanssa jo sen verran edistyneitä, että pystyvät viittaamaan omakohtaisiin ko-  
kemuksiin natiivipuhujien kanssa. Lisäksi tällä taitotasolla olevat henkilöt pys-  
tyvät ehkä aistimaan helpommin myös natiivien kielelliseen käyttäytymiseen ja  
kommunikaation sujumiseen liittyviä asenteita. Hieman yllättäviä ehkä ovat  
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 oppi-  
jan suullisen kielitaidon taso ei voisi ylittää samalle tasolle kuin jolla kyseisen  
kielen natiivipuhujan kielitaito on. Sen vuoksi osa oppijoista kokee, että vieraan  
kielen oppija ei pysty kielellisesti olemaan täysin yhdenvertainen natiivipuhu-  
jan kanssa suullisessa kommunikaatiotilanteessa.

Todettakoon lisäksi, että väittämää 36 ei kommentoinut yksikään oppija  
taitotasoasteikon ääripäistä eli tasolta A1 tai tasolta C2. Tämän pohjalta voidaan  
esittää ainakin seuraavanlainen kysymys: Eivätkö väittämässä esitellyt asiat  
koske täysin aloittelevia kielenoppijoita tai erinomaisen suullisen ranskan kie-  
len taidon omaavia? Onkin varsin selvää, että aloittelevat ranskan kielen puhu-  
jat eivät pysty tunnistamaan sitä, toimiiko ranskan kielen natiivipuhuja heitä  
kohtaan samalla tavalla kuin toista ranskan kielen natiivia kohtaan. Erinomai-  
sen ranskan kielen taidon omaavat ei-natiivipuhujat taas mitä luultavimmin  
kokevat itsensä sen verran varmoiksi kommunikaatiotilanteissa natiivipuhujan  
kanssa ollessaan, etteivät he tunne itseään millään tavalla vaivaantuneiksi tä-  
mäntilanteissa.

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.

Tämän jälkeen tarkastellaan väittämää 13.

### **Väittämä 13**

(taso C1; korrelaatio .51; Infit Mnsq .70; Infit ZStd -1.8; vastausjakauma 21–188)

Osaan ilmaista ajatuksiani ja mielipiteitäni selvästi ja tarkasti sekä esittää monimutkaisia perusteluja vakuuttavasti.

Huomattakoon aluksi, että vaikka väittämä 13 on merkitty huonosti toimineiden väittämien joukkoon edellä olevassa taulukossa, väittämää ei ole sen saamien arvojen pohjalta syytä pitää yhtä huonosti toimivana kuin edellä esiteltyjä väittämiä. Väittämä 13 sijoittuu taitotasoasteikon yläpäähän tasolle C1. Väittämän 13 Infit mean square -arvo on suositeltavan raja-arvon rajalla (.70), sen Infit Zstd -arvo on hyväksyttävä (-1.8) ja sen korrelaatiokerroin on .51 eli ihan hyvä. Winstep-analyysin mukaan väittämä 13 pysyy hyvin tasolla C1 eli oppijat ovat pitäneet sitä tämääntasoista osaamista edustavana väittämänä (kuvio 5).

Väittämä on kerännyt 6 kommenttia ylimpien tasojen oppijoilta. Kaikki tätä väittämää kommentoineet ovat itsearvioineet oman kielitaitonsa taitotasolle B2 tai sitä ylempäs taitotasoasteikolla. Väittämässä 13 on kysymyksessä melko vaativa toiminta jopa äidinkielellä. Lisäksi oppijoiden kommentteista käy ilmi, että tässäkin väittämässä osa oppijoista olisi toivonut väittämässä käytettyjen käsitteiden tarkempaa selkeyttämistä. Tässä muutamia oppijoiden kommentteja:

*”Väittämä 13 oli hankala. Vaikea arvioida, millaiset perustelut ovat monimutkaisia” (4).*

*”Ne kysymykset, joissa kysyttiin erikoistaitoja tai syvempää kielitaitoa, olivat hankalampia, koska niissä vastaus ei ollut itsestään selvä. Hankalimpia väittämiä: 3, 6, 13, 18, 24, 29” (5).*

*”Väittämät 13 ja 38 tuntuivat sikäli hankalilta, etten ole varma kyseisistä taidoistani edes puhuessani äidinkielelläni” (4).*

Eräs oppijoista kommentoi väittämää myös siitä näkökulmasta, ettei väittämässä olisikaan kyse vain kieli- vaan muusta taidosta:

*”Ongelmallista oli, jos väittämä mittasi pikemminkin viestintätaitoa kuin kielitaitoa esim. itsensä ilmaiseminen selkeästi, vakuuttavasti ja tarkasti” (5).*

Väittämän 13 voidaan todeta aiheuttaneen kommentteja ja pohdintaa selvästi ylimpien tasojen oppijoiden keskuudessa. Muille tähän C1-tasoiseen melko vaa-

tivaan väittämään vastaaminen on mitä luultavimmin ollut mutkatonta (= väittäjä ei kuvaa minun osaamistani eli vastaus on ollut "ei").

Loput huonosti toimineisiin osioihin Rasch-analyysin perusteella luokitelluista väittämistä ovat osioita, joiden Infit mean square -arvo on hyväksyttävissä rajoissa, mutta joiden Infit Zstd -arvo ei pysy täysin hyväksyttävissä rajoissa. Alla olevista väittämistä puhuttaessa on kuitenkin syytä korostaa, että näiden väittämien esille ottaminen huonosti toimineiden väittämien joukossa on perusteltua tutkimukseen liittyvän yksityiskohtaisen analyysin näkökulmasta. Samalla on kuitenkin syytä korostaa, että Rasch-analyysin käyttöoppaan mukaan perussääntö Winstep-analyyseissä on se, että jos mean squares -arvot (Infit ja Outfit) ovat hyväksyttäviä, Zstd -arvo voidaan jättää huomiotta<sup>18</sup>. Kahden (väittämät 31 ja 33) Infit Zstd -arvo on hyväksyttävän rajalla ja kolmen muun väittämän (22, 25 ja 37) vastaava arvo on -2.4 (väittämät 22 ja 25) tai -2.6 (väittäjä 37). Tässä kohdassa ei käsitellä uudelleen väittämää 19, joka siis osoittautui huonosti toimivaksi molemmissa osioanalyyseissä ja jota on käsitelty edellä perinteisen osioanalyysin huonosti toimineiden väittämien yhteydessä.

Alla olevaan taulukkoon on listattu paperilla tehdyn itsearviointin mahdollisesti ongelmallisia osioita sekä perinteisen osioanalyysin että moderniin testiteoriaan perustuvan osioanalyysin perusteella. Vertailuarvoina on käytetty korrelaatiokerrointa (perinteinen osioanalyysi) ja Infit mean square -arvoa (moderniin testiteoriaan perustuva Rasch-analyysi).

TAULUKKO 9 Mahdollisesti ongelmalliset väittämät perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta (paperilla tehty itsearviointi).

ONGELMALLISET OSIOT				
	Perinteinen osioanalyysi		Moderniin testiteoriaan perustuva Rasch-analyysi	
	Väittäjä	Osio-testi - korrelaatio	Väittäjä	Infit Mnsq-arvo
Paperilla tehty itsearviointi	41	.11	13	.70
	19	.12	36	1.33
	9	.18	23	1.40
	27	.19	5	1.68
	21	.21	19	1.30
	1	.22		

Kuten taulukosta selviää, osioanalyysit nostavat esiin eri osioita, joiden toimivuudessa saattaa olla jonkinlaisia ongelmia. Osio 19 näyttäisi kuitenkin molempien osioanalyysien tulosten pohjalta olevan jollakin tapaa ongelmallinen toimivuudeltaan, sillä sen osio-testi-korrelaatio on hyvin alhainen (.12) ja lisäksi sen Infit mean square -arvo nousee epäilyttävän osion rajalle (1.30).

<sup>18</sup> ([www.winsteps.com/winman/index.htm?diagnosingmisfit.htm](http://www.winsteps.com/winman/index.htm?diagnosingmisfit.htm))

Mitä tulee lomakkeen kuuteen perinteisen osioanalyysin esille tuomaan huonosti toimineeseen väittämään, joiden korrelaatioarvo oli alhainen, voidaan todeta, että ne edustavat kielitaitoasteikon ylä- tai alapäähän sijoittuvia väittämiä (mittarista on poistettu negatiivisen korrelaatioarvon omaava väittäjä 29). Näistä väittämistä neljä sijoittuu tasolle A1, yksi tasolle A2 ja yksi taitotason yläpäähän tasolle C2. On myös huomattava, että syy asteikon ääripäihin sijoittuvien osioiden toimimattomuuteen reliabiliteettianalyyseissä on usein yksinkertaisesti tekninen tai laskennallinen. Kun valtaosa tutkimukseen osallistujista vastaa väittämiin joko ”kyllä” tai ”ei”, niin vastauksiin ei saada paljoakaan hajontaa ja siten kyseeseen tulee variaation vähyyys vastauksissa. Variaation vähyyys puolestaan aiheuttaa sen, että huonosti toimineen osion ja muiden testiosioiden välillä ei ole yhteisvaihtelua eli korrelaatiota. Hajonnan vähyyys vastauksissa saattaa siten hyvin yksinkin selittää tiettyjen osioiden heikon toimimisen, tai ainakin sen voidaan todeta olleen mahdollisesti vahva osasy kyseisten osioiden toimimattomuuteen.

Rasch-analyysin esille tuomat kolme selkeästi huonosti toiminutta väittämää (5, 23 ja 36) sen sijaan eivät edusta taitotasoasteikon ääripäitä. Väittäjä 5 on tasolta A2, väittäjä 23 tasolta B1 ja väittäjä 36 taitotasolta B2. Muutkin mainitut väittämät, joiden Rasch-analyyseissä saamat arvot olivat joko hyväksyttävän rajalla tai joista vain Infit Zstd -arvo oli hyväksyttävien raja-arvojen ulkopuolella, sijoittuvat tasaisesti taitotasoasteikolle tasojen A2 ja C1 sisälle.

Yhteenvetona paperilla tehdyn itsearviointilomakkeen väittämistä voidaan todeta, että suurin osa lomakkeen hyvin toimineista väittämistä täytti Viitekehysten suositukset, joiden mukaan hyvin muotoillut väittämät ovat muun muassa täsmällisiä, selkeitä ja lyhyitä (EVK 2004, 280–281). On lisäksi syytä ottaa huomioon, että osioiden toimimattomuus on aina jonkin verran suhteellinen asia. Jos tutkimuksessa olisi ollut muutama aivan alatasen kielitaitoa edustava oppija enemmän vastaamassa alatasen väittämiin ”ei”, niin kyseisten väittämien korrelaatiot olisivat olleet parempia. Onkin mahdotonta suoralta kädeltä väittää, että ne väittämät, jotka näyttivät toimivan huonosti mittarissa, eivät voisi toimia normaalisti, jos niitä käytettäisiin sellaisilla oppijoilla, jotka ovat todellisia aloittelijoita. Toisaalta on kuitenkin todettava, että huonosti toiminut väittäjä 29 oli sisällöltään huono eikä sen toimimattomuus siten selity edellä selostetun asteikon ääripäihin liittyvän vaihtelun vähyydellä.

#### 5.1.1.6 Väittämien vaikeustason tarkastelua

Lopuksi tarkastellaan vielä paperilla tehdyssä itsearvioinnissa käytetyn mittarin eli itsearviointilomakkeen itsearviointiväittämien ”vaikeutta”. Väittämien vaikeus on laskettu oppijoiden vastausten perusteella. Kuvio 5 esittää, miten itsearviointiväittämät asettautuivat Winstep-analyysin perusteella ohjelman laatimalle välimatka-asteikolliselle mitta-asteikolle, joka kuviossa ulottuu miinus seitsemästä plus kahdeksaan. Kuvion perusteella saa nopeasti käsityksen siitä, missä määrin aikaisemmassa DIALANG-hankkeessa ja muissa vastaavissa tutkimuksissa Viitekehysten eri tasoille asetellut väittämät käyttäytyivät odotuksenmukaisesti tässä tutkimuksessa: olivatko esimerkiksi kaikki B2-tason

väittämät vaikeampia kuin B1-tason väittämät, kuten voisi olettaa? Kuviosta 5 voi myös tehdä päätelmiä oppijoiden taitotasosta suhteessa väittämiä vaatimustasoon, koska Winstepin Rasch-analyysin luoma mitta-asteikko on yhteinen sekä oppijoille että testin osioille (tässä tapauksessa väittämille).

```

Person - MAP - Item
<more>|<rare>
8      . | t21_C2
      |T
7      +
      .# |
6      + t34_C1
      .# T|
      ## | t28_C2
5      + t15_C2
      .## | t13_C1
      .# | t6_C2
4      .# +
      .## |S t16_C1 t18_C2 t3_C1
      ## |
      S|
3      .##### + t24_C1 t38_C2
      ### | t33_C1
      .### |
      t35_B2 t36_B2
2      .##### +
      .## |
      ### | t40_B1 t8_B2
      .## |
1      .### +
      ## M| t2_B2 t42_B2
      .##### | t37_B2
0      .### +M
      .##### |
      .#### | t17_B1
      ## | t22_B1 t23_B1 t7_B1
-1     . +
      ##### | t25_B1 t26_B1
      ##### | t39_B1 t5_A2
      .### | t11_A2 t4_B1
-2     .## S+
      ### |
      | t31_A2
      ## |
-3     .### +
      . | t32_A1
      | t14_A2
      ## |S
-4     ## +
      | t10_A2 t12_A1 t30_A1
      . |
      T|
-5     . +
      # | t9_A1
      | t1_A2 t27_A1
-6     . +
      | t19_A1
      | t41_A1
-7     +
      <less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

```

KUVIO 5 Paperilla tehty itsearviointi: väittämien vaikeus oppijoiden vastausten perusteella.

Kuvio 5 havainnollistaa hyvin sitä, kuinka lomakkeen väittämät sijoittuivat välimatka-asteikolla toisaalta eri tasoille sijoittuviin väittämiin nähden ja toisaalta oman taitotasonsa väittämien kesken. Kuvioista 5 selviää, että oppijat eivät kokeneet tiettyjen väittämien kuuluvan sille taitotasolle, jolle ne oli määritelty kuuluviksi. Esimerkkinä väittämistä, joita oppijat pitivät vaikeampina kuin niiden määritelty taitotaso on, ovat väittämät 12 (A1), 32 (A1) ja 40 (B1), joita käsitelen alla:

#### **Väittämä 12**

(taso A1; korrelaatiokerroin .40; Infit mean square .85; vastausjakauma 195–14 )

Osaan kysyä itseeni liittyviä ja toisia ihmisiä koskevia kysymyksiä, jotka liittyvät asuinpaikkaani, tuntemiini ihmisiin ja tavaroihini. Osaan myös vastata tällaisiin kysymyksiin.

#### **Väittämä 32**

(taso A1; korrelaatiokerroin .49; Infit mean square .87; vastausjakauma 184–25)

Osaan esittää yksinkertaisia kysymyksiä ja vastata niihin, esittää yksinkertaisia väittämiä ja ottaa niihin kantaa, kun kyseessä ovat jokapäiväiset asiat.

Molemmat yllä olevat A1-tasoiset väittämät sijoittuvat kuviossa 5 ylemmäksi kuin muut A1-tasoiset väittämät; väittämä 12 näyttää sijoittuvan suunnilleen A1- ja A2 -tasojen välimaastoon ja väittämä 32 (A1), jota oppijat ovat pitäneet vieläkin vaikeampana kuin väittämää 12, sijoittuu selvästi A2-tason yläpäähän. Väittämien muotoilua ja sisältöä analysoidessa voidaan havaita, että kyseiset väittämät sisältävät seuraavia yhteisiä elementtejä:

#### **Väittämät 12 ja 32**

- liittyvät kysymiseen ja kysymyksiin vastaamiseen
- sisältävät viittauksia aloittelevalle kielen oppijalle tuttuihin ja yleensä helppoina pidettyihin asioihin (esim. oppija itse, asuinpaikka, omat tavarat, jokapäiväiset asiat)
- kuuluvat molemmat itsearviointilomakkeen pisimpiin väittämiin, joiden sisältö on melko runsas.

Lisäksi väittämä 32:n sisältö on jonkin verran tulkinnanvarainen: mitä itse asiassa tarkoitetaan yksinkertaisilla kysymyksillä, joita esitetään ja joihin vastataan tai yksinkertaisilla väittämillä, joita esitetään tai joihin otetaan kantaa? Miten yksinkertaisuus tulisi ymmärtää tässä yhteydessä? Väittämä 32 on ollut mukana myös jo edellä mainitussa Euroopan komission kielitaitoindikaattoritutkimuksessa. Tutkimuksen loppuraportista käy ilmi, että väittämä 32 sijoittuu puheen ymmärtämistä kuvaavan itsearviointiväittämän rinnalla huomattavasti ylemmäksi vaikeustasoa kuvaavalla janalla kuin luetun ymmärtämisen ja kirjoittamisen A1-tasoiset väittämät (First European Survey on Language Compe-



tences. Final Report 2012, 29). Väittämä 32 sijoittuu kielitaitoindikaattoritutkimuksessa hyvin lähelle A2-tasoisia väittämiä. Euroopan Komission tekemän ja käsillä olevan tutkimuksen mukaan väittämä 32 vaikuttaisikin siten edustavan vaikeustasoltaan pikemminkin A2- kuin A1-tasoista kielen osaamista.

Myös väittämä 40 on oppijoiden vastausten perusteella vaikeampi kuin sen määritelty taso B1:

**Väittämä 40**

(taso B1 ; korrelaatiokerroin .70; Infit mean square .82; vastausjakauma 77–131)

Osaan kertoa melko yksityiskohtaisesti kokemuksistani ja kuvata tunteita ja reaktioita.

Väittämä 40 on ollut esillä jo edempänä, sillä se edustaa väittämää, joka on toiminut hyvin sekä perinteisen että Rasch-analyysin tunnuslukujen valossa. Tässä väittämässä oppijoilla näyttää selvästikin olleen käsitys, että tunteiden ja reaktioiden kuvaaminen sekä kokemusten melko yksityiskohtainen kertominen vaatii ylemmälle kuin B1-tasolle sijoittuvaa kielitaitoa. Väittämä 40 sijoittuikin kuviossa 5 aivan selkeästi tasolle B2 lähelle C1-tasoisia väittämiä. Tämä väittämä kuuluu edellä tarkastellun lomakkeen hyvin toimineiden väittämien joukkoon. Tästä huolimatta voidaan todeta, että kun osioita tarkastellaan eri näkökulmista (tässä tarkastelun lähtökohtana on oppijoiden kokema osioiden vaikeustaso), saadaan viitteitä osioiden mahdollisista ongelmista ainakin suhteessa niiden aiottuun vaikeustasoon tai verrattuna aikaisemmissa tutkimuksissa, erilaisilla kielenoppijoilla tehdyissä tutkimuksissa havaittuun vaikeustasoon. Kun yhdistetään tilastoanalyysin tulos väittämän sijoittumisesta asteikolle oppijoiden kommentteihin, voidaan näin ollen paljastaa näissäkin osioissa joitakin tulkintaongelmia, vaikka jotkut niistä käyttäytyivätkin hyvin sekä perinteisissä että moderneissa osionaalyseissa.

Kuvion 5 kaksi ylimmäiseksi sijoittuvaa väittämää ovat väittämät 21 (C2) ja 34 (C1):

**Väittämä 21**

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

Puhun ranskan kieltä melkein yhtä hyvin kuin äidinkieltäni.

Väittämä 21 erottuu sijaintinsa vuoksi asteikolla kaikista muista väittämistä siinä suhteessa, että sitä on pidetty hyvin vaativana, ja se sijoittuikin siten hyvin kauas seuraavasta sitä alemmaksi asteikolle sijoittuvasta väittämästä 34. Väittämä 21 liittyy tärkeään seikkaan oppijoiden asenteiden ja mielikuvien kannalta eli kysymykseen siitä, voiko vierasta kieltä koskaan oppia samantasoisesti kuin omaa äidinkieltään. Tässä yhteydessä kannattaa muistaa, että vieraan kielen ja äidinkielen samantasoiseen oppimiseen saatiin oppijoiden vapaita kommentteja erityisesti niiltä oppijoilta, joiden kielitaidon taso sijoittui melko korkealle taito-

tasoasteikossa. Muistettakoon myös, että väittämä 21 kuului edellä tarkasteltujen huonosti toimineiden väittämien joukkoon perinteisen osioanalyysin mukaan.

Kuviota 5 tarkastellessa voidaan todeta, että C1- ja C2-tasolle sijoittuvat väittämät eivät näytä pysyvän niille määritellyllä tasolla, vaan sekoittuvat keskenään. Oppijat ovatkin pitäneet monia näille tasolle sijoittuvia väittämiä toiselle tasolle kuuluvina. Otettakoon esimerkiksi tasoaan vaikeammaksi mielletty väittäjä numero 34:

#### **Väittäjä 34**

(taso C1 ; korrelaatiokerroin .31; Infit mean square .93; vastausjakauma 10–199)

Pystyn selviytymään helposti väittelystä, vaikka siinä käsiteltäisiin abstrakteja ja vaikeita asioita, joita en tunne.

Tarkemmin analysoitaessa väittäjä 34 vaikuttaakin sisältönsä vaativuuden puolesta selkeästi kuuluvan pikemminkin taitotasolle C2 kuin C1. On suhteellisen hankala kuvitella (suomalaisen) kielenkäyttäjän kannalta vaativampaa suullista kielenkäyttötilannetta kuin väittäjän kuvaama tilanne, jossa kielenkäyttäjä selviytyy vieraalla kielellä *helposti* niinkin haasteellisesta kielenkäyttötilanteesta kuin mitä väittelyä sisältävä tilanne on. Väittäjän kuvaaman kielenkäyttötilanteen vaativuus korostuu entisestään, kun edellä mainittujen seikkojen lisäksi väittämässä viitataan abstrakteihin ja vaikeisiin asioihin, joita kielenkäyttäjä ei tunne.

C2-tasoiset väittämät 6, 18 ja 38 puolestaan ovat esimerkkejä väittämistä, joita tutkimukseen osallistujat pitivät helpompina kuin C2-tasolle kuuluvina väittäminä. Mikään näistä väittämistä ei näyttänyt toimivan huonosti, kun niitä analysoitiin tilastollisin keinoin, mutta sen sijaan kahdesta näistä väittämistä (6 ja 18) on oppijoiden kommentteja, jotka saattavat valaista asiaa.

#### **Väittäjä 6**

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

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

Väittäjä 6 kommentoitiin 12 kertaa. Kaikista kommentteista käy esille oppijoiden hämmennys heidän mielestään epäselvästi muotoillun väittäjän sisällön tulkinnasta. Tässä muutama esimerkki saaduista kommentteista:

*6: vaikeisiin asioihin kielellisesti, vai muuten? Miksi palaisin asiaan? Miten niin toisten huomaamatta? (3).*

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

*En ymmärtänyt tilannetta (4).*

*Kohta 6 oli vaikeasti ymmärrettävää. 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 täysin oikeutetusti kysyä, mitä itse asiassa tarkoittavat väittämässä mainitut vaikeat kohdat, joihin palataan. Koska väittämän muotoilu on selvästikin tulkinnanvarainen, jokainen tutkimukseen osallistuja on tulkinnut väittämän sisältöä omalla tavallaan. Lisäksi on mainittava, että erityisen huolestuttavaa väittämän ymmärrettävyyden ja toimivuuden kannalta on se, että edes kielitaitonsa taitotasosteikon ylimmille tasoille (tasot 4 ja 5) arvioineet oppijat eivät ole välttämättä ymmärtäneet puheena olevan väittämän sisältöä. On siten selvää, että väittämä 6 ei täytä hyvin toimivan väittämän kriteereitä ainakaan selkeyden ja yksiselitteisyyden osalta.

Väittämä 18 keräsi niin ikään kommentteja kahdeltatoista oppijalta. Tarkastelemme alla myös näitä kommentteja.

#### **Väittämä 18**

(taso C2 ; korrelaatiokerroin .52; Infit mean square 1.06; vastausjakauma 34–175)

Puhun ranskaa sujuvasti ja selkeästi niin että kieli ei aiheuta mitään ongelmia sosiaaliseen ja henkilökohtaiseen elämäni.

Tutkimukseen osallistuneille oppijoille korostettiin ennen lomakkeen väittämiin vastaamista, että heidän tulisi tarvittaessa väittämiin vastatessaan kuvitella itsensä ranskankieliseen ympäristöön. Silti monet kommentoijista tuntuvat läheneen liikkeelle senhetkisestä todellisesta tilanteestaan ja toimintaympäristöstään eli suomenkielisestä ympäristöstä. Väittämä 18 edustaa kuitenkin selkeästi kielenkäyttötilannetta, jossa kieltä käytetään ranskankielisessä ympäristössä. Jos oppija siis vastasi tähän väittämään suomenkielisestä kielenkäyttöympäristöstä käsin, väittämän sisältö saattoi hänestä tuntua hieman oudolta. Vaikuttaa myös siltä, että joissakin tapauksissa oppijat eivät näytä kokeneen väittämien sisältöä relevanteiksi vieraan kielen hallinnan kannalta. Tässä muutama esimerkki kommentteista:

*18: en oikein ymmärtänyt miten sosiaalinen ja henkilökohtainen elämäni voisivat muuttua jonkin kielen vuoksi. Kysymysmuoto oli hieman kummallinen (2).*

*Tosielämässä ei olisi mitään ongelmia, vaikken koskaan sanoisi ranskan sanaa...(3).*

Kommenttien joukossa oli myös kommentteja, joissa pohdittiin sekä väittämän sisällön tarkoitusta että joidenkin väittämässä käytettyjen sanojen merkitystä:

*18: miten määritellään 'sujuva'? Toisaalta vaikka ei osaisikaan puhua sujuvasti, niin sen ei tarvitse aiheuttaa ongelmia sosiaaliseen ja henkilökohtaiseen elämäni (4).*

18: kieli ja kulttuuri ovat niin sidoksissa, että tähän oli vaikea vastata – tein sen ajatellen, että kieli = kulttuuri (5).

18: sosiaalisen elämän ongelmat? Mitä oikeastaan tarkoitetaan? (4).

Oppijoiden kommentteja tarkasteltaessa tuleekin väistämättä mieleen kysymys, pitäisikö kirjallisten itsearviointiväittämien yhteydessä määritellä tarkemmin väittämän kuvaaman kielenkäyttötilanteen lokalisointiin liittyvä ympäristö tai konteksti. Kun osa oppijoista vastaa väittämään sijoittaen kielellisen toiminnan ranskankieliseen ympäristöön ja toinen osa taas suomenkieliseen todelliseen ympäristöönsä viitaten, on selvää, että saadut vastaukset ovat hyvin heterogeenisiä. Oppijat tulkitsevat siis todistettavasti tätä väittämää eri lähtökohdista, mikä osittain vähentää tämän väittämän toimivuutta ja luotettavuutta.

Itsearviointilomakkeen kolmas C2-tasoinen väittäjä, johon edellä viitattiin ja jota oppijat ovat pitäneet taitotasoaan helpompana, on väittäjä 38. Väittäjä 38 sijoittuu kolmesta mainitusta väittäjästä alimmaksi taitotasoasteikolla, jolla se sijoittuu niinkin alas kuin lähelle B2/C1-tasojen rajaa. Tästä väittäjästä ei valitettavasti löydy kuitenkaan oppijoiden vapaita kommentteja.

#### **Väittäjä 38**

(taso C2 ; korrelaatiokerroin .54; Infit mean square 1.13; vastausjakauma 48-161)

Osaan tuottaa selkeän, sujuvan ja loogisesti rakentuneen puheen.

### **5.1.1.7 Yhteenvedoa paperilla tehdyn itsearviointimittarin luotettavuudesta**

Yhteenvedona paperilla tehdyssä itsearvioinnissa käytetystä lomakkeesta voidaan todeta, että lomake näyttää olevan riittävän luotettava, jotta sen pohjalta voidaan odottaa saatavan luotettavia tutkimustuloksia, kunhan väittämät 29 (sisällöllisiä ongelmia) ja 20 (ei vaihtelua) poistetaan. Lomakkeen laatu ilmenee myös osio-testi-korrelaatioita ja Winstepin Infit mean square-arvoja tarkasteltaessa. Osio-testi-korrelaatioiden tasoista 60 % sijoittui arvojen .51–.73 välille ja 25 % arvojen .31–.40 välille. Siten yhteensä 85 % paperilla tehdyn itsearviointilomakkeen väittämistä ylitti arvon .30, jota yleisesti pidetään osioiden hyväksyttävien korrelaatioarvojen alarajana. Alle rajan .30 jäi siten 40:stä väittäjästä 6 väittäjää (15 %), joista enemmistö oli alimman A1-tason väittäjiä.

Winstepin Infit mean square-arvoista valtaosa eli noin 92 % oli hyväksyttävissä rajoissa eli .7 ja .13 välillä. Vain kolmessa näytteessä oli epäilystä niiden sopivuudesta mittarin taustalla olevaan tilastolliseen malliin.

Paperilla suoritettussa itsearvioinnissa käytetyn lomakkeen laadun arviointi paljasti monta mielenkiintoista seikkaa. Ensimmäinen havainto oli se, että perinteiseen osioanalyysiin ja moderniin testiteoriaan perustuvat osioanalyysit nostivat osittain esille eri asioita (väittäjiä). Täten tässä tutkimuksessa tuli käytännössä testattua ja todeksi havaittua se, että nämä kaksi analyysiä todellakin

tuovat esille erilaisen näkökannan (*different "ways of seeing"*) tutkittavaan ilmiöön (Gyagenda & Engelhard 2009, 242). Samantyyppisiä huomioita esittivät Lynch & Namara (1998, 176–177) maahanmuuttajien suullista kielitaitoa koskevassa tutkimuksessaan, jonka aineiston analysoinnissa käytettiin GENOVA ja Facets -analyysijä. Heidän mukaansa Facets tarjoaa yksityiskohtaisempaa tietoa kuin GENOVA. Lisäksi Facets tarkastelee aineistoa mikroskooppisen tarkasti paljastaen aineistosta mahdolliset pienetkin "viat". Tutkijat totesivat kuitenkin, että mainitut analyysit mahdollistavat tasapainoisen ja perusteellisen tulkinnan tarkasteltavasta aineistosta, kun niitä käytetään yhdessä.

Tutkimukseen osallistuneiden oppijoiden laadulliset kommentit valaisivat omalta osaltaan sitä puolta, että jokin lomakkeessa käytetty kirjallinen itsearviointiväittäjä ei välttämättä missään vaiheessa nouse esille tilastollisissa analyysissä, vaikka lomaketta itsearviointiin käyttäneet oppijat ovatkin selvästi ilmaisseet jonkin väittämän olleen vaikeatulkintainen tai sisällöltään moniselitteinen. Eri menetelmät tarkastelevat väittämiä eri näkökulmista ja antavat siten toisiaan täydentävää hyödyllistä tietoa väittämien toimivuudesta sekä käytetyn itsearviointimittarin luotettavuudesta. Menetelmien tuottaman tiedon eroista voidaan kuitenkin vetää joitakin yleisempiäkin johtopäätöksiä, joita käsitellään tarkemmin luvussa 6.

### 5.1.2 Tehtävä 1: kuunneltaviin puhenäytteisiin perustuva itsearviointi

Kuunneltaviin näytteisiin perustuvan itsearviointilomakkeen luotettavuus analysoitiin samaan tapaan tilastollisesti kuin edellä käsitelty paperilla tehdyssä itsearvioinnissa käytetty lomake. Ero näissä kahdessa itsearviointityypissä on siis se, että paperilla tehdyssä itsearvioinnissa osallistujilla ei ollut auditiivista tukea itsearvioinnissaan, vaan itsearviointi tehtiin pelkästään kirjallisten itsearviointiväittämien pohjalta, kun taas kuunneltaviin puhenäytteisiin perustuvassa itsearvioinnissa osallistujilla oli kuunneltavanaan tietty määrä puhenäytteitä molemmista tutkimuksessa käytetyistä tehtävistä. Muistutettakoon vielä, että tehtävää 1 koskevia kuunneltavia puhenäytteitä oli 26 ja tutkimukseen osallistuvan oppijan tehtävänä oli valita, selviytyikö hän omasta mielestään paremmin tehtävästä kuin nauhalta kuullun näytteen puhuja (= kyllä) vai oliko oppija sitä mieltä, että hän itse ei selviytynyt tehtävästä paremmin (tai selviytyi samantasoisesti) kuin puhenäytteen puhuja (= ei). Käsittelemme alla tehtävään 1 liittyvien puhenäytteiden toimimista mittainstrumentin osana samaan tapaan kuin edellä on jo tehty paperilla tehdyn itsearviointilomakkeen suhteen.

#### 5.1.2.1 Mittarin toimivuus kokonaisuutena

Tehtävässä 1 käytetyn itsearviointilomakkeen toimivuus kokonaisuutena analysoitiin samoin kuin paperilla tehdyn itsearviointilomakkeen toimivuus käyttäen hyväksi sekä perinteistä osioanalyysiä (SPSS:n reliabiliteettianalyysi) että moderniin testiteoriaan perustuvaa osioanalyysiä (Rasch-analyysi Winstepillä). Myös tämän itsearviointilomakkeen havaittiin toimivan hyvin. Cronbachin alfa tälle 26-osioiselle mittarille oli .94. Mittarin erottelukyky oppijoille oli 3.65 ja

osioille 10.94 Winstep-analyysin mukaan (ks. Liite 12). Tämä tarkoittaa, että mittari pystyy erottelemaan oppijat luotettavasti 3–4 ryhmään heidän itsearvioitun kielitaitonsa perusteella. Mittarin osiot taas voidaan aineiston perusteella jakaa ainakin kymmeneen eri luokkaan (taitotasoon).

### 5.1.2.2 Puhenäytteiden laadun analyysi

Samoin kuten edellä tehtiin paperilla tehdyn itsearvioinnin kanssa, tarkastellaan seuraavaksi tehtävän 1 kuunneltujen puhenäytteiden ominaisuuksia ja laadua sekä perinteiseen testiteoriaan kuuluvien osio-testi-korrelaatioiden että moderniin testiteoriaan kuuluvan Rasch-analyysin avulla. Samalla analysoidaan näytteiden sisältöä laadullisesti. Sekä osio-testi-korrelaatioanalyysin että Rasch-analyysin tarkoituksena on selvittää, näyttääkö jokin tietty puhenäyte toimivan samansuuntaisesti kuin samaan kokonaisuuteen kuuluvat muut puhenäytteet eli mittaavan samaa ominaisuutta kuin muut. Näytteiden laadullisen / sisällön analyysin avulla pyritään sitten osoittamaan, löytyykö puhenäytteiden väliltä jokin tai joitakin yhteisiä elementtejä, jotka selittäisivät niiden toimivuuden tai toimimattomuuden.

Tehtävän 1 puhenäytteitä käsiteltiin samoin kuin edellä jo tehtiin paperilla tehdyn itsearvioinnin kanssa. Toisin sanoen puhenäytteet ryhmiteltiin taulukoihin 10 ja 11 taitotasoittain (A1–C2) siten, että taulukon ensimmäisessä sarakkeessa on puhenäytteiden taitotasot alkaen taitotasosta A1. Toisessa sarakkeessa on huonosti toimineiden puhenäytteiden numerot sekä sulkeissa niiden vastausjakaumat (1. luku = ”kyllä”-vastausten ja 2. luku = ”ei”-vastausten lukumäärät kyseistä näytettä koskien). Kolmannessa sarakkeessa on tarkasteltavien tunnuslukujen arvot (perinteisessä osioanalyysissä osio-testi-korrelaatiot ja Rasch-analyysissä Infit mean square ja Infit Zstd -arvot). Taulukon oikeanpuoleisessa reunassa esitellään jälleen hyvin toimineet puhenäytteet taitotasoittain.

Näytteen toimivuutta tarkasteltiin ensin perinteisen osioanalyysin näkökulmasta hyödyntämällä osio-testi-korrelaatioita, jotka saatiin sekä SPSS:n reliabiliteettianalyysistä että Winstepin osio-testi-korrelaatioanalyysistä. Jos osio-testi-korrelaation arvo oli yli .30, näytteen voidaan todeta arvon suuruudesta riippuen toimineen joko kohtalaisesti tai hyvin. Kaikki arvon .30 alle jäävät arvot puolestaan kertovat, että puhenäytteen toimivuudessa on mahdollisesti ongelmia – se on ehkä epäselvästi muotoiltu, tai se mittaa jotakin muuta ominaisuutta kuin muut puhenäytteet.

Perinteisen osioanalyysin jälkeen puhenäytteitä tarkasteltiin moderniin testiteoriaan perustuvan osioanalyysin (Rasch-analyysi) näkökulmasta, jolloin tunnuslukuina käytettiin Infit mean square ja Infit Zstd -arvoja. Rasch-analyysin tunnuslukujen lähtökohtana käytettiin edelleen sitä, että väittämien Infit mean square -arvon tulee sijaita arvojen .7 ja 1.3 välillä ja Infit Zstd -arvon rajojen +2 ja -2 sisällä, jotta niitä voidaan pitää hyvin toimivina mittarin osioina. Samoin kuin paperilla tehdyn itsearvioinnin yhteydessä, on syytä muistaa, että huonosti toimivien puhenäytteiden kokonaislukumäärä voi helposti vai-

kuttaa suuremmalta Rasch-analyysillä tehdyssä tarkastelussa, koska tarkasteltavia tunnuslukuja on kaksi perinteisen osio-testi-korrelaation sijaan.

Seuraavaksi esitellään vertailun helpottamiseksi kahteen peräkkäiseen taulukkoon (taulukot 10 ja 11) koottuina tehtävän 1 hyvin ja huonosti toimineita puhenäytteitä niiden saamien tilastollisten tunnuslukujen avulla. Taulukossa 10 ovat tehtävää 1 koskevat puhenäytteet, jotka on analysoitu perinteisen osio-analyysin mukaisesti. Taulukko 11 sisältää tehtävän 1 puhenäytteiden tilastotieteelliset analyysit Rasch-analyysin näkökulmasta. Taulukoiden 10 ja 11 jälkeen esitellään ensin molempien testiteorioiden näkökulmasta joitakin hyvin toimineita puhenäytteitä. Sen jälkeen tarkasteluvuorossa ovat instrumentissa huonosti toimineet puhenäytteet.

TAULUKKO 10 Kuunneltuihin puhenäytteisiin perustuva itsearviointi: perinteisen osioanalyysin mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 1).

Näytteen taito- taso	Huonosti toimi- neet näytteet	Korjattu osio-testi - korrelaatio	Hyvin toimineet näytteet	Korjattu osio-testi - korrelaatio
<b>A1 (4)</b>	7 (201- 7)	.22	3	.41
A1	17 (197 -11)	.23		
A1	10 (192 -16)	.30		
A1				
A1				
<b>A2 (7)</b>			11	.75
A2			8	.75
A2			4	.70
A2			19	.69
A2			1	.66
A2			22	.62
			26	.42
<b>B1 (10)</b>				
B1			2	.79
B1			18	.78
B1			25	.74
B1			23	.70
B1			15	.69
B1			6	.69
B1			21	.69
B1			20	.68
B1			16	.65
B1			13	.61
<b>B2 (2)</b>			14	.69
B2			24	.57
<b>C1 (3)</b>	12 (6-203)	.24	5 (30 - 179)	.48
C1	9 (7-202)	.26		
<b>C2 (-)</b>				



TAULUKKO 11 Kuunneltuihin puhenäytteisiin perustuva itsearviointi: modernin osioanalyysin (Rasch-analyysi) mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 1).

Näytteen taitotaso	Huonosti toimineet näytteet	Infit mean square	Infit Zstd	Hyvin toimineet näytteet	Infit mean square	Infit Zstd
<b>A1 (4)</b>				7	1.01	.1
A1				17	.99	.1
A1				10	1.18	.9
A1				3	.94	-.3
A1						
<b>A2 (7)</b>						
A2	26	1.41	2.9	11	.82	-.7
A2	22	.73	-2.4	4	.95	-.5
A2	8	.78	-2.3	19	.91	-.9
A2				1	.86	-1.3
A2						
<b>B1 (10)</b>						
B1	2	.65	-3.0	16	1.14	1.0
B1	18	.74	-2.4	13	1.08	.8
B1				25	.78	-1.7
B1				23	.98	-.1
B1				15	1.03	.3
B1				6	1.07	.6
B1				21	1.08	.7
B1				20	1.09	.8
B1						
B1						
<b>B2 (2)</b>						
B2	24	1.38	3.4	14	.88	-.8
<b>C1 (3)</b>						
C1				12	1.06	.3
C1				9	.82	-.3
				5	1.24	1.3
<b>C2 (-)</b>						

### 5.1.2.3 Perinteisen osioanalyysin ja Rasch-analyysin mukaiset hyvin toimineet puhenäytteet

Kun tarkastellaan tehtävän 1 puhenäytteitä perinteisen osioanalyysin mukaan, voidaan todeta, että tason A2 näytteiden (7 kpl) osiot-testi-korrelaatioiden arvot ovat erittäin korkeita. Seitsemästä puhenäytteestä kuuden korrelaatioarvo on yli 0.60. Samaisen itsearviointilomakkeen B1-tasolla olevien 10 puhenäytteen korrelaatioarvot sijoittuvat kaikki välille 0.61–.079, joten niiden voidaan todeta toimineen erittäin hyvin. Lomakkeen korrelaatioarvojen määrää tarkasteltaessa todettakoon lisäksi, että seitsemällä lomakkeen näytteistä korrelaatiokerroin on yli .70 ja kymmenellä yli .60.

Myös Rasch-analyysin mukaan tehtävän 1 puhenäytteet toimivat pääosin hyvin. 26 näytteestä vain kolmen Infit mean square- ja Infit Zstd -arvot yhdessä olivat hyväksyttävien rajojen ulkopuolella. Taitotasoilla A1 ja C1 olevat näytteet toimivat kaikki Rasch-analyysin mukaan hyvin.

Aluksi tarkastellaan yksityiskohtaisesti joitakin eritasoisia näytteitä, jotka ovat toimineet hyvin sekä perinteisen että Rasch-analyysin näkökulmasta. Tarkastelun tarkoituksena on yrittää selvittää, sisältävätkö kyseiset näytteet mahdollisesti sellaisia elementtejä, jotka selittäisivät niiden hyvän toimivuuden. Tarkastelun kohteena ovat näytteet 11, 21 ja 15. Näytteiden litteroinnissa käytetyt merkinnät löytyvät taulukosta 28 kohdasta 5.3.

### Näyte 11

(A2; korrelaatio .75; Infit mean square .82; Infit Zstd -1.7; 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.)

Yleisluonnehdinta puhetuotoksesta:

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.

### Näyte 21

(B1; korrelaatio .69 .; Infit mean square 1.08 ; Infit Zstd -.7; vastausjakauma 75–134 )

Euh\_ je xxx (ei saa selvää) avec François\_ qui est \_ \_un garçon\_ il se repose [ääntyy repo:z] après \_ euh \_il est joué \_ au foot \_ \_euh\_ il est été\_ parce que: \_ il a seulement un \_ un t-shirt et jeans et chaussures du tennis \_ euh [mieltii] \_ \_ et \_ probablement \_ et \_ il est un \_étudiant\_ euh\_ qui ne travaille pas \_ pendant \_ l'été ( 34 sek.)

Yleisluonnehdinta puhetuotoksesta:

naisääni, melko lyhyt näyte, jättää melko sujuvan vaikutelman, mutta toisaalta puhuja mieltii jonkin verran (ei keksi sanottavaa?), näytteen lyhyiden vuoksi vaikea arvioida sujuvuuden todellista tasoa, esitys selkeä, joitakin ongelmia peruskieliopin hallinnassa (mm. artikkelien käyttö), ääntäminen kokonaisuutena hyvä, hyvätasoinen B1-tason näyte.

Näytettä 21 tarkastellaan tarkemmin myöhemmin työssä kohdassa, jossa käsitellään tehtävää 1 koskevia näytteitä, jotka keräsivät eniten ”vaikea päättää” -rasteja tutkimukseen osallistuneilta oppijoilta.

### Näyte 15

(B1; korrelaatio .69; Infit mean square 1.03 ; Infit Zstd .3; vastausjakauma 70–139 )

Mais\_ \_ je crois que François\_ euh il a l’air un peu triste parce que\_ il voudrait jouer du \_basket ball et \_il a demandé ses camarades et \_\_ou euh\_ ses ses amis\_ euh pour jouer le\_ basket \_et\_euh\_ maintenant\_ il doit seulement s’asseoir et \_ attendre si quelqu’un voudrait venir parce que \_ tout le monde [mon] euh \_ \_ \_ tout le monde [mon] \_ sont dit que \_ \_ qu’on doit mang – manger \_ et après ça\_ peut-être\_ on peut jouer le basket ball \_ et \_ il peut\_ seulement attendre (49 s.)  
Yleisluonnehdinta puhetuotoksesta:

miesääni, melko sujuvaa, hyvätasoinen B1-tasoa edustava näyte, hyviä rakenteita, melko hyvä ääntäminen, tuotos pidempi kuin edellä (49 s.).

Myös näyte 15 kuuluu niihin näytteisiin, joiden osalta oppijoilla on ollut vaikeuksia päättää, vastaavatko he ”kyllä” vai ”ei” kuunneltuaan näytteen. Näytettä tarkastellaan siten yksityiskohtaisemmin jäljempänä tekstissä kohdassa, jossa käsitellään tehtävän 1 eniten ”vaikea päättää”-rasteja keränneitä näytteitä.

Tehtävän 1:n aihepiiri on melko tavallinen. Kuvasta kertomisen tehtävänanto jätettiin tarkoituksellisesti avoimeksi, jotta kukin oppija pystyi tuottamaan puhetta mahdollisimman paljon ja pitkään. A2-tasoinen näyte 11 on heikompi kuin kaksi muuta, mikä onkin täysin odotuksenmukaista A-tason puhenäytteelle. Henkilön ääntäminen on yleensä aika selvää, mutta sanan loppujen lausuminen saattaisi joissakin tilanteissa aiheuttaa ymmärtämisongelmia. Puhe ei ole kovin sujuvaa, mutta se on ymmärrettävää, ainakin lyhyehkössä tuotoksessa, jonka aihepiiri on melko helppo. Tauot tekevät puheesta vähän katkonaisen. Puhuja pystyy kuitenkin kertomaan tarinan yksinkertaisena luettelona.

Kaksi muuta näytettä, näytteet 21 ja 15, edustavat molemmat hyvin B1-tasoista kielitaitoa (= kynnystaso), jolle on tyypillistä se, että henkilö on ikään kuin jo astunut tietyn ”osaamiskynnyksen” yli ja pystyy siten ilmaisemaan itseään huomattavan paljon sujuvammin kuin taitotasoa A2 edustava näytteen 11 puhuja. Molempia näytteitä leimaa tietynlainen sujuvuus, vaikka kieliopillisiä ongelmia esiintyykin.

Yhteenvetona tehtävä 1:n hyvin toimineista esimerkinäytteistä voidaan todeta ensinnäkin se, että ne pysyvät kaikki kolme niille määritellyllä taitotasolla. Tosin näyte numero 11 sijoittuu vaikeustasonsa puolesta aivan A2-tason ylärajalle, missä se on hyvin lähellä B1-tasoisia näytteitä (kuvio 6). Toiseksi voidaan todeta, että tarkastellut puhenäytteet vaikuttaisivat vastaavan hyvin Viitekehysten kyseisille tasoille (A2 ja B1) määrittelemää puhumisen osaamista. Todettakoon siten, että jos puhenäyte on tyypillinen tietyn taitotason näyte, on

odotettavaa, että se toimii mittainstrumentin osana moitteettomasti. Se, että jokin näyte edustaa tyypillisesti tiettyä taitotasoa ja toimii hyvin mittainstrumentin osana, ei kuitenkaan välttämättä tarkoita sitä, että kaikkien oppijoiden olisi helppo verrata omaa puhetuotostaan kyseiseen näytteeseen ja valita ”kyllä”- ja ”ei” -vaihtoehtojen väliltä. Tätä aihetta käsitellään tarkemmin ”vaikea päättää”-rasteja esittelevässä tutkimuksen kohdassa jäljempänä.

#### *Yleistä huonosti toimineista puhenäytteistä*

Kun tarkastellaan yllä olevaa perinteistä osioanalyysiä koskevaa taulukkoa (taulukko 10), huomio kiinnittyy samantyyliisiin seikkoihin kuin paperilla tehdyssä itsearviointinnissakin. Huonommin toimineet puhenäytteet sijoittuvat myös studiossa kuunneltavien puhenäytteiden pohjalta tehdyssä itsearviointinnissa taitotasoasteikon ääripäihin eli tasoille A1 (2 kpl) ja C1 (2 kpl)<sup>19</sup>. Huonosti toimineiden puhenäytteiden korrelaatioarvot ovat kuitenkin kaikki yli .20 ja yleisellä tasolla korkeampia kuin paperilla tehdyn itsearviointilomakkeen väittämien vastaavat arvot. Tämä saattaa ainakin osittain johtua siitä, että vaikka jakaumat näissäkin ovat epätasaisia, ne eivät kuitenkaan ole niin epätasaisia kuin kirjallisissa itsearviointiväittämässä. Huonosti toimineita puhenäytteitä on siis neljä: näytteet 7, 9, 12 ja 17 (näytteen 30 korrelaatioarvo on juuri hyväksyttävän alarajalla eli .30).

Alla olevassa korrelaatiokertoimia koskevassa tarkastelussa etsitään mahdollisia yhteisiä syitä näytteiden toimimattomuuteen. Näytteet on järjestetty korrelaatiokertoimensa perusteella järjestykseen siten, että ensin esitellään näyte, jolla on alin korrelaatiokerroin. Näytteiden tarkastelun pohjaksi käytössä on kolmenlaista tietoa: 1. tilastollisista analyyseistä saatu mittalukuihin pohjautuva tieto, 2. Yki-korpuksen arvioijilta saatu puhenäytteiden tasoa koskeva tieto ja 3. näytteiden vaikeustasosta saatu tieto.

Rasch-analyysin, jossa siis puhenäytteitä tarkasteltiin Infit mean square ja Infit Zstd -tunnuslukujen avulla, pohjalta huonosti toimivia väittämiä on tehtävässä 1 yhteensä 6 kappaletta. Varsinaisia ”huonoja” näytteitä, joissa siis molemmat tarkasteltavat tunnusluvut ovat suositeltavien raja-arvojen ulkopuolella, on kuitenkin tästä vain puolet eli kolme puhenäytettä. Muut kolme näytettä ovat näytteitä, joiden Infit mean square -arvo pysyy hyväksyttävissä rajoissa, mutta Infit Zstd -arvo ei pysy välillä +2 ja -2. Tämä saattaa tarkoittaa joitain ongelmia näytteen toimimisessa instrumentin osana, mutta asian vahvistaminen vaatisi lisätutkimuksia eikä kuulu tämän tutkimuksen piiriin. Tasoilla A1 ja C1 ei ole lainkaan huonosti toimineita puhenäytteitä tehtävässä 1. Huonosti toimivat näytteet ovat 26 (A2), 2 (B1) ja 24 (B2), eli jälleen Rasch-analyysi pystyi identifioimaan asteikon keskivaiheille sijoittuvien osioiden mahdollista ongelmallisuutta. Tarkastelemme niitä erikseen perinteisen osioanalyysin huonosti toimineiden väittämien jälkeen.

<sup>19</sup> Taitotasoa C2 edustava näyte puuttui näytteistä.

#### 5.1.2.4 Perinteisen osioanalyysin mukaiset huonosti toimineet puhenäytteet

Perinteinen osioanalyysi erottelee selkeästi neljä puhenäytettä huonosti toimineiksi näytteiksi. Yksi näytteistä (näyte 10) on korrelaatiokertoimensa vuoksi hyväksyttävän rajalla (.30), ja siksi sitä ei analysoida huonosti toimineena näytteenä. Merkille pantavaa on, että huonosti toimineet näytteet sijoittuvat kaikki joko taitotasosteikon alapäähän (A1: näytteet 7 ja 17) tai päinvastoin asteikon yläpäähän (C1: näytteet 9 ja 12). Kaikkien neljän tarkasteltavan näytteen korrelaatioarvot ovat selkeästi suositeltavan rajan .30 alapuolella (.22 – .26). Seuraavaksi tarkastellaan yksityiskohtaisesti näitä neljää puhenäytettä perinteisen osioanalyysin näkökulmasta.

##### Näyte 7

(A1; korrelaatiokerroin .22; Infit mean square 1.01; Infit Zstd .1; vastausjakauma 201–7)

Kyseessä on alimman tason A1 näyte, jonka tasosta kaikki Yki-korpuksen arvioijat ovat olleet täydellisen yksimielisiä. Näytteiden vaikeustasoa kuvaavassa kuviossa (kuvio 6) näyte 7 sijoittuu alimmaiseksi A1-tason näytteeksi ja pysyy siten hyvin sille määritellyllä tasolla. Näytteen puhuja puhuu erittäin hitaasti ja katkonaisesti. Lähestulkoon kaikki tutkimukseen osallistujat valitsivat ”kyllä”-vaihtoehdon kuunneltuaan tämän näytteen, eli he olivat sitä mieltä, että selvisivät tehtävästä paremmin kuin tämän näytteen puhuja. Näyte ei siten erottele.

##### Näyte 17

(A1; korrelaatiokerroin .23; Infit mean square .99 ; Infit Zstd .1 ; vastausjakauma 197–11)

Yki-korpuksen viiden arvioijan työtä tarkastellessa voidaan havaita, että heidän arvionsa ovat jakautuneet kahteen ryhmään: kolme arvioijista on sijoittanut näytteen tasolle A1 ja kaksi tasolle A2. Näyte sijoittuu vaikeustasoa kuvaavassa kuviossa (kuvio 6) taitotasolle A1. Näytteessä kuuluu olevan erityisesti puutteelliseen ääntämiseen ja katkonaisuuteen liittyviä ongelmia. Valtaosa tutkimukseen osallistujista (197) on valinnut ”kyllä”-vaihtoehdon, ja siten myöskään tämä näyte ei erottele riittävän hyvin.

##### Näyte 12

(C1; korrelaatiokerroin .24; Infit mean square 1.06; Infit Zstd .3; vastausjakauma 6–203)

Myös näytteen 12 osalta viiden arvioijan arviot jakautuivat kahteen ryhmään seuraavasti: kolme arvioijaa on sijoittanut näytteen tasolle C1 ja kaksi tasolle C2. Tilastollisten analyysien tarpeista johtuen näytteen taso on tässä tutkimuksessa määritelty enemmistöperiaatteella; koska kolme arvioijaa viidestä on sijoittanut näytteen tasolle C1, näytteen tasoksi on määritelty tämä taso. Näytteen

vaikeustasoa tarkasteltaessa voidaan todeta sen pysyvän sille määritellyllä taitotasolla C1 (kuvio 6). Tutkimukseen osallistuneiden oppijoiden vastauksia tarkasteltaessa voidaan jälleen huomata, että näyte ei erottele hyvin osallistujien kielitaitoa: ainoastaan kuusi osallistujaa on vastannut tähän näytteeseen ”kyllä”. On kuitenkin huomattava, että vähäisestä ”kyllä”-vastausten määrästä huolimatta osion korrelaatio on kuitenkin yli .20, mistä voi päätellä että kuusi ”kyllä”-vastauksen antanutta oppijaa oli juuri niitä, joiden oletettiin vastaavan positiivisesti tämäläisyyteen osioon (= oppijoita, joiden kielitaito oli hyvin korkealla taitotasosteikolla). Siten ne oppijat, joiden kielitaito oli hyvällä (mutta ei erinomaisella) tasolla tai jotka kokivat selviytyvänsä tehtävästä yhtä hyvin kuin nauhan erittäin hyvätasoinen puhuja, joutuivat tutkimuksen sääntöjen mukaan vastamaan tähän näytteeseen ”ei”. Tämän pohjalta voidaan todeta, että menetelmä, jolla tutkimusaineisto hankittiin, näyttäisi vaikuttavan jonkin verran tutkimuksesta saataviin tuloksiin. Tämä näyttäisi koskevan varsinkin erittäin hyviä ja erittäin heikkoja näytteitä.

### Näyte 9

(C1; korrelaatiokerroin .26; Infit mean square .82 ; Infit Zstd-.3 ; vastausjakauma 7-202)

Puhenäytteen numero 9 korrelaatioarvo on melko lähellä toivottua .30 rajaa. Neljä alkuperäiskorpuksen viidestä arvioijasta on sijoittanut tämän näytteen tasolle C1 ja yksi arvioija tasolle C2. Näytteen vaikeustasoa kuvaavassa kuviossa näyte pysyy taitotasolla C1 (kuvio 6). Vain seitsemän tutkimukseen osallistujaa on valinnut ”kyllä”-vaihtoehdon kuunneltuaan näytteen numero 9.

Yhteenvedona korrelaatioarvojen valossa huonosti toimineista puhenäytteistä tehtävässä 1 voidaan todeta, että mitä todennäköisimmin näytteet 9, 12, ja 17 ovat kahden eri taitotason rajalle sijoitettavia näytteitä; näytteet 9 ja 12 taitotason C1 ja C2 ja näyte 17 tason A1 ja A2 rajalle. Tämä tulkinta saa vahvistusta viiden arvioijan tasolle sijoittamista tarkasteltaessa. Näyte numero 7 on liian helppo, koska kukaan yli 200 osallistujasta ei ole valinnut sen kuultuaan ”ei”-vaihtoehtoa. On todennäköistä, että näytteet, jotka sijoittuvat kahden eri taitotason rajalle joko asteikon yläpäässä (C1–C2) tai alapäässä (A1–A2) käyttäytyvät ”epävakaasti”. Niillä näyttää olevan selvä vaikutus tietyn osion korrelaatioon suhteessa koko käytetyn instrumentin kokonaispistemäärään.

Kaikkien neljän mittarissa huonosti toimineen väittämän osio-testi-korrelaatioiden arvo jää hieman alle toivotun rajan .30 olematta silti kovin kaukana tästä raja-arvosta. Sen sijaan mittarin Infit mean square ja Infit zstd -arvot eivät osoita merkkejä siitä, että näiden hieman alhaisen korrelaatiokerroimen omaavien näytteiden kohdalla olisi syytä epäillä vakavampia toimimattomuusongelmia, sillä Infit mean square -arvot pysyvät kaikki hyväksyttävien raja-arvojen eli .7 ja 1.3 välillä. Myös Infit zstd -arvot pysyvät neljän kyseessä olevan osion suhteen suositeltavissa rajoissa, minkä pohjalta voidaan todeta, että kyseiset osiot eivät osoita merkkejä sopimattomuudesta mittariin. Myöskään näytteisiin liittyvissä vaikeustasoissa ei tapahdu muutoksia, eli oppijat

ovat pitäneet käsiteltyjä näytteitä niille määritellyille taitotasolle kuuluvina eivätkä helpompina tai vaikeampina.

### 5.1.2.5 Rasch-analyysin mukaiset huonosti toimineet puhenäytteet

Tehtävän 1 puhenäytteistä selkeästi ongelmallisilta Rasch-analyysin näkökulmasta vaikuttivat puhenäytteet 2, 24 ja 26, joiden Infit mean square ja Zstd -arvot eivät pysy hyväksyttävissä rajoissa. Näytteen 2 Infit mean square -arvo jää alle hyväksyttävän alarajan .70 ja näytteiden 24 ja 26 vastaavat arvot ylittävät suositellun ylärajan (1.3). Samoin mikään mainituista näytteistä ei pysy Infit Zstd -arvonsa puolesta +2: n ja -2:n välillä. Näytteitä tarkastellaan Yki-korpuksen arvioijien antamien kyseistä näytettä koskevien tasoarviointien, Winstepin osioiden vaikeutta osoittavan analyysin, mahdollisten oppijoiden kommenttien sekä näytteen yleisluonnehdinnan avulla.

#### Näyte 2

(taso B1; korrelaatio .79; Infit Mnsq .65; Infit ZStd -3.0; vastausjakauma 68–141)

Yki-korpuksen arvioijat ovat jakautuneet kahteen ryhmään arvioidessaan näytteen 2 tasoa. Kolme arvioijista on sijoittanut näytteen tasolle B1, kun taas 2 arvioijaa on pitänyt näytettä pikemminkin A2-tasoisena. Winstep-analyysin osoittaman osioiden vaikeutta kuvaavan analyysin mukaan tämä näyte vaikuttaa sijoittuvan oman tasoistensa B1-näytteiden joukkoon.

Kyseessä on naispuhujia, joka tuottaa puhetta melko sujuvasti, mutta jonka ääntäminen ei ole vielä täysin kohdekielen mukaista. Puhujan artikkelien käyttö on lisäksi satunnaista, ja hän tuottaa puhetta hieman katkonaisesti. Näytteen puhujia edustaa todennäköisesti pikemmin B1-tason alareunaan kuuluvaa osaamista kuin B1-tason vahvaa osaamista, joka olisi jo melko sujuvaa.

#### Näyte 24

(taso B2; korrelaatio .57; Infit Mnsq 1.38/ZStd 3.4; vastausjakauma 104–105)

Yki-korpuksesta tulevien näytteiden tasoarvioinnit jakautuvat jälleen kahteen ryhmään siten, että 3 arvioijaa on sijoittanut tämän näytteen taitotasolle B2 ja 2 arvioijaa tasoa alemmaksi eli tasolle B1. On huomattava, että tämä näyte nousi esille työssä jäljempänä kohdassa, jossa tarkastellaan osioiden vaikeutta Winstep-analyysin avulla. Näyte sijoittui näet mittarin osioiden sijoittumista kuvaavassa kuviossa silmiin pistävän alas lähelle taitotasojen A2–B1 rajaa, vaikka kyseessä arvioijien enemmistön mukaan olisi pitänyt olla taitotasoa B2 edustava näyte (tarkempi selostus seuraa jäljempänä osioiden vaikeutta tarkasteltaessa). Onkin siis hyvin todennäköistä, että tämä näyte sisältää joitakin sellaisia kielen piirteitä, jotka oppijoiden näkemyksen mukaan tekevät siitä huomattavasti helpomman kuin tasoa B2 edustavan näytteen. Näyte saattaa siten edustaa pikemminkin taitotasoa B1 kuin tasoa B2 tai ainakin siitä voidaan perustellusti olla kahta eri mieltä riippuen siitä, mitä kriteeriä painotetaan.

Näytteen miespuhujia ääntää hyvin ja käyttää hyviä lauserakenteita. Tuotos on lyhyt (41 sekuntia), mutta sen ajan, kun henkilö puhuu, hän tuottaa puhetta sujuvasti ja hänen mielikuvituksensa vaikuttaa rikkaalta. Puhuja lopettaa tuotoksensa puhekielenomaiseen humoristiseen ilmaisuun (*il va faire une carrière terrible*<sup>20</sup>), jota monet tutkimukseen osallistuneet aloittelevat ranskan oppijat pitivät outona, mutta jota taas monet pidemmälle kieliopinnoissaan edenneet oppijat pitivät hauskana lopetuksena.

### Näyte 26

(taso A2; korrelaatio .42; Infit Mnsq 1.41/ZStd 2.9; vastausjakauma 159–50)

Yki-korpuksen arvioijien taitotasot näytettä 26 koskien ovat seuraavat: taitotaso A2 (4 arvioijaa) ja taitotaso A1 (1 arvioija). Osioiden vaikeutta tarkastelevassa Winstep-analyysissä näyte 26 pysyy hyvin omalla taitotasollaan sijoittuen kuitenkin aivan tason A2 alapäähän. Kaikki Yki-korpuksen arvioijat ovat siis olleet hyvin samoilla linjoilla näytteen vaikeustason suhteen.

Kyseessä on miesääni, jonka tuotos on hyvin lyhyt (29 sekuntia), ja se sisältää paljon miettimistä ja taukoja. Ääntäminen on melko hyvää, mutta tuotoksessa on paljon ongelmia oikeakielisyyden kannalta (mm. artikkelien puuttuminen ja pronominien käyttö). Henkilö ei kerro kuin muutama lauseen verran kuvasta ja sen jälkeen hän lopettaa tuotoksensa listamaisesti ulkoa opeteltuihin fraaseihin ja toivotuksiin (*oui, c'est ça, c'est tout, merci, au revoir, courage!*). Kokonaisuutena tuotos on niin lyhyt, että oppijoilla on voinut olla vaikeuksia verrata omaa tuotostaan tähän näytteeseen yksinkertaisesti sen vähäisen sisällön vuoksi. Vähäinen sisältö vaikuttaa myös siihen, että voi olla vaikeaa arvioida puhujan todellista kielitaidon tasoa, sillä luettelemalla peräkkäin ulkoa opeteltua kielialainesta yllä kuvattuun tapaan tuotos saattaa antaa kuvan tietynlaisesta sujuvuudesta, mitä on kuitenkin erittäin vaikea todentaa näin lyhyen tuotoksen pohjalta.

Yhteenvetona näytteistä 2, 24 ja 26 todettakoon seuraavaa: näytteiden 24 ja 26 puhujat ovat molemmat miehiä eivätkä kovin monisanaisia. Molempien tuotos on melko lyhyt, mikä saattaa vaikeuttaa niiden tasolle sijoittamista ja oppijan oman tuotoksen vertaamista näytteisiin, kun otetaan huomioon, että oppijat kuulivat näytteet vain kerran. Miespuhujien puhenäytteiden sijoittamista tasolle saattaa vaikeuttaa myös se, että molemmat puhujat olivat omalla tavallaan ja tasollaan suhteellisen sujuvan oloisia. Näytteen 2 naispuhujan tuotos on hyvä esimerkki siitä, että jopa saman tason sisällä olevien näytteiden taso voi olla hyvinkin erilainen. Edempänä työssä käsitellyt B1-tasoiset näytteet ovat sijoittuneet korkeammalle kuin tämä B1-tason näyte, jossa on kuultavissa vielä paljon A2-tason suorituksen piirteitä.

Taulukkoon 12 on listattu tehtävän 1 osalta kuunneltaviin puhenäytteisiin perustuvan itsearviointin mahdollisesti ongelmallisia osioita sekä perinteisen osioanalyysin että moderniin testiteoriaan perustuvan Rasch-analyysin perus-

<sup>20</sup> Suora käänös: 'Hän tulee tekemään hirvittävän uran [jalkapalloilijana]'.



teella. Vertailuarvoina on käytetty korrelaatiokerrointa (perinteinen osioanalyysi) ja Infit mean square -arvoa (Rasch-analyysi).

TAULUKKO 12 Mahdollisesti ongelmalliset puhenäytteet perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta (tehtävä 1).

ONGELMALLISET OSIOT				
	Perinteinen osioanalyysi		Moderniin testiteoriaan perustuva Rasch-analyysi	
	Näyte	Korrelaatio	Näyte	Infit mean square
Itsearviointi:	7	.22	2	.65
Kuunneltavat	17	.23	24	1.38
näytteet,	12	.24	26	1.41
tehtävä 1	9	.26		

#### 5.1.2.6 Puhenäytteiden vaikeustason tarkastelua

Lopuksi tarkastellaan vielä tehtävän 1 kuunneltaviin puhenäytteisiin perustuvan itsearviointilomakkeen näytteiden 'vaikeutta' vertaamalla niiden ennalta määritellyjä vaikeustasoja oppijoiden vastauksiin.

Kuviosta 6 (ks. s 170) näkyy, että taitotasolle A1, A2 ja B1 sijoitetut näytteet erottelevat hyvin ja näyttävät pysyvän omalla paikallaan eli niiden voidaan todeta edustavan sen tasoista kielitaitoa, jota niiden odotetaan edustavan. Osiot, jotka kiinnittävät huomiota sijoittumisellaan asteikossa, ovat näytteet 24 (B2) ja 13 (B1), joita tarkastellaan alla erikseen.

#### Näyte 24 (B2)

(taso B2; korrelaatio .57; Infit Mnsq 1.38/ZStd 3.4; vastausjakauma 104–105)

Aluksi on syytä huomauttaa, että näyte 24 nousi esille edellä Rasch-analyysin yhteydessä, jossa se luokiteltiin huonosti toimivaksi näytteeksi. Osioiden vaikeustasoja kuvaavaa kuviota katsottaessa huomio kiinnittyy siihen, että B2-tasoinen näyte 24 sijoittuu kuviossa A2-tasoinen näytteiden joukkoon. Lisäksi on huomattava, että koska näyte sijoittuu hyvin alas asteikossa, sen välimatka toiseen B2-tasoiseen näytteeseen 14, joka näyttää sijoittuvan hyvin B2-tasolle, on huomattavan pitkä. Siten näytteet 24 ja 14, joiden tulisi edustaa samantasoista kielitaitoa, eivät sijoitu edes vierekkäisille taitotasolle, vaan niiden väliin jää kokonainen taitotaso (B1). Näytteessä 24 on mitä luultavimmin joitain sellaisia piirteitä, jotka vaikuttavat siihen, että oppijat eivät ole pitäneet sitä niin vaativana kuin B2-tasoinen näytteiden voisi olettaa olevan. Näytettä on analysoitu edellä Rasch-analyysin osoitettua näytteen toimineen huonosti. Tämän analyysin pohjalta voidaan todeta, että näytteen lyhyys varmasti vaikuttaa siihen kokonaiskuvaan, joka näytteen tasosta jää oppijoille. On selvää, että näin suuret erot näytteiden tasoissa eivät ole toivottavia. Yleisesti ottaen on syytä vielä lisä-

tä, että mittari sisältää vain kaksi B2-tasoista näytettä, mikä myös saattaa hankaloittaa B2-tasoisen kielitaidon arviointia tämän mittarin avulla.

**Näyte 13 (B1)**

(taso B1; korrelaatio .61; Infit Mnsq 1.08 / ZStd .8; vastausjakauma 123–86 )

Tämä B1-tasoinen näyte ei myöskään näytä pysyvän omalla taitotasollaan, vaan se sijoittuu asteikossa määritellyä tasoaan alemmaksi eli A2-tasoisten näytteiden joukkoon. Osallistujat ovat siten pitäneet myös tätä näytettä hieman helpompana kuin mikä on sen määritelty taitotaso. Näytettä 13 tarkastellaan lähemmin jäljempänä tutkimuksessa sen saamien ”vaikea päättää” -rastien suuren lukumäärän vuoksi.

Lopuksi voidaan vielä todeta, että tehtävää 1 koskevassa mittarissa ei näyttäisi olevan näytteitä taitotasolta C2, mikä ei luonnollisestikaan ole hyvä asia mittarin tasapainoisuuden kannalta. Toisaalta kuviota 6 tarkastellessa tulee mieleen mahdollisuus, että näytteet 9 ja 12 edustaisivatkin taitotasoa C2. Niiden etäisyys seuraavaksi alempaan näytteeseen 5 (C1) on sen verran suuri ja selkeä, että niiden tason voitaisiin olettaa olevan C2. Tätä mahdollista tulkintaa tukevat myös arvioijien kahdelle C-tasolle jakautuvat kyseisiä näytteitä koskevat taitoarvioinnit.

```

Person - MAP - Item
<more>|<rare>
8      .# +
      # | t1_12_C1
      # | t1_9_C1
7      +
      | T
6      +
      | T
      .#####
5      +
      |
      ####
4      +
      | t1_5_C1
      ###
3      + S
      .#### S
      #####
2      ## + t1_14_B2
      | | t1_23_B1 t1_25_B1
      .### | t1_16_B1
      ### | t1_15_B1 t1_2_B1
1      ### + t1_21_B1
      ### | t1_20_B1
      | | t1_18_B1 t1_6_B1
      .# | t1_11_A2
0      # +M
      ##### M | t1_24_B2 t1_4_A2 t1_8_A2
      #####
      ##### | t1_19_A2
-1     + | t1_13_B1
      #####
      .#####
-2     ## +
      |
      .#### | t1_22_A2
      | | t1_26_A2
-3     +
      .##### S | S
      #####
-4     +
      | t1_3_A1
      ###
-5     + | t1_10_A1
      . |
      ## | t1_17_A1
-6     +
      | T
      | T | t1_7_A1
      .#
-7     . +
      <less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

```

KUVIO 6 Tehtävä 1: puhenäytteiden vaikeus oppijoiden vastausten perusteella.

### 5.1.2.7 Yhteenvedoa tehtävän 1 itsearviointimittarin luotettavuudesta

Yhteenvedona tehtävän 1 kuunnelluista puhenäytteistä (26 kappaletta) voidaan todeta, että puhenäytteet näyttävät toimivan itsearviointitarkoituksessa kokonaisuutena erittäin hyvin. Kuten paperilla tehdyssä itsearvioinnissa, myös tehtävää 1 koskevan lomakkeen laatua tarkasteltiin osio-testi-korrelaatioiden sekä Winstepin Infit mean square ja Infit zstd -arvojen avulla. Osio-testi-korrelaatioiden tasoista noin 69 % sijoittui arvojen .57–.75 välille ja 15 % arvojen .30–.48 välille. Siten 84 % tämän itsearviointimittarin osioista ylitti tason .30. Mittarin osioista 4 (15 %) jäi suositellun raja-arvon .30 alle. Paperilla tehdyn itsearvioinnin osio-testi-korrelaatioista hieman pienempi osa eli 60 % sijoittui suunnilleen samalle välille kuin tehtävää 1 koskevat osio-testi-korrelaatiot. Suositellun raja-arvon .30 ylittäneiden osioiden määrä on suunnilleen sama molemmissa tähän asti tarkastelluissa mittareissa (paperilla tehty itsearviointi 85 %, tehtävää 1 koskeva itsearviointi 84 %). Kun verrataan erittäin pienten (alle .20) osio-testi-korrelaatioiden kokonaismäärää paperilla tehdyn itsearvioinnin ja tehtävän 1 kuunneltaviin näytteisiin perustuvan itsearvioinnin välillä, voidaan todeta, että paperilla tehdyssä itsearvioinnissa alle .20 korrelaatioita oli paperilla tehdyssä itsearvioinnissa 4 kappaletta, kun taas tehtävässä 1 niitä ei ollut yhtään.

Winstepin Infit Mean square -arvoista valtaosa eli noin 88 % oli hyväksyttävissä rajoissa eli .7 ja .1.3 välillä. Kolmen näytteen (näytteet 2, 24 ja 26) kohdalla oli epäilystä niiden sopivuudesta mittarin taustalla olevaan tilastolliseen malliin.

### 5.1.3 Tehtävä 2: kuunneltaviin puhenäytteisiin perustuva itsearviointi

Tehtävää 2 koskevan kuunneltaviin näytteisiin perustuvan itsearviointilomakkeen luotettavuus analysoitiin samaan tapaan tilastollisesti kuin edellä on selostettu (paperilla tehty itsearviointi ja tehtävää 1 koskeva itsearviointi). Itsearviointi tehtävässä 2 suoritettiin samoin kuin tehtävässä 1 eli oppijat itsearvioivat omaa suoriutumistaan jälleen kuunneltaviin puhenäytteisiin verraten. Puhenäytteitä oli tehtävässä 2 kolme vähemmän kuin tehtävässä 1 eli yhteensä 23 kappaletta. Tehtävän 2 puhenäytteet olivat myös huomattavasti lyhyempiä. Oman suoriutumisen vertaaminen nauhalta kuunneltavaan näytteeseen tapahtui samojen sääntöjen mukaan kuin tehtävässä 1: oppijan tuli siis valita kuunneltavien näytteiden osalta, selviytyikö hän omasta mielestään paremmin tehtävästä kuin kuunnellun näytteen puhuja (= kyllä) vai oliko osallistuja sitä mieltä, että hän itse ei selviytynyt tehtävästä paremmin (tai selviytyi samantasoisesti) kuin näytteen puhuja (= ei).

#### 5.1.3.1 Mittarin toimivuus kokonaisuutena

Tehtävän 2 puhenäytteisiin perustuvan itsearviointilomakkeen toimivuus kokonaisuutena analysoitiin samoin kuin paperilla tehdyn itsearviointilomakkeen ja tehtävässä 1 käytetyn itsearviointilomakkeen toimivuus. Analyysissä käytet-

tiin siten edelleen hyväksi perinteistä SPSS:n reliabiliteettianalyysia ja Winstepin avulla toteutettua moderniin testiteoriaan perustuvaa Rasch-analyysia. Myös tehtävää 2 koskeva itsearviointilomake toimi hyvin, sillä sen Cronbachin alfan arvo oli .89. Mittarin erottelukyky oppijoille oli 2.98 ja osioille 10.23 Winstep-analyysin mukaan (ks. Liite 13). Mittari näyttäisi siten pystyvän erottelemaan oppijat luotettavasti noin 3 ryhmään heidän itsearvioidun kielitaitonsa perusteella, kun taas mittarin osiot voidaan aineiston perusteella jakaa ainakin kymmeneen eri luokkaan (taitotasoon).

### 5.1.3.2 Puhenäytteiden laadun analyysi

Tässä luvussa tarkastellaan tehtävän 2 kuunneltujen puhenäytteiden ominaisuuksia ja laatua sekä perinteiseen testiteoriaan kuuluvien osio-testi-korrelaatioiden että moderniin testiteoriaan kuuluvan Rasch-analyysin avulla. Samalla analysoidaan näytteiden sisältöä laadullisesti. Puhenäytteiden toimivuutta eli sitä, näyttääkö jokin puhenäyte mittaavan samaa ominaisuutta kuin muut tehtävän 2 puhenäytteet, pyritään selvittämään edellä mainittujen analyysitapojen avulla. Näytteiden laadullisen ja sisällöllisen analyysin avulla pyritään sitten osoittamaan, löytyykö puhenäytteiden väliltä jokin tai joitakin yhteisiä elementtejä, jotka selittäisivät niiden toimivuuden tai mahdollisen toimimattomuuden.

Tehtävän 2 puhenäytteet ryhmiteltiin taulukkoihin 13 ja 14 taitotasoin (A1–C2) siten, että taulukon ensimmäisessä sarakkeessa ovat puhenäytteiden taitotasot alkaen taitotasosta A1. Toisessa sarakkeessa ovat huonosti toimineiden puhenäytteiden numerot sekä sulkeissa niiden vastausjakaumat (1. luku = ”kyllä”-vastausten ja 2. luku = ”ei”-vastausten lukumäärä). Kolmannessa sarakkeessa on tarkasteltavien tunnuslukujen arvot (perinteisessä osioanalyysissä osio-testi-korrelaatiot ja Rasch-analyysissä Infit mean square ja Infit Zstd -arvot). Taulukon oikeanpuoleisessa reunassa esitellään hyvin toimineet puhenäytteet taitotasoin.

Näytteen toimivuutta tarkasteltiin siten kuin edellä on jo tehty eli ensin perinteisen osioanalyysin näkökulmasta osio-testi-korrelaatioiden avulla, jotka saatiin sekä SPSS:n reliabiliteettianalyysistä että Winstepin osio-testi korrelaatioanalyysistä. Jos osio-testi-korrelaation arvo oli yli .30, näytteen voidaan todeta arvon suuruudesta riippuen toimineen joko kohtalaisesti tai hyvin. Kaikki arvon .30 alle jäävät arvot puolestaan kertovat, että puhenäytteen toimivuudessa on mahdollisesti ongelmia – se on ehkä epäselvästi muotoiltu tai mittaa jokin muuta ominaisuutta kuin muut puhenäytteet.

Perinteisen osioanalyysin jälkeen puhenäytteitä tarkasteltiin moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta, jolloin tunnuslukuina käytettiin edelleen Infit mean square ja Infit Zstd -arvoja. Rasch-analyysin tunnuslukujen lähtökohtana käytettiin sitä, että väittämien Infit mean square -arvon tulee sijaita arvojen .7 ja 1.3 välillä ja Infit Zstd -arvon rajojen +2 ja -2 sisällä, jotta niitä voitaisiin pitää hyvin toimivina mittarin osioina. Samoin kuin paperilla tehdyn itsearvioinnin ja tehtävää 1 koskevan itsearvioinnin yhteydessä on syytä muistaa, että huonosti toimivien puhenäytteiden kokonaislukumää-

rä voi helposti vaikuttaa suuremmalta Rasch-analyysillä tehdyssä tarkastelussa, koska tarkasteltavia tunnuslukuja on 2 perinteisen osio-testi-korrelaation sijaan.

Taulukkojen 13 ja 14 jälkeen esitellään ensin molempien testiteorioiden näkökulmasta joitakin hyvin toimineita väittämiä, minkä jälkeen tarkasteluvuorossa ovat instrumentissa huonosti toimineet väittämät molempien osio-analyysien näkökulmasta.

TAULUKKO 13

Kuunneltuihin puhenäytteisiin perustuva itsearviointi: perinteisen osioanalyysin mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 2).

Näytteen taitotaso	Huonosti toimineet näytteet	Osio-testi -korrelaatio	Hyvin toimineet näytteet	Osio-testi -korrelaatio
<b>A1 (3)</b>	1 (203-6)	.22		
A1			10	.54
A1			2	.66
<b>A2 (7)</b>			9	.61
A2			13	.60
A2			20	.54
A2			5	.48
A2			14	.47
A2			19	.40
A2			16	.34
<b>B1 (9)</b>			11	.59
B1			22	.58
B1			3	.57
B1			7	.54
B1			17	.48
B1			12	.47
B1			15	.39
B1			6	.60
B1			23	.61
<b>B2 (1)</b>			4	.41
<b>C1 (2)</b>			8	.49
C1	21 (14-195)	.22		
<b>C2 (1)</b>	18 (9-200)	.07		

TAULUKKO 14 Kuunneltuihin puhenäytteisiin perustuva itsearviointi: modernin osioanalyysin (Rasch-analyysi) mukaiset hyvin ja huonosti toimineet puhenäytteet (tehtävä 2).

Näytteen taitotaso	Huonosti toimineet näytteet	Infit mean square	Infit Zstd	Hyvin toimineet näytteet	Infit mean square	Infit Zstd
<b>A1 (3)</b>						
A1	1 (203-6)	1.42	1.2	10	.87	-.9
A1	2 (144-64)	.72	-3.1			
<b>A2 (7)</b>						
A2				9	.90	-1.2
A2				13	.88	-1.3
A2				20	1.05	.6
A2				5	1.06	.5
A2				14	.82	-1.1
A2				19	.99	.0
				16	1.14	.8
<b>B1 (9)</b>						
B1				11	.95	-.5
B1				22	.95	-.5
B1				3	.88	-1.2
B1				7	1.07	.9
B1				17	1.14	1.5
B1				12	1.17	1.7
B1				15	1.09	.7
B1				6	.85	-1.5
				23	.85	-1.6
<b>B2 (1)</b>						
				4	.97	-.2
<b>C1 (2)</b>						
C1				8	.87	-1.0
				21	1.21	.9
<b>C2 (1)</b>	18 (9-200)	1.43	1.4			

### 5.1.3.3 Perinteisen osioanalyysin ja Rasch-analyysin mukaiset hyvin toimineet puhenäytteet

Myös tehtävän 2 puhenäytteet toimivat valtaosin hyvin sekä perinteisen osioanalyysin että Rasch-analyysin näkökulmasta. Otan nyt yksityiskohtaisempaan tarkasteluun joitakin eritasoisia näytteitä, jotka ovat toimineet hyvin molemmissa analyyseissä. Tarkastelun tarkoituksena on yrittää selvittää, sisältävätkö kyseiset näytteet mahdollisesti sellaisia elementtejä, jotka selittäisivät niiden hyvän toimivuuden. Tarkastelussa käytetään hyväksi kaikkea tietoa, joka koskee tarkasteltavana olevia näytteitä. Tällaista tietoa on Yki-korpuksen arvioijien arviointitieto, osioiden vaikeutta koskeva tieto, mahdolliset oppijoiden kom-

mentit ja muut näytteen kuuntelun jälkeiset kommentit. Esimerkkinä hyvin toimineista näytteistä tehtävässä 2 ovat näytteet 11, 23 ja 9.

### Näyte 11

(taso B1; korrelaatio .59; Infit Mnsq .95; Infit ZStd -.5; vastausjakauma 91–118)

Kyseessä on näyte, joka kuuluu eniten ”vaikea päättää” -rasteja keränneiden näytteiden joukkoon, ja sen vuoksi näytettä koskeva tarkempi litteraatio löytyy jäljempänä työssä. Yki-korpuksen arvioijat ovat jakautuneet kahteen ryhmään sijoittaessaan tätä näytettä taitotasolle: 3 arvioijaa on sijoittanut näytteen taitotasolle B1 ja 2 arvioijaa tasolle A2. Kun tarkastellaan tehtävää 2 koskevien osioiden vaikeustasoa (kuinka vaikeana oppijat ovat pitäneet kyseisiä näytteitä), voidaan todeta, että näyte 11 pysyy hyvin tasollaan muiden B1-tasoisten näytteiden joukossa (kuvio 7).

Kyseessä on näyte, jonka puhuja puhuu tasaisesti ja melko varman tuntuisesti ja jonka ääntäminen on hyvää. Valitettavasti viralliseen ammatilliseen kontekstiin liittyvään esittelytilanteeseen täysin sopimattoman sanan käyttö (*copine* = kaveri p.o. kollega) vaikuttaa negatiivisesti kuvaan, joka näytteestä jää.

### Näyte 23

(taso B1 ; korrelaatio .61; Infit Mnsq .85; Infit ZStd -1.6; vastausjakauma 79–129)

Näytteen numero 23 tasossa Yki-korpuksen arvioijien välillä on ollut vaihtelua. Kaksi arvioijaa on sijoittanut näytteen tasolle A2, kaksi tasolle B1 ja yksi arvioija jopa tasolle B2. Onkin todennäköistä, että näyte 23 sisältää joitain sellaisia piirteitä, jotka vaikeuttavat sen arviointia tai ainakin aiheuttavat arvioijien mielipiteiden jakautumista. Kun tarkastellaan näytteiden vaikeustasoa koskevaa analyysiä, voidaan kuitenkin todeta, että oppijoilla ei näytä olleen vaikeuksia kyseisen näytteen vaikeustason kanssa. Oppijat ovat pitäneet näytettä B1-tasoisena, sillä se sijoittuu hyvin B1-tasoisten näytteiden joukkoon (kuvio 7).

Kyse on puhujasta, joka saa esittelyn tehtyä melko sujuvasti, vaikka tuotoksessa onkin sekoitettu tiettyjä sanoja keskenään (*collège/collègue*). Ääntäminen on hyvää. Tuotos sisältää joitakin kielioppivirheitä.

### Näyte 9

(taso A2 ; korrelaatio .61; Infit Mnsq .90; Infit ZStd -1.2; vastausjakauma 107–102)

Yki-korpuksen arvioijien arvioinnit näytteestä 9 ovat jälleen jakautuneet kahteen ryhmään: kolme arvioijaa on pitänyt näytettä A2-tasoisena, kun taas kahden arvioijan mielestä näyte on tasoa B1. Jäljempänä näytteiden vaikeustasoa koskevaa analyysiä tarkasteltaessa voidaan havaita, että näyte 9 sijoittuu analyysissä aivan A2-tason näytteiden ryhmän yläreunaan tai itse asiassa jopa B1-tason näytteiden joukkoon. Oppijat ovat siten pitäneet näytettä vaikeustasoltaan pikemminkin B1- kuin A2-tasoisena näytteenä.



Näytteen 9 puhujan ääntäminen on melko hyvää ja selkeää, mutta hän tuottaa puhetta hieman katkonaisesti. Myös tämän näytteen yleiskuvaa häiritsee se, että puhuja ei nähtävästi ole muistanut oikeaa *esitellä*-verbiä (*présenter*) ranskaksi, vaan käyttää englannin kielestä muunnettua *introducter*-verbiä. Eriytyisesti jotkut ylimpien tasojen oppijat ovat pitäneet väärän verbin käyttöä pahana virheenä ("*tyhmän kuuloista*"). Lisäksi puhujan tuotoksessa esiintyy muutama muukin sanaston ja kieliopin hallintaan liittyvä ongelma, jotka ovat kiinnittäneet ylimpien tasojen oppijoiden huomiota.

Yhteenvetona tehtävän 2 hyvin toimineista näytteistä voidaan todeta, että ne edustavat hyvin tyyppillistä B1-tason suullista osaamista (näytettä 9 voidaan hyvin tulkita B1-tasoisena näytteenä). Näytteiden ääntäminen on selvästi ymmärrettävää, vaikkakin silloin tällöin saattaa esiintyä ei-kohdekielenomaista ääntämistä. Kieliopillisista ja sanastollisista kömmähdyksistä huolimatta näytteiden puhujien viesti tulee ymmärretyksi. Kaikkien näytteiden puhujat ovat selvillä tärkeimmistä kohteliaisuussäännöistä, mutta sanaston hallintaan liittyvät vaikeudet aiheuttavat joskus ongelmia.

#### 5.1.3.4 Yleistä perinteisen osioanalyysin ja Rasch-analyysin mukaisista huonosti toimineista puhenäytteistä

Tehtävän 2 huonosti toimineita puhenäytteitä koskevat huomautukset ovat varsinkin perinteisen osioanalyysin näkökulmasta tarkasteltuina osittain samoja kuin edellä tehtävässä 1. Tehtävän 23 näytteestä huonosti toimineet puhenäytteet sijoittuvat taitotasoasteikon ääripäihin eli tasoille A1 (1 kpl) ja C1 (1 kpl) sekä C2 (1 kpl). Huonosti toimineet näytteet ovat näyte 18, 21 ja 1. Näytteen 18 korrelaatiokerroin on erittäin alhainen (.07), ja molempien näytteiden 21 ja 1 korrelaatiokerroin on .22.

Rasch-analyysin tarkastelun pohjalta tehtävässä 2 on vain kaksi näytettä (näytteet 1 ja 18), joiden Infit mean square -arvo ylitti suositellun 1.3 raja-arvon ja yhden näytteen Infit Zstd -arvo oli hyväksyttävän alarajan -2 alapuolella. Eriytyisesti on huomattava, että minkään tehtävän 2 puhenäytteen Rasch-analyysissä tarkastellut tunnusluvut eivät molemmat ole samanaikaisesti mahdollisista toimimattomuusongelmista ennustavien suositeltavien raja-arvojen ulkopuolella. Tehtävän 2 Rasch-analyysi ei tällä kertaa osoittanut ongelmallisia näytteitä olevan lainkaan mittarin keskitasoilla eli tasoilla A2, B1, B2 eikä myöskään taitotasolla C1. Sen sijaan perinteisessä osioanalyysissä ongelmallisia näytteitä ei ollut tasoilla A2, B1 ja B2, joten perinteisen osioanalyysin esille nostamat ongelmallisilta vaikuttavat osiot sijoittuvat edelleen mittarin ääripäihin.

Tehtävässä 2 huomionarvoista on se, että tällä kertaa sekä perinteinen osioanalyysi että Rasch-analyysi nostavat molemmat esiin mittarin ongelmallisina osioina näytteet 1 ja 18. Tämän lisäksi analyysit nostivat esille vielä kaksi eri näytettä, näytteet 21 (perinteinen analyysi) ja 2 (Rasch-analyysi). Perinteisessä osioanalyysissä huomio kiinnittyy näytteen 21 alhaiseen korrelaatiokerrotimeen (.22). Rasch-analyysin mukaan näytteen 2 Infit Zstd -arvo (-3.1) on suositellun raja-arvon ulkopuolella, mutta samalla sen Infit mean square -arvo on kuitenkin hyväksyttävissä rajoissa (.72). Kaikkien näiden

kommenttien jälkeen todettakoon vielä, että molempien analyysien esille nostamat mahdollisesti ongelmalliset näytteet edustavat tällä kertaa kaikki taitotasoasteikon ääripäitä: näyte 1 (A1), näyte 18 (C2), näyte 21 (C1) ja näyte 2 (A1).

Seuraavaksi tarkastellaan kolmea huonosti toiminutta näytettä (näytteet 18, 1, ja 21). Näytteiden tarkastelun pohjaksi käytössä on pääosin neljänlaista tietoa: 1. tilastollisista analyyseistä saatu mittalukuihin pohjautuva tieto, 2. Yki-korpuksen arvioijilta saatu puhenäytteiden tasoa koskeva tieto, 3. mahdolliset osallistujien kommentit näytteistä ja 4. näytteiden vaikeustasoa koskeva tieto.

### **Näyte 18**

(taso C1; korrelaatio .07; Infit Mnsq 1.43; Infit Zstd 1.4 ; vastausjakauma 9–200 )

Näytteen 18 korrelaatiokerroin on koko lomakkeen heikoin (.07). Yki-korpuksen arvioijien arviot näytteestä jakautuvat kahteen ryhmään: kolme arvioijaa on sijoittanut näytteen tasolle C2 ja kaksi tasolle C1. Valtaosa tutkimukseen osallistujista (200) on valinnut "ei"-vaihtoehdon näytteen kuunneltuaan, mikä aiheuttaa sen, että variaatio oppijoiden valinnoissa on ollut varsin vähäistä. Yhdeksän oppijaa on vastannut tämän näytteen kohdalla "kyllä", eli he ovat täten ilmaisseet selviytyneensä paremmin tehtävästä kuin C1-tasoisien näytteen puhuja. Tämän vuoksi voitaisiin olettaa, että kyseessä olisivat olleet kielitaidoltaan C2-tasoiset oppijat. Näytteen erityisen matala korrelaatio ja erityisesti korkea mean square -arvo (1.43) kertoo kuitenkin siitä, että vastaajat ovat olleet siinä mielessä "väärä", että heidän ei olisi kielitaitonsa puolesta pitänyt vastata selviytyneensä tehtävästä paremmin kuin näytteen puhuja.

Näytteen vaikeustasoa tarkasteltaessa voidaan havaita, että oppijat ovat sijoittaneet näytteen 18 kuuluvaksi taitotasolle C2 eli tasoa ylempänä kuin näytteen määritelty taitotaso on.

Osallistujien kommenttien ja näytteen kuuntelun pohjalta on todettava, että kyseessä on näyte, jonka puhuja on erittäin sujuva ja erityisen luonteva. Hän hallitsee hyvin idiomaattisen kielen ja puhekielen ilmaukset sekä kohteliaisuussäännöt. Lisäksi puhujan sanaston käyttö on johdonmukaisesti virheetöntä, ja hänellä on ranskan kielelle ominainen intonaatio ja ääntäminen.

### **Näyte 1**

(taso A1; korrelaatio .22; Infit Mnsq 1.42; Infit Zstd 1.2 ; vastausjakauma 203–6 )

Kyse on alinta taitotasoa A1 edustavasta näytteestä, jonka tasosta kaikki viisi arvioijaa ovat olleet yksimielisiä. Näytteen puhuja on epäröivä ja hänen puheensa kuulostaa katkonaiselta sekä siltä, että hänellä on vaikeuksia tuottaa puhuttua ranskan kieltä. Melkein kaikki tutkimukseen osallistujat (203 kpl) ovat valinneet tämän näytteen kuunneltuaan "kyllä"-vaihtoehdon eli he ovat kokeneet selvinneensä paremmin tehtävästä kuin näytteen puhuja. Tästä näyte ei siten erottele oppijoita, ja myös sen variaatio on vähäistä.

On syytä huomioida, että tämä näyte kuuluu niihin näytteisiin, joita myöhemmin työssä tarkastellaan "vaikea päättää" -rastien lukumäärien valossa. Siten tätä näytettä koskeva litteraatio löytyy jäljempänä tutkimuksen siitä osasta, jossa käsitellään tehtävään 2 liittyviä "vaikea päättää" -rasteja keränneitä näytteitä.

Oppijoiden kommentit näytteestä 1 paljastavat, että oppijoilla ei ole yleisesti ollut minkäänlaista ongelmaa päättää, suoriutuivatko he tehtävästä paremmin vai huonommin kuin näytteen 1 puhuja. Vain kaksi henkilöä kaikista tutkimukseen osallistuneista on merkinnyt rastin "vaikea päättää" -sarakeeseen. Suurin osa (14 kommenttia) näytettä kommentoineista oppijoista on kiinnittänyt huomiota näytteessä joko siihen, että sen puhe ei ole sujuvaa (katkeava puhe, takeltelu sanoissa, tökkivä jne.) tai siihen, että näytteen puhuja ei onnistu lainkaan suorittamaan tehtävässä edellytetyä kahden henkilön toisilleen esittelyä (10 kommenttia).

Näytteen vaikeustasoa tarkastelevassa tutkimuksen kohdassa jäljempänä näyte 1 sijoittuu kaikista tehtävän 2 näytteistä alimmaiseksi hyvin kauas seuraavasta A1-tason näytteestä.

Puhuja on epäroivä ja hänellä on selviä vaikeuksia suoriutua oikeastaan millään tasolla tehtävänannon mukaisesta esittelystä. Puhujan tuotos ei etene, eikä hän nähtävästi löydä sanoja, joilla esittelyn saisi hoidettua.

### **Näyte 21**

(taso C1; korrelaatio .,22; Infit Mnsq 1.21; Infit Zstd .9 ; vastausjakauma 14–195 )

Näytteen taso on määritelty tasolle C1 arvioijien enemmistön sijoitettua sen tälle tasolle. On kuitenkin korostettava, että Yki-korpuksen viiden arvioijan välillä on ollut erimielisyyttä tämän näytteen taitotasosta: kolme viidestä arvioijasta on sijoittanut näytteen tasolle C1, mutta kaksi arvioijaa on sijoittanut saman näytteen tasolle B1. On luonnollisesti merkille pantavaa, että tasot, joille arvioijat ovat näytteen sijoittaneet, eivät ole vierekkäisiä tasoja, mikä tekee arvioinnista vähemmän luotettavaa. Vaikka näytteen taitotaso on määritelty enemmistöperiaatteen mukaisesti tasolle C1, on todennäköistä, että näyte sisältää joitakin sellaisia piirteitä, jotka ovat aiheuttaneet näytteen kuuntelijoille vaikeuksia arvioida tätä näytettä. Tämä toteamus voi siten koskea sekä Yki-korpuksen arvioijien arviointeja että tutkimukseen osallistuneiden oppijoiden itsearviointia suhteessa kyseisen näytteen tasoon. Kun katsotaan tämän näytteen vaikeutta oppijoiden kokemana, voidaan todeta, että näyte sijoittuu taitotasolle C1.

Näytteen 21 puhujan ääntäminen ja intonaatio ovat erittäin hyviä, ja hänen tuotoksensa on sujuvaa. Puhuja esittelee toisilleen kollegansa ja tärkeän vieraan kohteliaasti ja oikeaoppisesti. Tuotos on virheetöntä eikä sisällä sanastollisia ongelmia. Siten voidaankin todeta, että arvioijien enemmistön päätös siitä, että näyte edustaa taitotasoa C1, vaikuttaa oikeansuuntaiselta.

Taulukkoon 15 on listattu tehtävän 2 osalta kuunneltaviin puhenäytteisiin perustuvan itsearviointin mahdollisesti ongelmallisia osioita sekä perinteisen osioanalyysin että moderniin testiteoriaan perustuvan Rasch-analyysin perus-

teella. Vertailuarvoina on käytetty korrelaatiokerrointa (perinteinen osioanalyysi) ja Infit mean square -arvoa (moderniin testiteoriaan perustuva Rasch-analyysi).

Taulukossa esiteltyjä osio-testi-korrelaatioita tarkasteltaessa voidaan havaita, että näytteiden 1 ja 21 korrelaatioarvot jäävät yleisesti hyväksyttävänä pidetyn 0.30 raja-arvon alle. Nämä molemmat korrelaatioarvot ovat kuitenkin suhteellisen lähellä hyväksyttävän rajaa, kun taas näyte 18 korrelaatioarvo on hyvin heikko eli .07. Heikon korrelaatiokertoimen lisäksi näytteen 18 Infit mean square -arvo on 1.43 eikä siten pysy hyväksyttävissä rajoissa (.7–1.3). Vaikka näytteen 18 Infit zstd -arvo puolestaan pysyykin suositeltavissa rajoissa (–2–+2), on todettava, että kahden muun tunnusluvun heikkouden pohjalta näytteen toimivuutta mittarin osana on syytä epäillä.

Näytteen 1 Infit mean square -arvo (1.42) ylittää niinkään epäilyttävän osion raja-arvon 1.3, ja siten voidaan pitää todennäköisenä, että myöskään tämä näyte ei kaikilta osin sovi mittariin.

Lopuksi tarkastellaan vielä tehtävän 2 kuunneltaviin puhenäytteisiin perustuvan itsearviointilomakkeen näytteiden ”vaikeutta” vertaamalla niiden ennalta määritellyjä vaikeustasoja oppijoiden vastauksiin.

TAULUKKO 15

Mahdollisesti ongelmalliset puhenäytteet perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta (tehtävä 2).

ONGELMALLISET OSIOT				
	Perinteinen osioanalyysi		Moderniin testiteoriaan perustuva Rasch-analyysi	
	Näyte	Korrelaatio	Näyte	Infit mean square
Itsearviointi:	18	.07	18	1.43
Kuunneltavat näytteet, tehtävä 2	1	.22	1	1.42
	21	.22		

### 5.1.3.5 Puhenäytteiden vaikeustason tarkastelua

Seuraavaksi tarkastellaan tehtävän 2 puhenäytteiden vaikeustasoa. Kuvioista 7 ilmenee jo silmämääräisesti tarkastellen, että mittarin osiot keskittyvät asteikon alatasoille (A1–B1). Kun tarkastellaan yksittäisiä osioita, voidaan havaita, että oppijat ovat pitäneet monia näytteitä vaikeampina kuin mitä niiden määritelly taitotasoa on. Esimerkkeinä tällaisista näytteistä voidaan mainita näytteet 10 (A1), 2 (A1) ja 15 (B1). A1-tasoiset näytteet numero 10 ja 2 sijoittuvat A2-tasoisien näytteiden ryhmään. B1-tasoisista näytettä 15 oppijat ovat selvästikin pitäneet huomattavasti korkeammalle taitotasolle kuin B1 kuuluvana näytteenä, sillä näyte sijoittuu asteikolla B2–C1-tasoisien näytteiden välimaastoon. Koska näytteet 10 ja 15 kuuluvat molemmat niihin näytteisiin, joita myöhemmin työssä tarkastellaan niiden keräämien ”vaikea päättää” -rastien avulla ja

joista siten on olemassa paljon kommentteja, käsitellään poikkeuksellisesti jo tässä yhteydessä näitä näytteitä oppijoiden kommenttien valossa.

### Näyte 10

(taso A1; korrelaatio .54; Infit Mnsq .87; Infit Zstd -.9 ; vastausjakauma 171–37)

Näyte on saanut 15 tutkimukseen osallistujan merkitsemään rastin "vaikea päättää" -sarakeeseen. "Vaikea päättää" -rastien vähäinen määrä kertoo siitä, että valinta "kyllä"- ja "ei"-vaihtoehtojen välillä kuuntelutilanteessa ei ole ollut kovin vaikeaa valtaosalle tutkimukseen osallistuneista oppijoista. Sen sijaan näytteen joitakin piirteitä on kommentoitu melko runsaasti. A1-tasoinen näyte on saanut oppijoilta eniten kommentteja (17 kpl) jälleen puheen sujuvuuden puutteeseen liittyen. Huomiota sujuvuuden puutteen lisäksi kiinnitettiin muun muassa näytteen puhujan hitaaseen puheen tuottamiseen ja takelteluun. Sanaston valintaan liittyvät kommentit olivat seuraavaksi yleisimpiä (14 kpl); valtaosa näistä kommentteista kohdistuu siihen, että puhuja on erehtynyt käyttämään väärää verbiä esitellessään henkilöitä toisilleen ("*Je vous invite*" po. "*Je vous présente*"). Näytteen puhujan ääntämistä on myös kommentoitu paljon (13 kommenttia); ääntämistä on pidetty muun muassa ongelmallisena, huonona ja englannin kielen värittämänä.

### Näyte 15

(taso B1; korrelaatio .39; Infit Mnsq 1.09; Infit Zstd .7 ; vastausjakauma 38–171 )

Näyte 15 kuuluu niiden kolmen tarkastellun näytteen ryhmään, joihin tutkimukseen osallistujat ovat merkinneet eniten "vaikea päättää" -rasteja. Tämä näyte on kerännyt eniten oppijoiden kommentteja (49 kpl) sekä runsaasti "vaikea päättää" -rasteja (38 kpl) koko tarkastellun aineiston osalta. Huomionarvoista on, että 38:sta "vaikea päättää" -rastista huolimatta osallistujat eivät selvästikään ole osanneet eritellä, mikä piirre tai mitkä piirteet näytteessä aiheuttivat sen, että heidän oli vaikea valita "kyllä" tai "ei" -vaihtoehtojen väliltä. Tästä epävarmuudesta kertoo se, että 30 "vaikea päättää" -rastia 38:sta on jätetty täysin ilman kommenttia siitä, miksi osallistujilla on ollut vaikea päättää, selviytykö itse paremmin tilanteesta kuin näytteen puhuja vai ei.

Oppijoiden kommentit keskittyvät jälleen sujuvuuden kommentointiin, vaikka niitä onkin huomattavasti vähemmän (14 kpl) kuin ilman selitystä jääneitä "vaikea päättää"-rasteja (30 kpl). Ilmeisesti tämän B1-tasoinen näytteen puhujan sujuvuus ja verraten hyvä ääntäminen (10 kommenttia) sekä melko hyvin sujunut henkilöiden toisilleen esittely (8 kommenttia) ovat aiheuttaneet päänvaivaa osallistujille siinä mielessä, että näytteen puhujan suullisen kielitaidon taso on suhteellisen samantasoinen kuin valtaosan tutkimukseen osallistuneista.

Yllä tarkasteltuja näytteitä oppijat ovat siis pitäneet vaikeampina kuin niiden ennalta määriteltä taitotaso on. Niiden lisäksi esimerkkinä taitotasoaan

helpompana pidetystä näytteestä voidaan mainita näyte 22 (B1), joka sijoittuu A2-tasoisten näytteiden joukkoon.

Mittaria tarkasteltaessa voidaan todeta myös, että ylimpien tasojen näytteiden vähyys saattaa olla jonkin verran ongelmallista mittarin kokonaisuuden kannalta; mittari sisältää 2 näytettä tasolta C1 sekä yhden näytteen sekä tasolta B2 että tasolta C2.

```

Person - MAP - Item
<more>|<rare>
6      . +
      ## |
      |T t2_18_C2
5      +
      ## T | t2_21_C1
4      +
      .### |
3      ##### + t2_4_B2
      |      t2_15_B1
      |S t2_8_C1
      .#### S
2      .##### +
      .### |
      |      t2_3_B1 t2_6_B1
1      ##### + t2_17_B1 t2_23_B1
      ##### |
      .##### | t2_11_B1
      |      t2_7_B1
      ##### M | t2_9_A2
0      . +M
      ##### | t2_20_A2
      .##### |
      ##### | t2_22_B1
-1     +      t2_13_A2
      ##### | t2_12_B1 t2_2_A1
      .### |
-2     ##### S+ t2_5_A2
      .##### |
      |      t2_10_A1
      |S
-3     .# + t2_14_A2
      |      t2_16_A2 t2_19_A2
      ## |
-4     +
      .# T |
-5     +
      |T
      .## | t2_1_A1
-6     +
      <less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

```

KUVIO 7 Tehtävän 2 puhenäytteiden vaikeus oppijoiden vastausten perusteella.

### 5.1.3.6 Yhteenvetoa tehtävän 2 itsearviointimittarin luotettavuudesta

Yhteenvetona tehtävän 2 kuunnelluista puhenäytteistä (23 kappaletta) voidaan todeta, että puhenäytteet näyttävät toimivan itsearviointitarkoituksessa kokonaisuutena suhteellisen hyvin. Osio-testi-korrelaatioiden tasosta noin 48 % sijoittui välille .54–.66 ja 39 % välille .34–.49. Siten noin 87 % tehtävässä 2 käytetyn itsearviointimittarin osioista ylitti suositeltavan tason .30. Kolme (13 %) tämän mittarin osioista jäi alle rajan .30. Hyvin pienten (alle .20 jäävien) osio-testi-korrelaatioiden määrä tehtävässä 2 on yksi kappale.

Winstepin Infit Mean square -arvoista valtaosa eli noin 91 % oli hyväksyttävissä rajoissa eli arvojen .7 ja .1.3 välillä. Kahden näytteen (näytteet 1 ja 18) kohdalla oli epäilystä niiden sopivuudesta mittarin taustalla olevaan tilastolliseen malliin.

### 5.1.4 Yhteenvetoa kolmen itsearviointimittarin luotettavuudesta

Yhteenvetona kaikkia kolmea itsearviointimittaria koskien tarkastelen lopuksi seuraavia arvoja ja tunnuslukuja kolmen eri taulukon avulla:

1. perinteinen osioanalyysi: reliabiliteetti ja osio-testi-korrelaatioiden osuudet ja kappalemäärät suuruusluokittain (taulukko 16)
2. Rasch-analyysi: reliabiliteetti ja Infit mean square -arvojen osuudet ja kappalemäärät suuruusluokittain (taulukko 17)
3. ongelmalliset osiot perinteisen osioanalyysin ja moderniin testiteoriaan pohjautuvan osioanalyysin näkökulmasta (taulukko 18).

Jokaista taulukkoa kommentoidaan lyhyesti heti taulukon jälkeen.



TAULUKKO 16 Perinteinen osioanalyysi: kolmen itsearviointimittarin reliabiliteetti ja osio-testi-korrelaatiot (.30 ylittävien prosenttiosuus, kappalemäärät ja prosenttiosuudet suuruusluokittain).

Itsearviointityyppi	Cronbachin alfa	.30 ylittävien osio-testi-korrelaatioiden osuus (%)	Osio-testi-korrelaatiot		
			suuruusluokittain	kappalemäärittäin	%-osuuksittain
Paperilla tehty	.94	85	.10	4	10
			.20	2	5
			.30	5	12,5
			.40	5	12,5
			.50	11	27,5
			.60	11	27,5
			.70	2	5
			.80	-	-
			.90	-	-
			Kuunnellut näytteet, tehtävä 1	.94	85
.20	4	15,4			
.30	1	3,8			
.40	3	11,5			
.50	1	3,8			
.60	10	38,5			
.70	7	27,0			
.80	-	-			
.90	-	-			
Kuunnellut näytteet, tehtävä 2	.89	87			
			.10	-	-
			.20	2	8,7
			.30	2	8,7
			.40	7	30,4
			.50	6	26
			.60	5	21,7
			.70	-	-
			.80	-	-
			.90	-	-

Tutkimuksessa käytettyjen kolmen eri itsearviointilomakkeen reliabiliteetti on hyvä ja samansuuntainen, sillä reliabiliteettikertoimien arvo on alimmillaankin .89 (tehtävä 2, kuunnellut näytteet), ja kahden muun itsearviointilomakkeen (paperilla tehty itsearviointi ja tehtävä 1, kuunnellut näytteet) reliabiliteetti on .94.

Osio-testi-korrelaatioita koskevan hyväksyttävän raja-arvon .30 ylittävien osioiden osuus on korkea kaikissa kolmessa lomakkeessa. Tässä yhteydessä on

kuitenkin syytä huomioida, että eri lomakkeiden välillä on eroja toisaalta siinä, kuinka korkeita korrelaatioiden arvot ovat sekä toisaalta siinä, mikä on korkeimpien korrelaatioiden lukumäärä. Esimerkiksi tehtävässä 1 on useita arvon .70 ylittäviä korrelaatiokertoimia (7 kpl), ja myös paperilla tehdyssä itsearviointilomakkeessa on muutama (2 kpl) tämän arvon ylittävä korrelaatiokerroin, kun taas tehtävän 2 korkein kerroin sijaitsee tasolla .66. Siten tehtävässä 2 ei ole yhtäkään osiota, jonka korrelaatiokerroin ylittäisi .70:n. Korkeat korrelaatiot erityisesti tehtävässä 1 osoittavat omalta osaltaan, että tämänkaltainen konkreettisiin puhenäytteisiin perustuva itsearviointi toimii paremmin kuin paperilla tehty itsearviointi.

Niiden osioiden kokonaismäärä, joka ylitti perinteisen osioanalyysin suositellun raja-arvon .30 on sama eli 85 % sekä tehtävässä 1 että paperilla tehdyssä itsearviointilomakkeessa. Tehtävässä 2 rajan .30 ylittäviä osioita on 87 % eli hieman enemmän kuin kahdessa muussa itsearviointilomakkeessa, mutta tätä arvoa tarkastellessa on pidettävä mielessä, että korrelaatioarvot olivat yleisesti ottaen hieman alempia tehtävässä 2 kuin paperilla tehdyssä itsearvioinnissa ja tehtävässä 1. Paperilla tehdyn itsearvioinnin ja tehtävään 1 liittyvän itsearvioinnin prosentuaalisesti suurin korrelaatiokertymä sijoittuu tasoille .60 (38,5 %; tehtävä 1) ja .50 ja .60 (27,5 %; paperilla tehty itsearviointi). Sen sijaan tehtävä 2:ta koskeva prosentuaalisesti suurin korrelaatiokertymä jää alemmaksi kuin edellä (.40; 30,4 %).

TAULUKKO 17

Rasch-analyysi: kolmen itsearviointimittarin reliabiliteetti ja Infit mean square -arvot (.70–1.30 välille sijoittuvien osioiden prosenttiosuus, kappalemäärät ja prosenttiosuudet suuruusluokittain).

Itsearviointityyppi	Cronbachin alfa	Infit mean square			
		.70 -1.30 välille sijoittuvien osioiden %-osuus	suuruusluokittain	kappalemäärittäin	%-osuuksittain
Paperilla tehty	.94	92,5	.70	8	20,0
			.80	10	25,0
			.90	8	20,0
			1.0	5	12,5
			1.1	4	10,0
			1.2	1	2,5
			1.3	2	5
			1.4	1	2,5
			1.6	1	2,5
Kuunnellut näytteet, tehtävä 1	.94	88,4	.60	1	3,8
			.70	4	15,4
			.80	4	15,4
			.90	5	19,2
			1.0	7	27,0
			1.1	2	7,7
			1.2	1	3,8
			1.3	1	3,8
			1.4	1	3,8
Kuunnellut näytteet, tehtävä 2	.89	91,3	.70	1	4,3
			.80	7	30,4
			.90	5	22,0
			1.0	4	17,3
			1.1	3	13,0
			1.2	1	4,3
			1.4	2	8,7

Kun tarkastellaan yllä olevaa taulukkoa, voidaan havaita, että kaikissa kolmessa itsearviointimittarissa prosenttiosuus on Rasch-analyysin mukaan suuri niillä osioilla, joiden Infit mean square -arvo sijoittui välille .70–1.30. Kolmesta käytetystä itsearviointilomakkeesta paperilla tehdyssä itsearvioinnissa on prosentuaalisesti eniten (92,5 %) niitä osioita, jotka sijoittuvat edellä mainittujen raja-arvojen sisään. Seuraavaksi suurin edellä mainitulle välille sijoittuvien näytteiden

den prosenttiosuus on tehtävässä 2 (91,3 %). Kolmanneksi prosentuaalisessa kokonaistarkastelussa jää tehtävää 1 koskeva itsearviointilomake näytteineen (88,4 %).

Tämän toteamuksen jälkeen on kuitenkin jälleen paikallaan huomauttaa, että jos verrataan toisiinsa kolmea itsearviointilomaketta ja sitä, mille tasolle sijoittuu eniten .70–1.30 välille sijoittuvia arvoja, niin tulos on toisenlainen. Tässä tarkastelussa näet tehtävää 1 koskevien näytteiden Infit mean square -arvojen enemmistö (27 %) sijoittuu ihanteellisesti luokkaan 1.0. Jos verrataan vastaavia Infit mean square -arvoja samassa suuruusluokassa (1.0) paperilla tehdyn itsearvioinnin (12,5 %) ja tehtävää 2 koskevien näytteiden osalta (17,3 %), voidaan todeta prosenttiosuuksien olevan pienempiä kuin tehtävässä 1 samassa suuruusluokassa. Sekä paperilla tehdyssä itsearvioinnissa että tehtävää 2 koskevassa itsearvioinnissa suurin Infit mean square -kertymä sijoittuukin suuruusluokkaan .80: paperilla tehty itsearviointi 25 % ja tehtävää 2 koskevat näytteet 30,4 %.

Niiden näytteiden prosenttiosuudet, jotka eivät sijoitu hyväksyttävien raja-arvojen .70–1.30 välille, sijoittuvat 7,5 %:n ja 11,5 %:n välille kolmessa eri mitarissa seuraavasti: paperilla tehty itsearviointi 7,5 % (3 kpl), kuunnellut näytteet (tehtävä 2) 8,7 % (2 kpl) ja kuunnellut näytteet (tehtävä 1) 11,5 % (3 kpl).

Taulukkoon 18 on kerätty vertailun vuoksi tiedot eri itsearviointilomakkeiden ongelmallisista osioista sekä perinteisen osioanalyysin että moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta. Taulukosta käy esille se tosiasia, että perinteisen ja Rasch-analyysin tunnusluvut tarkastelevat samaa aineistoa eri näkökulmista. Tämän vuoksi niiden paljastamat mahdollisesti mitarin kokonaisuuden kannalta ongelmalliset osiot voivat olla eri osioita. Alla olevassa tarkastelussa perinteisen osioanalyysin tunnuslukua edustaa korrelaatiokerroin, jonka tulee siis olla vähintään 0.30, jotta osio olisi hyväksyttävä. Rasch-analyysin tunnuslukuna on käytetty jo aiemmin työssä käytettyä Infit mean square -arvoa, jonka siis tulisi asettua .7–1.3 välille ollakseen hyväksyttävä. Taulukosta on jätetty tarkoituksella pois paperilla tehdyn itsearvioinnin 2 osiota: osio 29, joka poistettiin analyyseistä sen vuoksi, että molempien testiteorioiden mukaiset analyysit osoittivat kyseisen osion olevan erittäin ongelmallinen (negatiivinen korrelaatiokerroin (-.51); Infit mean square -arvo 3,32). Samoin tästä tarkastelusta pudotettiin pois osio 20, jossa ei ollut minkäänlaista vaihtelua.

TAULUKKO 18

Kolmen itsearviointimittarin ongelmalliset osiot perinteisen osioanalyysin ja moderniin testiteoriaan perustuvan Rasch-analyysin näkökulmasta.

ONGELMALLISET OSIOT				
Itsearviointityyppi	Perinteinen osioanalyysi		Moderniin testiteoriaan perustuva Rasch-analyysi	
	Väittämä/näyte	Osio-testi-korrelaatio	Väittämä/näyte	Infit mean square
Paperilla tehty	41	.11	13	.70
	19	.12	36	1.33
	9	.18	23	1.40
	27	.19	5	1.68
	21	.21	19	1.30
	1	.22		
Kuunnellut näytteet, tehtävä 1	7	.22	2	.65
	17	.23	24	1.38
	12	.24	26	1.41
	9	.26		
Kuunnellut näytteet, tehtävä 2	18	.07	18	1.43
	1	.22	1	1.42
	21	.22		

Kuten jo todettiin, eri testiteoriat nostavat tarkasteluun osioiden eri puolia ja tästä johtuen antavat viitteitä mahdollisesti ongelmallisista osioista, jotka eivät kuitenkaan välttämättä ole samoja osioita. Jonkinlaisena nyrkkisääntönä varmasti voidaan pitää sitä, että jos molempien teorioiden ilmaisemat tunnusluvut osoittavat jonkin osion mahdollisen ongelmallisuuden, osiossa on mitä todennäköisimmin jonkintasoinen ongelma.

Molempien testiteorioiden näkökulmista ongelmallisilta vaikuttavia osioita ovat seuraavat: väittämä 19 (paperilla tehty itsearviointi) sekä tehtävää 2 koskevat näytteet 1 ja 18 (tosin on muistettava edelleen, että nämä kaksi viimeksi mainittua osiota ovat asteikon ääripään osioita). Tämän tutkimuksen mittareiden laatua koskevien analyysien tarkoituksena on osoittaa, että mittareiden osiot ovat kokonaisuutena riittävän laadukkaita, jotta niitä voidaan käyttää tässä tutkimuksessa. Työn pääfokus on oppijoiden itsearvioinnin ja ulkoisen arvioinnin vastaavuudessa, ja tästä syystä tässä työssä ei mennä klassisen ja modernin testiteorian osoittamien eroavaisuuksien tarkastelussa erojen esille nostamista pidemmälle.

Tutkimuksessa käytettyjen kolmen eri itsearviointilomakkeen luotettavuutta ja laatua on tarkasteltu sekä perinteisen osioanalyysin että moderniin testiteoriaan perustuvan Rasch-analyysin avulla. Kuten on jo käynyt ilmi, eri testiteorioihin pohjautuvat analyysit painottavat eri asioita ja käyttävät eri tunnuslukuja analyyseissään, ja siksi kolmen käytetyn itsearviointimittarin pa-

remmuusjärjestykseen laittaminen onkin haasteellinen tehtävä. Mikäli aineisto olisi analysoitu pelkästään esimerkiksi perinteisen osioanalyysin menetelmiä ja tunnuslukuja käyttäen, olisi ollut helppo todeta vaikkapa Cronbachin alfan tai .30 ylittävien korrelaatiokertoimien prosenttiosuuksien mukaan, mikä olisi ollut itsearviointitapojen paremmuusjärjestys. Tässä tutkimuksessa Rasch-analyysi muodostaa kuitenkin tärkeän osan perinteisen osioanalyysin rinnalla, ja siten tyydynkin tarkastelemaan lopuksi eri mittareiden luotettavuuteen ja laatuun liittyviä aspekteja yrittämättäkään laittaa lomakkeita keinotekoisesti paremmuusjärjestykseen.

Kaikkien kolmen itsearviointimittarin Cronbachin alfan arvot olivat hyviä. 42-kohtaisen paperilla tehdyn itsearvioinnin ja tehtävää 1 koskevan 26-kohtaisen itsearviointilomakkeen alfan arvot olivat molemmissa .94 sen jälkeen, kun paperilla tehdyn itsearvioinnin lomakkeesta oli poistettu negatiivisen korrelaatiokertoimen omaava näyte 29 ja täysin erottelematon näyte 20. Paperilla tehdyn itsearvioinnin väittämien hyvä toimivuus ei liene yllätys, koska ne perustuvat Viitekehukseen, jonka puhumisen väittämät puolestaan perustuvat empiiriseen tutkimukseen. Tehtävää 2 koskevan 23-kohtaisen itsearviointilomakkeen Cronbachin alfa oli hieman edellisiä alempi (.89), mutta silti hyvä.

Verrattaessa perinteiseen osioanalyysiin kuuluvien .30 ylittävien osioiden korrelaatiokertoimien prosenttiosuuksia keskenään on todettava, että tässäkin suhteessa kolme eri itsearviointimittaria olivat melko tasaisia. Paperilla tehdyn itsearvioinnin ja tehtävää 1 koskevan itsearviointilomakkeen .30 ylittävien osioiden korrelaatioiden prosentuaalinen osuus oli sama eli 85 %, mistä seuraa siis, että näiden lomakkeiden osioista 15 % jäi osioiden toimivuustarkastelussa yleisesti suositellun raja-arvon .30 alle. Tehtävää 2 koskevan itsearviointilomakkeen .30 ylittävien osioiden korrelaatiokertoimien prosentuaalinen määrä oli hieman edellistä suurempi (.87), ja siten tämän lomakkeen osalta raja-arvon .30 alle jäävien osioiden prosenttiosuus on vastaavasti vähän pienempi kuin edellä eli 13 %.

Mutta kuten edellä jo huomautettiin, analyysien tuottamia arvoja on syytä tarkastella hieman laajemmin ja yksityiskohtaisemmin kuin pelkästään raja-arvon .30 ylittävien osioiden kokonaismäärän näkökulmasta. Tämä tarkastelu näet tuo esille myös sen tosiseikan, että eri lomakkeiden osioiden tiettyyn suuruusluokkaan sijoittumisessa on suuria eroja. Esimerkiksi paperilla tehdyssä itsearvioinnissa ja tehtävää 1 koskevassa itsearvioinnissa on molemmissa täsmälleen sama prosenttiosuus osioita, jotka ylittivät .30 raja-arvon (85 %). Tehtävää 1 koskevassa itsearviointilomakkeessa on 65,5 % suuruusluokkaan .60 tai yli sijoittuvia osioita, kun taas samassa suuruusluokassa paperilla tehdyssä itsearviointilomakkeessa on .30 ylittäviä osioita vain 28 %. Edelleen jos tarkastellaan tehtävää 2 koskevan itsearviointilomakkeen .30 ylittävien osioiden suuruusluokkaa samoin kuin edellä eli laskemalla .60 tai sen yli sijoittuneet korrelaatiokertoimet, voidaan todeta näitä olevan tässä lomakkeessa vain 21,7 %. Korrelaatiokertoimien suuruusluokissa todellakin on huomattavia määrällisiä eroja eri lomakkeiden välillä myös kappalemäärittäin asiaa tarkasteltaessa: paperilla tehdyssä itsearvioinnissa .70 ylittäviä korrelaatiokertoimia on vain 2

kappaletta, tehtävää 1 koskevassa itsearviointilomakkeessa niitä on 7 kappaletta ja tehtävää 2 koskevassa itsearviointilomakkeessa ei lainkaan.

Rasch-analyysin tuottaman osioiden Infit mean square -arvon, jonka tulisi sijoittua välille .70–1.30, prosentuaaliset osuudet ovat myöskin melko tasaisia: paperilla tehty itsearviointi 92,5 %, tehtävää 1 koskeva itsearviointilomake 88,4 % ja tehtävää 2 koskeva itsearviointilomake 91,3 %. Jälleen on kuitenkin syytä palauttaa mieliin se, että osioiden suuruusluokkiin sijoittumisissa on eroja jopa välillä .70–1.30. Kuten edellä jo todettiin, tehtävää 1 koskevien näytteiden Infit mean square -arvojen enemmistö (27 %) sijoittuu ihanteellisesti luokkaan 1.0, kun taas vastaavat prosenttiosuudet samassa suuruusluokassa (1.0) kahdessa muussa lomakkeessa ovat pienempiä: paperilla tehdyssä itsearvioinnissa 12,5 % ja tehtävää 2 koskevassa itsearviointilomakkeessa 17,3 %. Edellä tuli lisäksi jo todetuksi, että sekä paperilla tehdyssä itsearvioinnissa että tehtävää 2 koskevassa itsearvioinnissa suurin Infit mean square -kertymä sijoittuu tehtävää 1 koskevaa itsearviointilomaketta alemmaksi eli suuruusluokkaan .80 (paperilla tehty itsearviointi 25 % ja tehtävää 2 koskevat näytteet 30,4 %). Infit mean squaren raja-arvojen .70–1.30 ulkopuolelle sijoittuvien osioiden prosenttiosuudet olivat myös hyvin samansuuntaisia kaikissa kolmessa itsearviointilomakkeessa: paperilla tehty itsearviointi 7,5 % (3 kpl), tehtävää 2 koskevat näytteet 8,7 % (2 kpl) tehtävää 1 koskevat näytteet 11,5 % (3 kpl).

Loppuyhteenvedona todettakoon vielä, että luotettavuutta ja laatua on syytä tarkastella tarpeellisten yhteenvetävien kokonaisprosenttilukujen lisäksi muun muassa juuri eri analyysissä eri suuruusluokkaan sijoittuvien osioiden määrillä. Jos tyydytään vain tarkastelemaan kokonaisprosenttilukuja, moni seikka käytettyjen mittareiden toimivuuden ja laadun osalta saattaa jäädä huomiotta. Lisäksi on muistettava, että odotuksenmukaista on, että asteikon ääripäihin sijoittuvissa näytteissä tulee näkyviin vaihtelun puute vastauksissa. Tämä seikka on suoraan yhteydessä siihen, mikä on tutkimukseen osallistujien taitotasojakauma. Jos tutkimukseen osallistuneissa oppijoissa olisi ollut joitakin alle A1-tasoisia ja samoin lisää C2-tasoisia oppijoita, olisivat myös asteikon ääripäiden tasolle A1 ja C2 sijoittuvien näytteiden tai väittämien korjatun osio-testi-korrelaation ja reliabiliteetin arvot osoittautuneet paremmiksi.

Kuunneltaviin puhenäytteisiin perustuvassa itsearvioinnissa on vielä huomioitava se, että tämänkaltainen kuunteluun perustuva vertailutoiminta vaatii auditiivista erottelukykä. Olisi kiinnostavaa selvittää, onko osallistujien sukupuolella vaikutusta siihen, millainen heidän auditiivinen erottelukykynsä on. Voidaan pohtia myös sitä, olisiko kuunneltaviin puhenäytteisiin perustuvien itsearviointimittareiden luotettavuus osoittautunut erilaiseksi, jos tutkimukseen osallistujissa olisi ollut enemmän miehiä.

Sekä perinteisen osioanalyysin että Rasch-analyysin osoittamien tulosten pohjalta voidaan lopuksi todeta, että kaikki kolme tässä tutkimuksessa käytettyä itsearviointilomaketta osoittavat riittävän vakaata sisäistä konsistenssia, jotta niitä voidaan huoletta käyttää mittainstrumentteina tässä tutkimuksessa.

## 5.2 Ulkoisten arvioijien luotettavuus

Mittainstrumenttien luotettavuustarkastelun jälkeen keskitytään ulkopuolisen arvioinnin luotettavuuden tarkasteluun. Ulkopuolinen arviointi toteutettiin niin, että kaksi henkilöä arvioi kaikkien oppijoiden suoritukset molemmissa tutkimuksessa käytetyissä tehtävissä (209 oppijaa eli 418 puhesuoritusta). Nämä arvioijat olivat olleet mukana jo arvioimassa YKI-testeistä peräisin olevia näytteitä, joiden avulla tutkittavat ranskan kielen oppijat sitten arvioivat omaa suoritustaan. Oppijoiden suorituksille haluttiin saada kolmaskin arvio, mutta osoittautui mahdottomaksi löytää kolmatta ranskaa hyvin osaavaa henkilöä, joka olisi voinut arvioida kaikki suoritukset. Siksi päädyttiin ratkaisuun, jossa yhteensä viisi henkilöä arvioi kukin muutaman kymmenen oppijan suorituksen molemmissa tehtävissä (kukin arvioija arvioi siis eri oppilaita), jolloin yksittäisen arvioijan työtaakka pysyi varsin kohtuullisena. Toiveena oli pystyä yhdistämään nämä viisi arvioijaa analyysivaiheessa ja käsittelemään heidän tuloksiin ikään kuin kyse olisi ollut yhdestä arvioijasta. Kokonaisuutena jokainen suoritus siis kolmoisarvioitiin, mutta kolmas arvioija saattoi vaihdella oppijasta toiseen.

Edellä kuvattuun kolmanteen arviointiin liittyy tietysti monta epävarmuustekijää, joten sen laatu ei ole itsestään selvää. Koska kaikki arvioinnit kytkeytyvät toisiinsa niiden kahden arvioijan välityksellä, jotka arvioivat kaikkien oppijoiden suoritukset, oli mahdollista analysoida arviot ja tarvittaessa poistaa epäluotettavien arviointien vaikutus tuloksiin. Nämä analyysit tehtiin Facets-ohjelmalla, joka mahdollistaa arvioijien yhdenmukaisuuden ja kunkin arvioijan oman johdonmukaisuuden selvittämisen (Linacre 2011).

Arvioijien luotettavuutta tarkastellaan ensin käyttäen hyväksi samoja tunnuslukuja, joita on käytetty edellä tutkimuksen mittainstrumenttien luotettavuutta tarkasteltaessa (Cronbachin alfa ja korjattu osio-testi-korrelaatio). Näiden tunnuslukujen lisäksi arvioijien luotettavuutta selvitellään heidän yksimielisyytensä pohjalta.

### 5.2.1 Kolmen arvioijan luotettavuus, tehtävät 1 ja 2

Kolmen arvioijan välistä reliabiliteettia koskevan Cronbachin alfan arvo on .9413 tehtävässä 1 ja .8536 tehtävässä 2. Arvioijan korrelaatio toisiin arvioijiin sekä arvioijan poistamisen vaikutus arvioinnin luotettavuuteen esitetään taulukossa 19.



TAULUKKO 19 Arvioijan korrelaatio suhteessa muihin arvioijiin ja arvioijan poistamisen vaikutus reliabiliteettiin.

	Korjattu osio-testi-korrelaatio (arvioijan korrelaatio muihin arvioijiin)		Alfan arvo, jos arvioija poistetaan	
	Tehtävä 1	Tehtävä 2	Tehtävä 1	Tehtävä 2
Arvioija 1	,8127	,5610	,9741	,9667
Arvioija 2	,9141	,8385	,8916	,6961
Arvioija 3	,9222	,8139	,8806	,7146

Taulukosta 19 näkyy selvästi, että vaihtuvan arvioijan (arvioija 1) poistaminen nostaisi arviointeja koskevaa reliabiliteetin arvoa huomattavasti varsinkin tehtävässä 2. Jommankumman hyvin toimineen arvioijan (arvioija 2 tai 3) poistaminen aiheuttaisi puolestaan arviointeja koskevan alfan arvon merkittävän laskemisen. Arvioijan korrelaatio muihin arviointeihin vastaa aikaisemmin käsitellyä korjattua osio-testi-korrelaatiota. Se antaa hyvin samanlaisen kuvan kuin alfa-arvojen tarkastelu: arvioija 1, joka on kooste viiden eri arvioijan arvioinneista, korreloi jonkin verran heikommin kahteen muuhun arvioijaan, varsinkin tehtävässä 2, jossa korrelaatio jää .5610:een.

Tämän jälkeen tarkastellaan kolmen ulkoisen arvioijan välisiä korrelaatioita, jotka kertovat myös yleisellä tasolla arvioijien välisestä yhdenmukaisuudesta. Korrelaatiossa on kyse riippuvuudesta kahden muuttujan välillä, ja tavallisia tapoja ilmaista kahden muuttujan välinen riippuvuus on korrelaatiokerroin (Heikkilä 2001, 203). Korrelaatiokerroin vaihtelee  $+1:n$  ja  $-1:n$  välillä siten, että mitä lähempänä kertoimen arvo on  $+1:tä$ , sitä enemmän kaksi tarkasteltua muuttujaa korreloivat keskenään (Chanquoy 2005, 70). Korrelaatiokertoimen avulla osoitetaan siten toisaalta se, onko kahden muuttujan välillä riippuvuutta vai ei, ja toisaalta jos riippuvuutta on, kuinka voimakasta se on ilman, että viitataan minkäänlaiseen kyseisten muuttujien väliseen kausaalisuhteeseen (Chanquoy 2005, 68). Korrelaatioista puhuttaessa käytetään merkitsevyytensä eli riskitasoa. Se ilmoittaa, kuinka suuri sen riski on, että saatu ero tai riippuvuus johtuu sattumasta. Merkitsevyytensä mittaa tehdyn johtopäätöksen tilastollista luotettavuutta. Korrelaatioiden merkitsevyys voidaan testata, ja yleisimmin testauksen jälkeen korrelaatiot ilmaistaan 5 %:n tai 1 %:n merkitsevyydellä (Heikkilä 2001, 194). Tässä tutkimuksessa korrelaatioiden merkitsevyys on ilmaistu yleisellä tavalla seuraavasti:

\* = 5 %:n merkitsevyys

\*\* = 1 %:n merkitsevyys

TAULUKKO 20 Ulkoisten arvioijien väliset korrelaatiot tehtävissä 1 ja 2.

		Tehtävä 1		
		Arv. 1	Arv. 2	Arv. 3
Tehtävä 1	Arv 1	1,000	,796**	,809**
	Arv 2	,796**	1,000	,951**
	Arv 3	,809**	,951**	1,000

		Tehtävä 2		
		Arv. 1	Arv. 2	Arv. 3
Tehtävä 2	Arv 1	1,000	,564**	,540**
	Arv 2	,564**	1,000	,936**
	Arv 3	,540**	,936**	1,000

Taulukot 19 ja 20 osoittavat, että myös silloin, kun arvioijien välisiä korrelaatioita tarkastellaan arvioijapari kerrallaan (osio-testi-tyyppisessä korrelaatiossa kyse oli yhden arvioijan vertaamisesta kaikkiin muihin arvioijiin yhtä aikaa), arvioija 1:n korrelaatiot kahteen muuhun arvioijaan ovat matalimmat erityisesti tehtävässä kaksi. Korrelaatiot, jotka ovat .540 luokkaa, ovat varsin matalia, kun kyse on samojen suoritusten arvioinnista: tällainen korrelaatio merkitsee, että kahden mittauksen (= kahden arvioinnin) selitysaste on vain 29,1 % ( $.540 \times .540 = .291$ ). Ilmaistuna toisin kyseiset arvioinnit perustuvat vain n. 29% osalta samalle varianssille (vaihtelulle) suorituksissa. Arvioijat siis todennäköisesti perustavat arvionsa oppijoiden suoritusten laadusta varsin erilaisille kriteereille.

Yhteenvedon voidaan todeta, että arvioija 1 oli toisten arvioijien kanssa täsmälleen samaa mieltä tasoarviosta vain noin puolessa tapauksista tehtävässä yksi, kun taas arvioijat 2 ja 3 olivat samaa mieltä yli 70% suorituksista. Tehtävässä kaksi arvioija 1 oli samaa mieltä vain noin 30 prosentissa suorituksista. Edellä kuvattujen perinteisten arvioinnin luotettavuusanalyysien tulokset osoittavat selvästi, että yritys käyttää koostetta viiden arvioijan arvioinneista kol-

mantena arviointina ei sellaisenaan ole onnistunut. Koska on varsin mahdollista, että vain osa viidestä 'ylimääräisestä' arvioijasta on syynä ongelmallisiin tuloksiin, arvioinnit päätettiin analysoida myös Facets-ohjelmalla, joka sopii erityisen hyvin juuri arviointiaineistojen ja arvioijien laadun tarkasteluun (McNamara 1996; Linacre 2011).

Liitteissä 14 ja 15 raportoidaan Facets-analyysien tulokset, jotka suorimmin liittyvät arvioijien ominaisuuksiin. Tehtävän 1 ja tehtävän 2 arvioinnit analysoitiin erikseen, koska ne oli myös arvioitu erikseen. Taustalla on myös ajatus siitä, että tehtävät erosivat ainakin jonkin verran siinä, millaisia vaatimuksia ne asettivat oppijoiden kielitaidolle. Ensimmäisessä analyysissä kaikki arvioijat olivat mukana. Kummassakin tehtävässä kävi heti selväksi, että arvioijien välillä oli huomattavia linjaeroja, sillä jotkut heistä olivat selvästi toisia ankarampia (ks. erot Measure-sarakkeen arvoissa liitteiden 14 ja 15 taulukoissa ensimmäisessä analyysissä). Vaikka Facets-analyysi pystyykin jossain määrin ottamaan huomioon arvioijien linjaerot, ovat suuret, usean logit-arvon eroavuudet todennäköisesti liian suuria ja vääristävät näin oppijoiden saamia tuloksia.

Huolestuttavampaa on kuitenkin, jos arvioijat käyttäytyvät ailahtelevaisesti arvioinneissaan eli esimerkiksi peruslinjaltaan ankara arvioija antaakin välillä yllättävän helposti korkeita arvioita huonommille suorituksille tai päinvastoin. Tulokset osoittivat arvioinneissa vakavia ongelmia myös tässä suhteessa. Liitteiden 14 ja 15 taulukoiden Infit mean square -arvot osoittivat arvioijan johdonmukaisuutta tai sen puutetta. Täydellisen johdonmukainen arvioija (ainakin verrattuna muihin analyysissä mukana oleviin arvioijiin ja arvioitavien oppijoiden osaamistasoon) saisi arvoksi 1.0. Mitä enemmän arvo eroaa ykkösesestä, etenkin ylöspäin, sitä epäjohdonmukaisempi arvioija on. Yleisesti arvot, jotka ylittävät 1.3:n tai 1.5:n kertovat, että kyseisen arvioijan arviot alkavat merkittävästi vääristää tuloksia. Myös Infit Zstd -arvot voivat kertoa ongelmista arvioinnissa: mikäli kyseinen arvo on suurempi kuin +2 tai pienempi kuin -2, arvioijan työssä voi olla ongelmia.

Arvioinnit analysoitiin useassa eri vaiheessa. Kun ensimmäisen analyysin perusteella tietyt arvioijat osoittautuivat ongelmallisiksi, heidät poistettiin yksi kerrallaan analyysistä ja jäljelle jääneet arvioijat analysoitiin uudelleen. Analyysit 2 ja 3 liitteissä 14 ja 15 kuvaavat tällaisten välivaiheiden tuloksia tehtävien 1 ja 2 osalta. Kummassakin tapauksessa lopputulos oli se, että ainoastaan arvioija 2 ja arvioija 3 jäivät jäljelle; he kumpikin arvioivat tarpeeksi johdonmukaisesti ja lisäksi heidän yleinen arviointilinjansa oli varsin samanlainen ankaruudeltaan.

Tarkempi Facets-analyysi siis todisti sen, mitä jo perinteisemmätkin indikaattorit osoittivat, eli että yritys korvata kolmas arviointi/arvioija yhdistämällä viiden eri arvioijan työ ei osoittautunut onnistuneeksi ratkaisuksi. Lopulta yksikään viidestä arvioijasta ei osoittautunut niin johdonmukaiseksi, että hänen arviointinsa olisi kannattanut säilyttää lopullisessa aineistossa. Yksi syy tähän on se, että heistä jokainen arvioi varsin pienen määrän suorituksia, jolloin pie-nehkökin määrä ongelmia arvioinneissa voi näyttäytyä isona ongelmana, kun asiaa analysoitiin tilastollisesti. Mahdollisia syitä ovat myös näiden arvioijien erilaiset taustat ja odotukset oppijoiden kielitaidon suhteen: henkilö, joka on

tottunut työskentelemään pääasiassa hyvien kielenoppijoiden parissa, voi arvioida aloittelevia oppijoita aivan eri pohjalta kuin sellainen henkilö, joka on totunut aloittelijoihin, ja päinvastoin. Joissakin tapauksissa merkitystä voi olla niin ikään sillä, onko arvioija kyseisen kielen natiivi vai ei (arvioijista kaksi oli ranskan kielen natiiveja). Myös tottumattomuus arvioida puhumistaitoa ylipäänsä ja erityisesti tutkimuksessa käytetyn arviointiskaalan avulla voi johtaa arviointien poikkeavuuteen.

Vaikka yritys suoritusten kolmoisarviointiin epäonnistui, analyysien perusteella on kuitenkin selvää, että jäljelle jääneet kaksi arvioijaa tekivät varsin laadukasta ja johdonmukaista työtä. Tätä osoittavat sekä heidän arviointiensa samansuuntaisuus esimerkiksi korrelaatioiden ja ristiintaulukointien (ks.liitteet 14 ja 15) valossa että Facets-analyysit heidän arviointilinjojensa vakaudesta. Tutkimuksessa siis onnistuttiin luomaan laadukas ja uskottava ulkoinen arviointi oppijoiden suullisesta ranskan kielen taidosta. Sitä vasten on mahdollista peilata oppijoiden eri itsearviointivälineillä tekemää omaa arviointia.

### **5.3 Itsearviointien ja ulkoisten arviointien väliset korrelaatiot**

Tässä kappaleessa tarkastellaan oppijoiden itsearviointeihin liittyviä korrelaatioita vertaamalla keskenään eri itsearviointeja sekä itsearviointeja ja ulkoisia arviointeja. Ensimmäisessä taulukossa (taulukko 21) ulkoista arviointia edustaa arvioijien 2 ja 3 yhdistetyt arviot tehtävistä 1 ja 2, kun arvioija 1 poistettiin epäjohdonmukaisuuksien vuoksi. Kahdessa muussa taulukossa (taulukot 22 ja 23) ovat nähtävissä vastaavat vertailut sekä tehtävän 1 että tehtävän 2 osalta. Taulukkojen sisältöä kommentoidaan lyhyesti jokaisen taulukon jälkeen.

TAULUKKO 21 Yhdistettyjen ulkoisten arviointien (tehtävät 1 ja 2) ja itsearviointien väliset korrelaatiokertoimet.

		Ulkoinen arvio: oppijan puhumistaso tehtävissä 1 ja 2 yhteensä (Rasch-arvo)	Itsearviointi: oppijan puhumistaso, holistinen, 6-portainen EVK-asteikko	Itsearviointi: 42 väittämää (Rasch-arvo)	Itsearviointi: yhdistetyt tehtävän 1 ja 2 puhenäytteet (Rasch-arvo)
Ulkoinen arvio: oppijan puhumistaso tehtävissä 1 ja 2 yhteensä (Rasch - arvo)	Pearsonin korr.kerr. Merkitsevyystaso N	1  210	,748  209	,792  209	,745  209
Itsearviointi: oppijan puhumistaso, holistinen, 6-portainen EVK-asteikko	Pearsonin korr.kerr. Merkitsevyystaso N	,748  209	1  209	,902  209	,708  209
Itsearviointi: 42 väittämää (Rasch-arvo)	Pearsonin korr.kerr. Merkitsevyystaso N	,792  209	,902  209	1  209	,748  209
Itsearviointi: yhdistetyt tehtävän 1 ja 2 puhenäytteet (Rasch-arvo)	Pearsonin korr.kerr. Merkitsevyystaso N	,745  209	,708  209	,748  209	1  209

Yllä olevasta taulukosta näkyy, että verrattaessa oppijoiden itsearviointia ulkoiseen arviointiin, joka tässä tapauksessa tarkoittaa siis oppijan suorittamien tehtävien 1 ja 2 yhdistettyä puhumistasoa, korkein korrelaatiokerroin, .792, on ulkoisen arvioinnin ja paperilla tehdyn 42 väittämää sisältävän yksityiskohtaisen itsearvioinnin välillä. Toiseksi korkein korrelaatio, .748, vallitsee holistisen 6-portaisen asteikon avulla tehdyn itsearvioinnin sekä ulkoisen arvioinnin välillä. Tehtävien 1 ja 2 puhenäytteisiin perustuvan itsearvioinnin ja ulkoisen arvioinnin välinen korrelaatio jää hieman edellistä alemmaksi eli .745:een.

Eri itsearviointien välillä korrelaatiokerroin on korkeimmillaan (.902) kahden paperilla tehdyn itsearviointityypin eli 6-portaisen itsearvioinnin ja 42 kirjallista väittämää sisältäneen itsearvioinnin välillä. Toiseksi korkein korrelaatio, .748, on 42 väittämää sisältäneen itsearvioinnin ja tehtävän 1 ja 2 yhdistettyjen puhenäytteiden välillä. 6-portaisen yleisitsearvioinnin ja tehtävän 1 ja 2 yhdistettyjen puhenäytteiden välinen korrelaatiokerroin on .708.

TAULUKKO 22 Ulkoisten arviointien ja itsearviointien väliset korrelaatiokertoimet (tehtävä 1).

		Ulkoinen arvio: oppijan puhu- mistaso tehtä- vässä 1 (Rasch- arvo)	Itsearviointi: oppi- jan puhumistaso, holistinen, 6-portainen EVK- asteikko	Itsearviointi 42 väittä- mää (Rasch- arvo)	Itsearviointi: tehtävän 1 puhenäytteet (Rasch-arvo)
Ulkoinen arvio: oppijan puhu- mistaso tehtävä- ssä 1 (Rasch - arvo)	Pearsonin korr.kerr.	1	,737	,768	,753
	Merkitse- vyystaso	,000	,000	,000	,000
	N	210	209	209	209
Itsearviointi: oppi- ijan puhumista- so, holistinen, 6-portainen EVK-asteikko	Pearsonin korr.kerr.	,737	1	,902	,737
	Merkitse- vyystaso	,000	,000	,000	,000
	N	209	209	209	209
Itsearviointi: 42 väittämää (Rasch-arvo)	Pearsonin korr.kerr.	,768	,902	1	,763
	Merkitse- vyystaso	,000	,000	,000	,000
	N	209	209	209	209
Itsearviointi: tehtävän 1 puhenäytteet (Rasch-arvo)	Pearsonin korr.kerr.	,753	,737	,763	1
	Merkitse- vyystaso	,000	,000	,000	,000
	N	209	209	209	209

Tehtävän 1 osalta korkein korrelaatiokerroin, .768, vallitsee samoin kuin edellä ulkoisen arvioinnin ja 42 väittämän avulla tehdyn itsearvioinnin välillä. Edellä mainittu korrelaatiokerroin on tehtävän 1 osalta kuitenkin hieman alempi kuin edellä mainittu yhdistetty kahden tehtävän pohjalta laskettu kerroin (.792). Sen sijaan tehtävän 1 puhenäytteisiin perustuvan itsearvioinnin ja ulkoisen arvioinnin välinen korrelaatio .753 on korkeampi kuin edellä yhdistettyjen kahden tehtävän vastaava korrelaatio (.745). Kolmanneksi korrelaatiokertoimen suuruuden osalta jää holistisen 6-portaisen itsearvioinnin ja ulkoisen arvioinnin välinen korrelaatiokerroin, joka on tehtävän 1 osalta .737.

Eri itsearviointien välillä korrelaatiokerroin on tehtävän 1 osalta korkeampi eli .763 tehtävän 1 puhenäytteiden ja 42 käytetyn yksityiskohtaisen itsearviointiväittämän välillä kuin puhenäytteiden ja 6-portaisen itsearvioinnin välillä: siinä korrelaatio on .737.

TAULUKKO 23 Ulkoisten arviointien ja itsearviointien väliset korrelaatiokertoimet (tehtävä 2).

		Ulkoinen arvio: oppijan puhumistaso tehtävässä 2 (Rasch-arvo)	Itsearviointi: oppijan puhumistaso, holistinen, 6-portainen EVK-asteikko	Itsearviointi: 42 väittämää (Rasch-arvo)	Itsearviointi : tehtävän 2 puhenäytteet (Rasch-arvo)
Ulkoinen arvio: Oppijan puhumistaso tehtävässä 2 (Rasch-arvo)	Pearsonin korr.kerr. Merkitsevyystaso N	1  210	,700  209	,745  209	,648  209
Itsearviointi: oppijan puhumistaso, holistinen, 6-portainen EVK-asteikko	Pearsonin korr.kerr. Merkitsevyystaso N	,700  209	1  209	,902  209	,579  209
Itsearviointi: 42 väittämää (Rasch-arvo)	Pearsonin korr.kerr. Merkitsevyystaso N	,745  209	,902  209	1  209	,628  209
Itsearviointi: tehtävän 2 puhenäytteet (Rasch-arvo)	Pearsonin korr.kerr. Merkitsevyystaso N	,648  209	,579  209	,628  209	1  209

Tehtävässä 2 korkein korrelaatiokerroin on samoin kuin kahdessa edellisessä tarkastelussa ulkoisen arvioinnin ja 42 yksityiskohtaisen itsearviointiväittämän avulla tehdyn itsearvioinnin välillä (.745). Seuraavaksi korkein korrelaatiokerroin itsearvioinnin ja ulkoisen arvioinnin välillä tehtävässä 2 on 6-portaisen yleisitsearvioinnin yhteydessä (.700). Matalin korrelaatio itsearvioinnin ja ulkoisen

arviointien välillä on oppijan puhumistason ja tehtävää 2 koskevien puhenäytteiden välillä, missä korrelaatio laskee alle .70:n (.648).

Yhteenvedona korrelaatiokertoimien tarkastelusta kolmen edellä esitellyn taulukon pohjalta voidaan todeta seuraavaa:

Itsearviointien ja ulkoisen arvioinnin välinen korrelaatio on suurin (.792) silloin, kun ulkoinen arviointi muodostuu oppijan yhdistetystä puhumistasosta tehtävissä 1 ja 2 ja kun itsearviointi suoritetaan paperilla käyttäen 42 yksityiskohtaista itsearviointiväittämää. Myös ulkoisen arvioinnin, joka muodostuu oppijan puhumistasosta tehtävissä 1 ja saman edellä mainitun 42-kohtaisen itsearviointilomakkeen yhdistelmästä, korrelaatiokerroin on korkeahko (.768). Kolmanneksi korkein korrelaatiokerroin (.753) ulkoisen arvioinnin ja itsearviointien välillä toteutuu tehtävän 1 osalta silloin, kun oppijan tehtävää 1 koskevaa puhumistaitoa verrataan tehtävää 1 koskevaan 26 puhenäytteeseen.

Tehtävää 2 koskevat korrelaatiokertoimet ovat yleisesti matalampia kuin tehtävää 1 tai molempia tehtäviä yhdistettynä koskevat kertoimet. Syynä tähän saattaa olla se, että suoriutuminen tehtävissä 2 perustui tehtävää 1 selvemmin tilanteeseen sopivan kielen käyttöön. Oppijoiden ja osittain myös arvioijien on saattanut olla vaikeampi arvioida tämänkaltaista tilannekohtaista kielenkäyttöä kuin kielitaitoa yleensä.

Tutkimuksen korkein korrelaatio eri itsearviointien välillä toteutuu silloin, kun itsearviointia tehdään käyttäen paperilla tehtyyn itsearviointiin tarkoitettuja lomakkeita. Käytössä ovat tällöin paperilla tehty 6-portainen yleistaso määrittävä itsearviointilomake ja 42 yksityiskohtaista itsearviointiväittämää sisältävä itsearviointilomake (.902).

Seuraavaksi korkeimmat korrelaatiokertoimet ovat tehtävää 1 koskevia: korkein kahdesta kertoimesta on .763, joka ilmaisee 42 itsearviointiväittämän ja tehtävä 1:tä koskevan 26 puhenäytteen välistä suhdetta. Myös 6-portaisen yleistaso määrittävän lomakkeen avulla tehdyn itsearviointien suhde tehtävän 1 puhenäytteisiin on melko hyvä eli .737.

Tehtävää 2 koskevien itsearviointien väliset kertoimet (.579 ja .628) ovat jälleen alempia kuin tehtävää 1 ja molempia tehtäviä yhdistettynä koskevat kertoimet, jotka vaihtelevat .708 ja .763 välillä (jos jätetään huomiotta kahden paperilla tehdyn itsearviointien välinen .902-suuruinen kerroin).

## 5.4 Tutkimukseen osallistuneiden laadullisia kommentteja

Luvussa 5.4 käsitellään tutkimukseen osallistuneiden oppijoiden itsearviointiin liittyviä seikkoja laadullisesta näkökulmasta. Luvussa ei siten enää palata työssä aiemmin käsiteltyihin yksittäisiin väittämiin, joiden toimivuutta on selvitetty tilastotieteellisten analyysien avulla luvussa 5.

Aluksi luvussa 5.4.1 tarkastellaan itsearviointi paperilla -mittaria koskevia oppijoiden kommentteja, jotka jakautuvat kolmeen pääkategoriaan. 1-



kategorian kommentit ovat kielitaidon yleistason valintaa koskevia kommentteja. 2-kategorian kommentit edustavat oppijoiden kommentteja, joissa he ovat ilmaisseet yleisellä tasolla mielipiteensä 42 yksityiskohtaisesta lomakkeen itsearviointiväittämästä kohdistamatta kuitenkaan kommenttiaan mihinkään tiettyyn lomakkeen väittämään. 3-kategoriaan ryhmitellyt kommentit eivät koske suoraan itsearviointi paperilla -lomaketta. Viimeisen kategorian kommentit sisältävät pikemminkin oppijoiden pohdintaa siitä, millaisiksi he kokevat itsensä oman kielitaitonsa arvioijina. Tähän kategoriaan on koottu myös ne kommentit, joissa oppijat pohtivat itsearviointiinsa vaikuttavia tekijöitä.

Luvussa 5.4.2 esitellään tehtäviä 1 ja 2 koskevat kommenttikategoriat, joihin on luokiteltu osallistujilta saadut puhenäytteitä koskevat kommentit sekä puhenäytteiden litteroinnissa käytetyt merkit. Luvun 5.4.2 alaluvuissa 5.4.2.1 (tehtävä 1) ja 5.4.2.2 (tehtävä 2) ovat vuorossa molempia tehtäviä koskevat kommenttianalyytit. Tarkastelussa keskitytään aluksi esittelemään taulukkomuodossa eniten tai vähiten "vaikea päättää" -rasteja saaneita näytteitä, minkä jälkeen lähempään tarkasteluun otetaan molempien tehtävien kolme eniten ja vähiten rasteja kerännyttä näytettä. Luku 5.4.2.3 sisältää yhteenvedon "vaikea päättää" -rasteja keränneiden puhenäytteiden ominaisuuksista.

#### **5.4.1 Itsearviointi paperilla -mittarin kommenttikategoriat ja osallistujien kommentit**

Tässä osassa käsitellään paperilla tehdyn itsearvioinnin tuloksia. Käsittelyssä hyödynnetään tilastollisten analyysien antaman tiedon lisäksi tutkimukseen osallistujilta kerättyä palautetta, joka koski käytössä olleen itsearviointilomakkeen sisältöä ja toimivuutta.

Paperilla tehdyssä itsearvioinnissa oli siis käytössä 2-osainen lomake, jonka ensimmäinen osa sisälsi nk. yleisitsearviointia kuudella Viitekehyksen perustaitotasolla (A1-C2). Lomakkeen toinen osa koostui 42 tilannekohtaisesta itsearviointiväittämästä, jotka kaikki liittyivät kielenoppijan taitoon puhua ranskan kieltä. Lomakkeessa käytetyt yksityiskohtaiset itsearviointiväittämät on laadittu DIALANG-hankkeessa, jossa ne on myös empiirisesti validoitu.

Nk. yleisitsearvioinnissa tutkimukseen oppijan tuli valita yksi kuudesta perustaitotasosta, joka parhaiten kuvasi hänen ranskan kielen suullista taitoaan tutkimukseen osallistumishetkellä. Koska tämä valinta tehtiin ensimmäisenä tehtävänä kielistudiossa, kyseessä oli osallistujan yleiskuva omasta suullisesta kielitaidostaan ranskan kielessä. Taulukossa 24 esitellään prosenttilukuina, mille tasolle tutkimukseen osallistuneet oppijat arvioivat oman suullisen ranskan kielen taitonsa. Yleisitsearvioitun kielitaidon keskiarvo 6-portaisella EVK-asteikolla arvioituna oli 2,86, mediaani 3 ja keskihajonta 1.18.

TAULUKKO 24 Oppijoiden ranskan kielen puhumistaito itsearvioituna holistisesti 6-portaisella EVK-asteikolla.

	Frekvenssi	%
A1	29	13,9
A2	52	24,9
B1	68	32,5
B2	42	20,1
C1	16	7,7
C2	2	1,0
Yhteensä	209	100,00

Suurin osa (32,5 %) tutkimukseen osallistuneista oppijoista arvioi oman suullisen ranskan kielen taitonsa yleistaitotasolle B1. Toiseksi eniten osallistujat valitsivat tasoa A2 (24,9 %) ja kolmanneksi tasoa B2 (20,1 %). Alimmalle taitotasolle A1 oman kielitaitonsa arvioineita oli tutkimuksessa 13,9 %. Kahden ylimmän tason valinnoita oli hyvin vähän: 7,7 % tasolla C1 ja 1,0 % tasolla C2.

Vieraiden kielten<sup>21</sup> tavoitetaitotasot puhumisen osalta sijoittuvat lukion opetussuunnitelmien perusteiden mukaan taitotasolle A2.1 (B3-oppimäärä), A2.1–A2.2 (B2-oppimäärä) ja B1.1 (A-oppimäärä; LOPS 2003, 100). Ranskan kielen suullisen taidon harjoittelumäärät ja -mahdollisuudet opintojen yhteydessä Suomessa ovat verraten vähäiset. Suomalaisia, joiden ranskan kielen suullinen taito sijoittuu ylimmille taitotasolle, on melko vähän. Täten voidaan todeta, että on hyvin odotuksenmukaista, että oppijat valitsivat tutkimuksen itsearviointiosassa eniten taitotasoa B1 ja A2.

#### *Osallistujien kommentteja*

Edempänä on analysoitu itsearviointi paperilla -lomaketta tilastollisten tunnuslukujen avulla ja saatu siten monenlaista tietoa väittämien toimimisesta. Tässä osassa tarkastellaan samaa lomaketta laadullisesti oppijoiden antamien kommenttien valossa. Tarkoitus on luoda katsaus siihen, miten tutkimukseen osallistuneet oppijat kokivat kirjallisten itsearviointiväittämien käytön ja millaisia kommentteja heillä mahdollisesti oli lomakkeesta yleensä sekä sen sisältämistä väittämistä.

30 % tutkimukseen osallistuneista oppijoista kommentoi kirjallista puhumisen itsearviointi -lomaketta väittämien sisällön, ymmärrettävyyden, selkeyden ja yleisen toimivuuden näkökulmasta. Tässä vaiheessa en enää käsittele niitä kommentteja, jotka liittyvät lomakkeen yksityiskohtaisiin väittämiin, sillä niitä on käsitelty jo edempänä tilastollisten analyysien yhteydessä, kun on yritetty löytää syitä tai selityksiä esimerkiksi tiettyjen väittämien toimimattomuudelle itsearviointimittarin osana.

<sup>21</sup> Ei koske englannin kieltä.

Oppijoiden kommentit sisältävät pääkategoriat ovat seuraavat:

- yleistason valintaa koskevat kommentit (tasot A1–C2)
- yleiskommentit, jotka liittyvät itsearviointiväittämiin ilman tarkempaa viittausta tiettyyn väittämään (42 itsearviointiväittämää)
- oppijaa itseään ja/tai hänen itsearviointiaan koskevat kommentit.

*Yleistason valintaa koskevat kommentit (tasot A1– C2)*

10 osallistujaa kommentoi yleistason valintaa. Kuusi näistä henkilöistä kertoi epäröineensä kahden vierekkäisen tason valinnan välillä, minkä jälkeen neljä heistä totesi harkinnan jälkeen päätyneensä alempaan kahdesta harkitusta taitotasosta. Alla on malliksi kaksi tähän asiaan liittyvää kommenttia, jotka tuntuvat pyörivän varmuus-käsitteen ympärillä<sup>22</sup>:

*Taitotasokuvauksissa valitsin 3, koska en tänä vuonna ole käyttänyt ranskaa paljon, mutta jos olisin opiskellut sitä tänä vuonnakin, olisin laittanut 4, kun varmuus oli parempi. (Tällä hetkellä opiskelen espanjaa ja se vähän sekoittaa) (3).*

*Olin 3 ja 4 arvioiden välillä, ja alempi taso tuli valittua kuin varmuuden vuoksi. 4-taso olisi ollut mahdollinen, koska kyseessä on yhä "tutut tilanteet" – en kuitenkaan ole joutunut kovin usein vieraisiin tilanteisiin nykyisellä kielitaidollani, ja tottumattomuus varmasti hankaloittikin alussa ennen todell. 4 -tasolle pääsemistä (3).*

Monet tämän kohdan kommentoijista totesivat kokeneensa yleistaitotason valinnan hankalaksi ja/tai jopa sattumanvaraiseksi. Kommentoijien mukaan sama väittämä saattoi sisältää eri kielitaidon tasoille kuuluvia elementtejä, mikä heidän mukaansa vaikeutti oikean tason valintaa. Väittämien kritisoitiin myöskin olevan liikaa toisiaan muistuttavia.

*Yleiskommentit, jotka liittyvät itsearviointiväittämiin ilman tarkempaa viittausta tiettyyn väittämään (42 itsearviointiväittämää)*

Yksityiskohtaisia itsearviointiväittämiä kommentoitiin huomattavasti enemmän kuin yleistason valintaa. Tutkimukseen osallistuneet oppijat antoivat 184 tähän kategoriaan luokiteltavaa kommenttia. Suurin osa kommenteista käsittelee seuraavia asioita (luku sulkeissa tarkoittaa kommenttien määrää):

- "kyllä" tai "ei" -vaihtoehdon valinta ei tuottanut vaikeuksia (15)
- väittämät olivat liian samanlaisia (14)

---

<sup>22</sup> Numero sulkeissa osoittaa, mille taitotasolle osallistuja arvioi oman kielitaitonsa.

- valinta "kyllä" tai "ei" -vaihtoehdon välillä oli liian rajoitettu (8)
- väittämässä käytetyt käsitteet eivät olleet riittävän tarkasti määriteltyjä (5).

Oppijan näkökulmasta onkin varsin ymmärrettävää, että valinta dikotomisten "kyllä" ja "ei" -vaihtoehtojen väliltä saattaa tuntua hankalalta. Viitekehyksessäkin korostetaan paljon osittaisten kompetenssien hallinnan tärkeyttä (EVK 2004, 187), ja tältäkin kannalta asiaa tarkastellen "kyllä" ja "ei"-vaihtoehdot edustanevat ideaalitulanteen kannalta liian jyrkkiä ääripäitä. Lisäksi joidenkin oppijoiden saattaa olla vaikea asettua konkreettisesti väittämän esittämään tilanteeseen ja arvioida sen pohjalta omaa suullista kielitaitoaan.

*Oppijaa itseään ja/tai hänen itsearviointiaan koskevat kommentit*

3-ryhmään sijoittuu yhteensä 47 kommenttia, jotka liittyvät jollakin tavalla joko oppijaan itseensä, hänen itsearviointiinsa ja/tai hänen kokemukseensa itsearvioinnista. Kommentit on jaoteltu seuraavaan viiteen alaryhmään, ja niitä käsitellään lyhyesti esimerkkien valossa alla:

- puuttuva kokemus väittämässä esitellystä tilanteesta
- oman itsearviointitaidon pohdinta
- itsearvioinnin yhdistäminen johonkin tiettyyn hetkeen tai aikaan
- oppijan luonteen ja/tai toimintatavan vaikutus itsearviointiin
- oppijan asenne omaa tasoa alempiin itsearviointiväittämiin.

*Puuttuva kokemus väittämässä esitellystä tilanteesta*

16 tutkimukseen osallistujaa kertoi vastauksissaan, että oli äärimmäisen vaikea kuvitella itseään sellaiseen kielelliseen tilanteeseen, josta itsellä ei ollut henkilökohtaista kokemusta. Kokemuksen puute sellaisesta todellisesta kielenkäyttötilanteesta, joka vastaisi väittämässä kerrottua tilannetta, teki näiden vastaajien kommenttien mukaan itsearvioinnista vaikeaa. Koska kokemus erilaisista kielenkäyttötilanteista karttuu kielitaidon kasvaessa, onkin ymmärrettävää, että suurin osa tähän ryhmään kuuluvista kommentteista saatiin oppijoilta, jotka sijoittivat kielitaitonsa pikemminkin taitotasoasteikon alatasoille eli tasoille A1 (4 kpl), A2 (7 kpl) tai B1 (4 kpl). Tässä muutama esimerkki tyyppillisistä kommentteista:

*"Oli hankala arvioida, koska en ole joutunut käyttämään koskaan ranskan kieltä käytännön tilanteessa. Lopuksi rupesin epäilemään, olenko antanut keskenään ristiriitaisia vastauksia" (2).*

*"Itsellänikään ei ole kunnollista kuvaa oman suullisen kielitaitoni tasosta, koska en ole oikeastaan koskaan käyttänyt suullista kielitaitoani ranskan tuntien ulkopuolella! Joten kielitaidon arviointi on aika hankalaa..." (3).*

Edellä esiteltyjen kaltaisten kommenttien pohjalta voidaankin pohtia, kuinka realistista on odottaa, että kieliopintojensa alkuvaiheessa olevalla oppijalla olisi todellinen kuva sellaisesta kuvitellusta kielenkäyttötilanteesta, josta hänellä ei ehkä ole aiempaa kokemusta edes omalla äidinkielellään. Pitäisikö kenties hyväksyä se, että tietynlaisten oppijoiden kanssa käytettäisiin tietynlaisia itsearviointiväittämiä? Vai voidaanko tilannetta lähestyä niin, että itsearvioinnin lähtöoletuksena on se, että oppija on kykenevä kuvittelemaan itsensä erilaisiin kielellisiin tilanteisiin siitäkkin huolimatta, ettei hänellä ole niistä todellista kokemusta? Varmaa ainakin on, että mikään eritasoisille oppijoille tarkoitettu itsearviointilomake ei voi kaikilta osin täyttää kaikkien erilaisten kielenkäyttäjien tarpeita.

#### *Oman itsearviointitaidon pohdinta*

10 tutkimukseen osallistunutta oppijaa kertoi itsearviointiprosessin käynnistyneen heissä reflektiivisen pohdinnan siitä, miten he kokevat kykynsä arvioida omaa osaamistaan. Monet kommentteista tuntuivatkin kyseenalaistavan sen, kuinka oikeaan osuvaa oma itsearviointi oli. Useat vastaajat tuntuivat epäilevän, että saattoivat tutkimuksen aikana joko yliarvioida tai aliarvioida omaa suullista ranskan kielen taitoaan. Joissakin tapauksissa tutkimukseen kuuluvien lomakkeiden täyttäminen tuntui saavan oppijan epäilemään entistä enemmän omaa kielitaitoaan. Tässä muutamia esimerkkejä tähän ryhmään kuuluvista kommentteista:

*"Olen aina ajatellut puhuvani sujuvaa ranskaa. Nyt kun kävin läpi kaikki nämä kysymykset yksityiskohtaisesti, en olekaan taidoistani täysin vakuuttunut" (3).*

*"Aina en ole varma, mitä on sujuva ja looginen puhe esimerkiksi olen Ranskassa ollessani tehnyt ja puhunut kaikkea kysyttyä mutta lukiessani määritelmiä paperilla en ole niinkään varma kyvyistäni" (3).*

*"Tuntui, että koko ajan joko yli- tai aliarvioi itseään" (2).*

Osa alimmille tasoille kielitaitonsa arvioineista oppijoista oli sitä mieltä, että oikeaan osuva itsearviointi edellyttää tiettyä kielitaidon tasoa:

*"Ei erityisiä vaikeuksia. Olen opiskellut ranskaa puoli vuotta joten en osaa välttämättä arvioida puhutun kielen osaamistani oikein" (1).*

Edellisten pohdintojen perusteella voidaan todeta olevan täysin luonnollista, että ihmiset oppivat kiinnittämään huomiota tiettyihin asioihin vasta sitten, kun joutuvat tekemisiin niiden kanssa joko todellisessa tai kuvitellussa kielenkäyttötilanteessa. Se, että joutuu miettimään ja punnitsemaan omaa osaamistaan ja mahdollista selviytymistään tietynlaisessa kielenkäyttötilanteessa, saattaa aiheuttaa osalle oppijoista epävarmuutta omista taidoistaan (ks. 3.4.1 oppijan persoonallisuus ja minäkuva).

*Itsearviointin yhdistäminen johonkin tiettyyn hetkeen tai aikaan*

Kuten aikaisemmin itsearviointilomakkeen väittämiä analysoitaessa todettiin, osallistujien kommentteista kävi selväksi, että itsearviointin ajallinen sijoittaminen tuntui olevan oppijoille joissakin tapauksissa haasteellista. Osa oppijoista vastasi väittämiin siten kuin kaikkia osallistujia oli pyydetty tekemään eli sen käsityksen mukaan, mikä heillä oli itsellään ranskan kielen suullisesta taidostaan tutkimukseen osallistumishetkellä (= millainen suullinen ranskan kielen taitoni on nyt). Sen sijaan vastauksista ilmeni selvästi myös se, että osa oppijoista vastasi väittämiin jonkun toisen ajanjakson tai hetken perusteella kuin tutkimukseen osallistumishetken. Näissä tapauksissa henkilöt viittasivat yleensä aikaan, jonka he olivat viettäneet ranskankielisessä ympäristössä (= millainen suullinen ranskan kielen taitoni oli silloin aiemmin). Tämänkaltaisiin kommentteihin näyttää usein liittyvän tietynlainen ehdollisuus, mikä ilmenee jos-sanan sisältävän ehtoja asettavan konditionaalilauseen esittämisenä (= jos tilanne olisi x, kielitaitoni olisi y). Näiden kommenttien esittäjät eivät siis välttämättä ole tehneet itsearviointiaan tutkimushetken kielitaitonsa pohjalta. Toisaalta tämänkaltaisia kommentteja esittäneiltä oppijoilta ei heidän itsearviointinsa lähtökohdaksi ole voitu varmistaa. Siten voi olla mahdollista myös, että kommentoijat ovat tehneet itsearviointin tutkimushetken kielitaitonsa pohjalta ja kommentoineet tämän lisäksi kielitaitonsa eri seikoista riippuvaa vaihtelua ikään kuin asian lisävalaistukseksi.

Tässä muutamia esimerkkejä tämäntyyppisistä kommentteista:

*”Kaikkihan tietysti riippuu tilanteesta ja esim. siitä, kuinka paljon kieltä olen käyttänyt lähiaikoina” (3).*

*”Tällä hetkellä vastaus on eri kuin mitä se olisi jos olisin esim. juuri tullut parin viikon matkalta Ranskasta tai jos testi tehtäisiin siellä. Aksentti ja kieli yleensä häviävät kun olen kauan Suomessa” (5).*

*”Minusta tuntuu, että juuri tällä hetkellä en ehkä pysty kaikkeen, mihin vastasin ’kyllä’, mutta jos hieman kertailisin aiemmin oppimaani, niin pystyisin” (2).*

Todettakoon jälleen, ettei ole toivottavaa, että tutkimukseen osallistajat ovat suullista kielitaitoaan arvioidessaan vastanneet väittämiin viitaten johonkin muuhun kuin tutkimuksen osallistumishetkellä olevaan kielitaitoonsa. Vastauksen yhdenmukaisuuden kannalta olisi tietenkin parempi, että kaikkien vastaukset perustuisivat siihen kielitaitoon, joka oppijoilla oli tutkimukseen osallistumisaikansa eikä jonakin muuna hetkenä tai muissa olosuhteissa mahdollisesti hallittuun kielitaitoon.

*Oppijan luonteen ja/tai toimintatavan vaikutus itsearviointiin*

Tiettyä kulttuuria - tämän tutkimuksen tapauksessa suomalaista kulttuuria - edustavat oppijat heijastavat kommentteissaan omaa kulttuuriaan, suomalaista luonteenlaatua ja toimintatapaa. Oppijan luonteella ja sillä kulttuurilla, josta hän on lähtöisin, on vaikutusta oppijan kuvaan itsestään ja hänen itsearviointiinsa. Suomalainen suhteellisen ilmeettömän ja tietyllä tavalla maltillinen ilmaisu-tapa eroaakin suuresti ranskalaisesta yleisesti ottaen suurieleisemmästä ja ekspressiivisemmästä tavasta ilmaista itseään. Omaan kulttuuriin liittyvät piirteet siirtyvät monien kielenkäyttäjien toimintaan ja käyttäytymiseen myös silloin, kun he ilmaisevat itseään vieraalla kielellä. Myös kielenkäyttäjän tai kielenoppijan sisäinen konteksti – kuten esimerkiksi muisti, mielikuviutus ja erilaiset sisäiset ja emotionaaliset prosessit – vaikuttaa siihen, miten henkilö hahmottaa kielenkäyttötilannetta. Sen vuoksi joissakin tapauksissa tuotettu kieli voi liittyä vain marginaalisesti siihen ulkoiseen kontekstiin, joka on havaittavissa (EVK 2004, 82).

Kuva omasta itsestä kielenkäyttäjänä vaikuttaa ilman muuta henkilön itsearviointin tuloksiin. Tähän tutkimukseen osallistuneiden oppijoiden kommentit eivät tue näkemystä, jonka mukaan kielenkäyttäjä voisi ikäänkuin ottaa eri persoonallisuuden käyttäessään vierasta kieltä suullisesti. Tässä joitakin kommentteja aiheeseen liittyen:

*”Osa itsearviointinista ei koske pelkkää ranskan kieltä vaan koko toimintatapaa. Ellei osallistu aktiivisesti keskusteluun suomeksi, ei takuulla muuta olemustaan kielen myötä niin, että ranskaksi olisikin vuolassanainen” (2).*

*”Itseni tuntien en pysty pitämään esim. virallista lyhyttäkään puhetta tai vakuuttelua ranskaksi, koska se on minulle suomeksikin ikävää tilanne (en viihdy sellaisessa!)” (3).*

*”En ole luonteeltani puhelias, sen vuoksi löytyy esimerkeistä paljon puhetilanteita, jollaisissa en ole ollut” (5).*

*Oppijan asenne omaa tasoa alempiin itsearviointiväittämiin*

Jotkut ylimmille taitotasoasteikon tasoille oman kielitaitonsa arvioineista oppijoista tuntuivat olevan epävarmoja siitä, kuinka vastata niihin itsearviointilomakkeen väittämiin, joissa ilmaistu puhetilanne edellytti alemmantasoista kielitaitoa kuin heidän oma kielitaitonsa oli. Tässä joitain esimerkkejä kommentteista:

*”Jotkut kysymykset olivat tyhmiä (?), koska niissä kysyttiin osaanko selviytyä jostain helposta tilanteesta. Pystyn selviytymään helposta, mutta myös vaikeasta tilanteesta (mitä vastata)?” (6)*

*"Tuntui välillä hassulta vastata 'kyllä' sellaisiin väittämiin, jotka olivat reilusti alle oman tason" (4).*

*"Joihinkin kysymyksiin en tiennyt miten vastata, esim. kysyessä peruskielitaitoon liittyviä tietoja (jotka jo on): vastaanko 'kyllä' = osaan vai 'ei' = osaan enemmän" (4).*

Tämän kohdan kommentit näyttäisivät olevan käänteisesti samantyyllisiä kuin kommentit, jotka koskivat puuttuvaa kokemusta väittämässä esitellystä tilanteesta. Tähän kohtaan kommenttejaan antaneet ylimpien tasojen oppijat kokivat siten itsellään olleen sekä kokemusta että kielitaitoa niin paljon, että he selviävät vaativista ylimmille tasoille sijoittuvista kielenkäyttötilanteista. Silti edellä olevista kommenteista käy ilmi se, että kommentoijat eivät ohjeistuksesta huolimatta ymmärtäneet vastata omaa kielitaitoaan alempitasoista osaamista kuvaaviin väittämiin "kyllä" (= selviydyn tilanteesta).

Yllä oleviin kommentteihin viitaten voidaan todeta, että käytettäessä itsearviointilomakkeita, joissa on kirjallisia itsearviointiväittämiä, oppijoille on ilmeisesti syytä selostaa lyhyesti taitotasoarvioinnin perusideologiaa. Nähtävästi kaikki oppilaat eivät ymmärrä, että jos oletetaan jonkun henkilön selviytyvän kielellisesti vaativasta ylimmille taitotasoille sijoittuvasta kielenkäyttötilanteesta, voidaan olettaa myös, että sama henkilö selviytyisi ainakin pääosin myös sitä alemmille tasoille sijoittuvista kielenkäyttötilanteista. Tähän tutkimukseen osallistujille annettiin ohjeeksi kielistudio-osuuden alussa, että kaikkiin niihin väittämiin, joiden ilmaisemasta tilanteesta selviytyminen tuntui mahdolliselta, vastataan "kyllä". Vastaavasti ohjeeksi annettiin, että niiden väittämien ilmaisemiin tilanteisiin, joista oppija ei kuvittelisi selviytyvänsä, vastataan "ei". Näyttäisi siten siltä, että oppijoiden ohjeistus voisi olla vieläkin yksityiskohtaisempaa, jotta oppijoille jäisi mahdollisimman vähän tulkinnanvaraa sen suhteen, miten konkreettisesti toteuttaa itsearviointia paperilla olevan mittavälilinen avulla.

#### *Yhteenvetoa itsearviointi paperilla -mittarista*

Tutkimukseen osallistuneiden oppijoiden vapaita kommentteja läpikäytessä voidaan havaita niiden antaneen tietynlaista hyödyllistä tietoa mm. väittämien varsinaisesta sisällöstä, puhumisen itsearviointilomakkeen toimimisesta sekä oppijoiden mietteistä ja asenteista itsearviointia kohtaan. Oppijoiden vapaiden kommenttien ei siis voida katsoa antavan vain tilastollisten analyysien tuloksia täydentävää lisätietoa, vaan ne nostavat esille esimerkiksi sen, että jotkut väittämät ovat saattaneet tuntua oppijoista hankalilta, vaikka nämä samaiset väittämät eivät erottuisi muiden lomakkeen väittämien joukosta missään tilastollisen tarkastelun vaiheessa. Oppijoiden antamia vapaita kommentteja voidaan siten pitää tilastollisten analyysien rinnalla tärkeinä indikaattoreina siitä, mikä oppijoiden mielestä toimii ja mikä ei.



#### 5.4.2 Kuunneltavat puhenäytteet -mittareiden kommenttikategoriat ja osallistujien kommentit

Tutkimuksen aineistoa kerätessä harjoitettiin siis kolmenlaista itsearviointia. Niistä kaksi tehtiin perinteisesti kirjallisia itsearviointiväittämiä apuna käyttäen (yleisitsearviointi ja yksityiskohtaisiin kielenkäyttötilanteisiin liittyvä itsearviointi). Tässä osassa tarkastellaan lyhyesti niitä osallistujien kommentteja, jotka liittyvät kielistudiossa toteutettuun itsearviointiin, jossa osallistujat vertasivat omaa suoritustaan nauhalta kuunneltaviin puhenäytteisiin.

Koska tutkimusaineisto on laaja ja sitä on pääosin analysoitu kvantitatiivisia tilastomenetelmiä käyttäen, päätettiin kuunneltaviin näytteisiin liittyvien kvalitatiivisesti tarkasteltavien kommenttien määrää rajoittaa. Yksityiskohtaiseen tarkasteluun otettavat näytteet valittiin molempien tehtävien osalta siten, että vastauslomakkeissa ollut "vaikea päättää" (vp) -sarakkeen rastien määrä näytettä kohden toimi näytteiden valintakriteerinä. Valintavaiheessa päätettiin, että tarkempaan tarkasteluun otetaan kolme prosentuaalisesti eniten ja kolme vähiten vp-rasteja kerännyttä näytettä kummastakin tehtävästä. Täten molemmista tehtävistä otettiin tarkasteluun yhteensä 6 näytettä. Jotta saatiin tietää, mitkä näytteet kummastakin tehtävästä edustaisivat eniten tai vähiten vp-rasteja keränneitä näytteitä, käytiin läpi osallistujien vastauslomakkeet 40 %:n osallistujamäärällä. Tämän toimenpiteen jälkeen saatiin selville näytteet, joista osallistujat olivat sitä mieltä, että oman suorituksen vertailu näytteeseen oli jostain syystä vaikeaa (= paljon vp-rasteja) tai että päinvastoin oman suorituksen vertailu kuunneltavaan näytteeseen ei aiheuttanut valintatilanteessa vaikeuksia (= vähän vp-rasteja).

Oppijoiden kommentteja varten laadittiin kommenttikategorioita, joiden tarkoitus oli luokitella kommentit sisältönsä perusteella riittävän yksityiskohtaisesti pääosin kielellisiin kategorioihin mutta myös muutamaaan muuhun kategoriaan. Kategorioiden laadinnassa otettiin huomioon yleisesti puhumisen testeissä käytetyt yleisimmät kriteerit kuten kieliopillinen tarkkuus, sujuvuus, sanasto(n määrä) ja ääntäminen (Underhill, 1987). On huomattava, että käytetty luokitus on hyvin karkea eikä sen tarkoitus ole olla eksplisiittinen. Luokittelussa käytetyt kahdeksan kommenttikategoriaa ovat seuraavat:

1. ääntäminen
2. sujuvuus
3. kielioppi
4. laatu
5. määrä
6. sisältö
7. puhuja
8. muut.

Kommenttikategorioiden määrä haluttiin pitää riittävän suppeana. Siitä johtuen jokaisen kategorian sisällä on hyvin monenlaista tiettyyn puhutun kielen ominaisuuteen liittyvää ainesta. Taulukossa 25 annetaan esimerkkejä sekä eri kategorioihin sijoitetuista kielen piirteistä että oppijoiden kommentteista.

TAULUKKO 25

Tehtävien 1 ja 2 kuunneltuja puhenäytteitä koskevat kommenttikategoriat.

Kategoriat	Esimerkkejä kategoriaan sijoitetuista kielen piirteistä	Esimerkkejä oppijoiden kommentteista
1. Ääntäminen	aksentit, nasaalit, puherytmi	<i>hirveä ääntäminen (mutta sujuva), aksentti suomalainen</i>
2. Sujuvuus	puhujan epäröinti, takeltelu, töksähtely, sanojen haku, hitaus, luontevuus, tauot	<i>sujuvampi kuin meikälainen, taukoja puheessa, ei-sujuva, puhe takertelevaa, hän puhui tönkösti</i>
3. Kielioppi	morfosyntaktiset, syntaktiset ja leksikaaliset kommentit (esim. pronominit, verbit, artikkelit, aikamuodot ja sanat)	<i>joitakin alkeellisia virheitä, muuten ok, vähemmän sanastoa kuin minulla, kielioppivirheitä</i>
4. Laatu	puhetuotoksen yleistä laatua kuvaavat kommentit, jotka eivät kuulu johonkin jom. kategoriaista	<i>hyvä, yksinkertainen, lyhyt, puhekielinen, selkeä, epälooginen</i>
5. Määrä	puhetuotoksen määrää kuvaavat kommentit (sis. yleensä jonkin määrää ilmaisevan sanan). Myös merkitykseltään puhetuotoksen määrää kommentoivat ilmaukset, joissa ei määrää ilmaisevia sanoja.	<i>hyvää tekstiä, mutta vähän; paljon asiaa</i>  <i>juttua riittää</i>
6. Sisältö	puhetuotoksen asiasisältö, tehtävänannon täyttyminen	<i>Outoja, sekavia juttuja; jäi pelkkään kuvailuun vaatteista; kertoo tarinan, hyvä lähestymistapa</i>
7. Puhuja	puhujan olemukseen tai esiintymiseen liittyvät kommentit	<i>hänellä on taito puhua; tuntuu, että puhuja jännittäisi kauheasti; kuin rautakangen niellyt</i>
8. Muut	sekalaisia kommentteja	<i>suuria vaikeuksia; kielitaito puutteellinen; ihan urpo, ei kovin kummoinen; kuin yö ja päivä ;) ; tavoitseeko muuta sanoa kuin "ranskalainen"?</i>

Taulukossa 25 esiteltyjen kategorioiden lisäksi tehtävässä 2 on tehtävän luonteen vuoksi käytössä sosiolingvistiseen kielenhallintaan liittyvä kategoria.

Kommenttien luokittelu eri kategorioihin oli osittain haasteellista ja joissakin tapauksissa myös tulkinnanvaraista. Erityisesti puhetuotoksen laatuun, määrään ja sisältöön liittyvien kommenttien luokittelussa saattaa esiintyä jonkin verran päällekkäisyyttä. Esimerkiksi kommentti *"aika lyhyt, mutta ok"* sijoitettiin laatu-kategoriaan sen vuoksi, että *"mutta ok"* tekee kommentista puhetuotoksen laatuun liittyvän kommentin. Sen sijaan *"näyte oli kovin lyhyt"* liittyy suoraan tuotoksen määrään ja siten se sijoitettiin määräkategoriaan.

Oppijat antoivat myös kommentteja, joita ei voitu sisältönsä puolesta sijoittaa mihinkään yllä mainituista kategorioista, vaan ne sijoitettiin muutkategoriaan. Tällaisia kommentteja ei ollut määrällisesti kovin paljon, mutta suurimman ryhmän näistä kommentteista muodostivat kommentit, joissa oppijat totesivat näytteen puhujan olevan hyvin samantasoinen tai samanlainen kuin hän itse oli (= valittu yleensä *"ei"*-vaihtoehto vastauslomakkeesta). Lopuksi on syytä korostaa, että kommenttikategorioiden sisältämien kommenttien tarkalla määrällä ei ole tässä työssä perustavaa laatua olevaa merkitystä, koska työn pääfokus ei ole kielen piirteiden määrällinen tai laadullinen luokittelu. Tämän osuuden tarkoituksena on tuoda esille hyvin karkealla tasolla tendenssejä siitä, mihin kielen piirteisiin oppijat näyttävät eniten kiinnittävän huomiota kuunnellessaan nauhalta puhenäytteitä ja verratessaan omaa tuotostaan ja selviytymistään näihin näytteisiin. Kommenttikategorioiden esittelyn jälkeen tarkastelen kolmen eniten ja vähiten *"vaikea päättää"*-rasteja keränneen näytteen valintaan liittyviä seikkoja. Sen jälkeen analysoin kyseisiä näytteitä yksityiskohtaisemmin osallistujien kommenttien valossa.

Jokaisesta näytteestä on tehty tarkastelun helpottamiseksi tavanomaisen kirjoitetun tekstin mukainen nopea litteraatio, jossa noudatetaan seuraavia merkintätapoja:

TAULUKKO 26 Puhenäytteiden litterointimerkintöjä.

Merkki	Seloste
[ ]	Selvästi ranskan kielen ääntämissäännöistä poikkeava ääntäminen on merkitty hakasulkeisiin.
:	Näytteen puhujan tuotoksessa esiintyvä sanojen loppujen venyttäminen on merkitty kaksoispisteellä.
≤ ≥	Tuotokseen tai puhujaan liittyvät kommentit on merkitty suurempi/pienempi kuin tai yhtä suuri -merkkien sisään.
_	Näytteen puhujan tuotoksessa esiintyvät tauot on merkitty _-viivoiin. Viivojen lukumäärä kertoo karkeasti taukojen keston.
( )	Mahdolliset epäselvät kohdat tuotoksessa on merkitty sulkeisiin ( ) ja yleensä varustettu kommentilla.
< >	Tuotoksen pituus on merkitty suurempi/pienempi -merkkien sisään tuotoksen perään.

### 5.4.2.1 Tehtävän 1 kommenttianalyysi

Taulukosta 26 selviää tehtävä 1:n kuunneltujen näytteiden (26 kpl) saamien vprastien määrä, joka perustuu 40 % otokseen koko aineistosta.

TAULUKKO 27 ”Vaikea päättää” -rastien prosentuaalinen määrä analyysiin valituista opiskelijoiden vastauslomakkeista (40 % kaikista lomakkeista), tehtävä 1.

Näyte	Kommentoijat yhteensä	Kommentit yhteensä	”Vaikea päättää” -rasteja yhteensä	%	Näytteen taso	Muita huomioita
1	56	81	4	4,9	2	
2	43	48	3	6,2	3	
3	52	71	3	4,2	1	
4	39	35	13	37,1	2	
5	44	57	12	21,0	5	
6	40	43	20	46,5	3	Yksimielisyys
7	41	58	3	5,1	1	Yksimielisyys
8	48	70	6	8,5	2	
9	52	94	3	3,1	5	
10	39	55	7	12,7	1	Yksimielisyys
11	31	37	12	32,4	2	
12	42	71	3	4,2	5	
13	36	28	17	60,7	3	
14	40	48	18	37,5	4	
15	35	29	15	51,7	3	
16	37	36	17	47,2	3	
17	35	47	4	8,51	1	
18	38	50	15	30,0	3	
19	34	34	11	32,3	2	
20	35	34	17	50,0	3	
21	32	23	15	65,2	3	
22	38	48	8	16,6	2	
23	43	56	15	26,7	3	
24	41	38	16	42,1	4	
25	35	54	7	12,9	3	
26	37	50	11	22,0	2	

Eniten vprasteja keräsi näyte 21 (65,2 %), toiseksi eniten näyte 13 (60,7 %) ja kolmanneksi eniten näyte 15 (51,7 %). Tarkasteltaessa näitä kolmea näytettä voidaan havaita, että kaikki kolme näytettä edustavat B1-tasoa. Suurin osa tutkimukseen osallistuneista oppijoista oli kielitaidoltaan tasoa A2–B1, ja siksi voidaankin todeta, että tutkimukseen osallistuneille oppijoille vaikeinta tuntui olevan verrata omaa suoritustaan nauhalta kuunneltaviin näytteisiin, jotka olivat likimain samantasoisia kuin heidän oma suullinen tuotoksensa.

Vähiten vp-rasteja keräsivät näytteet 9 (3,1 %), 3 (4,2 %) ja 12 (4,2 %). Nämä kolme näytettä sijoittuvat taitotasoasteikon ääripäihin siten, että näyte 3 edustaa alinta taitotasoa A1 ja kaksi muuta näytettä taitotasoa C1. Näiden huomioiden jälkeen voidaan todeta, että tutkimukseen osallistujilla ei näytä olleen vaikeuksia oman suorituksensa vertaamisessa joko taitotasoasteikon ylä- tai alapäähän sijoittuviin näytteisiin.

Merkille pantavaa vertailun vuoksi on se, että ne näytteet, joita tarkasteltiin mittareiden luotettavuutta koskevassa työn osassa ja jotka siinä osassa osoittautuivat huonosti toimiviksi (näytteet 7, 17, 10, 12 ja 9), eivät näytä olleen millään tavalla vaikeita tutkimukseen osallistuneille, kun he vertasivat omaa suoritustaan kyseisiin näytteisiin. Kuten taulukosta 26 näkyy, näiden näytteiden vp-rastien prosenttiosuudet ovat hyvin alhaisia ja vaihtelevat noin 3–12 %:n välillä. Jälleen on syytä todeta, että myös näiden näytteiden tarkastelu näyttää vahvistavan oikeaksi sen, että taitotasoasteikon ääripäitä (tässä tapauksessa taitotasoa A1 ja C1) edustaneet näytteet eivät näytä aiheuttaneen ongelmia oppijoille, kun he vertasivat omaa suoritustaan kyseisiin puhenäytteisiin.

Alla tarkastellaan lyhyesti niitä yhteensä kuutta näytettä, jotka keräsivät oppijoilta eniten tai vähiten "vaikea päättää" -rasteja molemmissa tehtävissä. Tuotoksen litteraation jälkeen seuraa lyhyt tutkimuksen tekijän yleisluonnehdinta puhetuotoksesta, minkä jälkeen näytteitä tarkastellaan oppijoilta saatujen kommenttien valossa erityisesti siitä näkökulmasta, mitkä piirteet kussakin tuotoksessa näytävät kiinnittäneen oppijoiden huomion.

#### *Eniten "vaikea päättää" -rasteja keränneet näytteet*

Seuraavassa tarkastellaan yksityiskohtaisemmin tehtävän 1 kolmea eniten "vaikea päättää" -rastia kerännyttä näytettä. Kaikki tässä luvussa tarkasteltavat näytteet edustavat siis taitotasoa B1. Ensimmäisenä tarkasteluvuorossa on näyte 21, joka keräsi jopa 65 % "vaikea päättää" -rasteja tutkimukseen osallistuneilta oppijoilta.

#### **Näyte 21 (B1)**

Euh\_ je xxx ≤ ei saa selvää ≥ avec François\_ qui est \_ \_un garçon\_ il se repose [ääntyy repo:z] après \_ euh \_il est joué \_ au foot \_ \_euh\_ il est été\_ parce que: \_ il a seulement un \_ un t-shirt et jeans et chaussures du tennis \_ euh ≤ miettii ≥ \_ \_ et \_ probablement \_ et \_ il est un \_étudiant\_ euh\_ qui ne travaille pas \_ pendant \_ l'été <34 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, melko lyhyt näyte, jättää melko sujuvan vaikutelman, mutta toisaalta puhuja miettii jonkin verran (ei keksi sanottavaa?), näytteen lyhyiden vuoksi vaikea arvioida sujuvuuden todellista tasoa, esitys selkeä, joitakin ongelmia peruskieliopin

hallinnassa (mm. artikkelien käyttö), ääntäminen kokonaisuutena hyvä, hyvätasoinen B1-tason näyte.

Oppijat ovat kiinnittäneet eniten huomiota näytteen puhujan sujuvuuteen (13 kommenttia) ja kielioppiin liittyviin seikkoihin (13 kpl). Alimman tason oppijat ovat pitäneet puhenäytettä sujuvana, kun taas ylimmän tason oppijoiden mielestä näytteen puhujan sujuvuudessa on ollut toivomisen varaa.

### Näyte 13 (B1)

François\_ aime les sports euh et il \_\_ le fait \_\_ euh \_\_ chaque jour\_ maintenant il est en train de jouer au vol volley ball \_ il a un t- [tee] shirt rouge et des pantalons gris elle\_ il a aussi des chaussures sportifs <25 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, erittäin lyhyt tuotos, tuottaa tekstiä todella hitaasti, vaikeuksia ilmaista itseään ranskan kielellä, käyttää aivan perussanastoa puhetta tuottaessaan.

Näytteen puhuja on saanut oppijoilta eniten ääntämiseen liittyviä kommentteja (17 kpl). Puhujan ääntämistä on pidetty yleisesti melko huonona. Sujuvuuteen liittyvät kommentit ovat seuraavaksi suurin ryhmä (13 kpl). Niissä korostetaan näytteen hitautta ja osittaista takeltelevuutta. Myös kieliopin hallinnan puutteeseen liittyviä kommentteja on melkein sama määrä kuin sujuvuuteen liittyviä kommentteja (12 kpl). Sisältökategoriasta löytyy kommentteja mm. siitä, että puhuja keskittyy vain kuvattavan henkilön ulkonäön kuvaamiseen.

### Näyte 15 (B1)

Mais\_ \_ je crois que François\_ euh il a l'air un peu triste parce que\_ il voudrait jouer du \_basket ball et \_ il a demandé ses camarades et \_\_ou euh\_ ses ses amis\_ euh pour jouer le\_ basket \_et\_euh\_ maintenant\_ il doit seulement s'asseoir et \_ attendre si quelqu'un voudrait venir parce que \_ tout le monde [mon] euh \_ \_ \_ tout le monde [mon] sont dit que \_ \_ qu'on doit mang - manger \_ et après ça\_ peut-être \_ ≤painottaa edellistä sanaa voimakkaasti ≥ on peut jouer le basket ball \_ et \_ il peut\_ seulement attendre <49 s.>

Yleisluonnehdinta puhetuotoksesta:

miesääni, melko sujuvaa, huomattavasti parempi tuotos kuin edellä, vaikka kaikki kolme näytettä edustavat B1-tasoa, hyviä rakenteita, melko hyvä ääntäminen, tuotos pidempi kuin edellä (49 s.).



Oppijoiden kommentit keskittyvät suurelta osin sujuvuus-kategoriaan (16 kpl). Alimpien tasojen oppijat pitivät näytettä erittäin sujuvana, kun taas ylimmän tason oppijoiden mielestä tämä B1-tasoisien näytteen puhuja ei selviytynyt tilanteesta sujuvuuden osalta heitä itseään paremmin. Seuraavaksi eniten saatiin sisältöön liittyviä kommentteja (13 kpl), joita oli harvinaisen paljon tämän näytteen osalta (muissa kahdessa edellä olevassa näytteessä vastaavia kommentteja oli vain 5 kappaletta (näyte 21) ja 6 kappaletta (näyte 13)). Osa oppijoista oli sitä mieltä, että näytteen puhujalla oli hyvä mielikuvitus ja että hän kertoi hyvin tarinaa, kun taas toiset moittivat puhujan käsitelleen aihetta liian yksipuolisesti (*puhui vain koripallosta, ei henkilöstä...*).

On huomattava, että näyte 15 on näyte, jonka tasosta alkuperäisen Yki-korpuksen 5 arvioijaa ovat olleet erimielisiä; 3 arvioijaa on sijoittanut näytteen B1-tasolle, 1 arvioija jopa tasolle B2 ja samoin yksi tasolle A2 eli arvioijien arvioinnit ovat jakaantuneet siten kolmelle eri tasolle. Tämänkaltainen arvioijien yksimielisyyden puute saattaa osittain selittää sitä, että myös oppijoiden on ollut vaikea sijoittaa näytettä tietylle tasolle ja verrata kyseistä näytettä omaan tuotukseensa.

Yhteenvetona eniten "vaikea päättää" -rasteja keränneistä tehtävä 1:n näytteistä todettakoon, että oppijat kommentoivat kaikkien näytteiden osalta paljon puheen sujuvuuteen liittyviä seikkoja. Merkillepantavaa on myös se, että sujuvuutta kommentoitaessa huomiota kiinnitettiin puheen tuottamisen hitauteen ja takelteluun erityisesti näytteessä 13. Kyseinen näyte sai myös osakseen melko negatiivisia ääntämiseen liittyviä kommentteja, joiden lisäksi kielioppiin liittyvät kommentit olivat yleisiä.

#### *Vähiten "vaikea päättää" -rasteja keränneet näytteet*

Tässä osassa tarkastellaan tehtävän 1 vähiten "vaikea päättää" -rasteja keränneitä näytteitä. Kuten edeltä jo selvisi, kaikkien kolmen näytteen puhujat edustavat taitotasoasteikon ääripäitä eli ovat joko aloittelevia ranskan oppijoita tai jo pitkälle suullisen kielitaidon hallinnassa yltäneitä oppijoita. Ensimmäisenä tarkasteluvuorossa on C1-tasoinen näyte 9, jonka keräämä "vaikea päättää" -rastien prosentuaalinen osuus oli 3,1 %.

#### **Näyte 9 (C1)**

Bon je vois François \_ \_ François est un garçon \_ il a environ vingt-cinq ans \_ ou en tout cas il a moins de trente ans \_ il a des cheveux blonds et des yeux \_ gris ou bleus et \_ il a l'air très sportif \_il\_ il porte des pantalons \_ des jeans \_ et \_ un t-shirt\_ \_à côté de lui il y a un basket ball \_ il a \_ \_ j'ai l'impression que\_ il va aller ou il vient de venir \_ \_ euh jouer \_ au basket avec ses copains \_ euh \_ \_ il euh \_ \_ il travaille dans un restaurant\_ il\_ \_euh \_ et euh \_ maintenant j'ai l'impression que c'est samedi parce qu'il travaille pas et et le samedi et le week-end\_ il \_ normalement il passe son temps avec ses copains les copains ils sont très importants pour lui \_ il\_ il fait du sport avec ses copains \_ il va à la

plage et au montagne en hiver \_ il fait du slalom \_ et tout ce qu'il fait il aime bien faire avec ses copains et tout ce qu'il fait c'est très sportif <1 min, 23 s.>

Yleisluonnehdinta puhetuotoksesta:

naisääni, pitkä tuotos, sujuvaa, vaikka puhuja miettiikin välillä selvästi, mitä sanoisi; hyvä ääntäminen.

Oppijat ovat kommentoineet eniten puhujan sujuvuutta (64 kommenttia), jota he ovat pitäneet yleisesti nopeana, luontevana, luonnollisena ja sujuvana. Jotkut tosin mainitsivat, että puhuja mietti jonkin verran, mistä aiheutuu pätkittäisyyden vaikutelma näytteessä. Seuraavaksi eniten (49 kommenttia) näyte on kerännyt ääntämistä koskevia kommentteja. Näytteen puhujaa epäiltiin ranskan kielen natiivipuhujaksi<sup>23</sup>, ja hänen ääntämistään keuhuttiin erittäin hyväksi ja sujuvaksi.

### Näyte 3 (A1)

Voilà François \_ \_ il est étudiant \_ \_ \_ \_ il est ici avec \_ son \_ basket ball \_ \_ il jouait [joye:e] au basket ball \_ \_ \_ \_ euh \_ \_ \_ \_ \_ il est très sportif \_ \_ \_ \_ \_ \_ \_ \_ il veut devenir [deveni:r] \_ \_ \_ \_ \_ un professionnel \_ au basket ball \_ \_ \_ \_ \_ \_ \_ \_ \_ il veut \_ \_ \_ jouer [joye] au basket ball \_ \_ \_ aux ≤ ei sidontaa ≥ Etats Unis \_ \_ \_ <1 minuutti>.

Yleisluonnehdinta puhetuotoksesta:

naisääni, puhetuotos erittäin katkonainen ja hidas, joutuu miettimään kauan ennen kuin saa sanottua, mitä haluaa sanoa, käyttää aivan perussanastoa ja rakenteita.

Oppijat ovat kommentoineet tämänkin näytteen osalta eniten ääntämistä ja sujuvuutta, joista molemmista on annettu 46 kommenttia. Ääntämistä on yleisesti pidetty huonona ja sujuvuus-kategoriassa on mainittu pitkät tauot, sanojen hakeminen ja pitkät miettimisajat, kommenttien pätkittäisyys, epäröinti ja hitaus.

### Näyte 12 (C1)

Il s'agit d'un garçon \_ il s'appelle François\_ il porte un pantalon\_ des jeans- probablement et \_ un t-shirt et il aussi \_ il porte aussi des basquets\_ les passe-temps préférés de François sont\_ jouer au basket avec ses amis \_ et il aime faire ça après l'école\_ il va à l'école encore parce que\_ il a pas passé encore son bac \_mais il le fera [fera] ce printemps \_ \_ ses projets d'avenir sont \_ \_ s'incriner à l'université\_ et il voudrait étudier \_ il voudrait étudier\_ les mathématiques mais il aime aussi les sports donc \_ \_comme il\_ il aime bien jouer au basket \_ c'est

<sup>23</sup> Kyseessä ei ollut äidinkieleltään ranskan-kielinen henkilö.

aussi \_ quelque chose\_ qu'il voudrait faire comme métier \_ François est un garçon très aimable\_ il est très ouvert envers \_ les autres \_ il aime bien \_ travailler et être avec ses amis <1 min, 11 sek.>

Yleisluonnehdinta puhetuotoksesta:

naisääni, erittäin hyvä ääntäminen, erittäin sujuva ja loogisesti rakentunut kokonaisuus.

Tätäkin näytettä on kommentoitu eniten sujuvuuden näkökulmasta (46 kommenttia). Sujuvuutta on kommentoitu muun muassa seuraavasti: erittäin sujuva, nopea, luonteva ja spontaani. Seuraavaksi eniten (32 kpl) kommentteja on kerännyt ääntämiskategoria; näytteen puhujan ääntämistä luonnehdittiin mm. erinomaiseksi, luonnolliseksi, luontevaksi ja ranskalaiseksi.

Yhteenvetona vähiten vaikea päättää -rasteja keränneistä tehtävän 1 näytteistä todettakoon, että oppijat kommentoivat kaikkien näytteiden osalta eniten puheen sujuvuuteen liittyviä seikkoja. Pitkät tauot puheessa näyttivät häiritsevän olennaisesti kommentoijia (näyte 3). Myös ääntämistä ja kielioppia koskevat kommentit olivat runsaita kaikissa kolmessa esitellyssä näytteessä.

### 5.4.2.2 Tehtävän 2 kommenttianalyysi

Taulukosta 28 selviää tehtävän 2 kuunneltujen puhenäytteiden (23 kpl) saamien ”vaikea päättää” -rastien lukumäärä, joka perustuu 40 % otokseen koko aineistosta.

TAULUKKO 28 ”Vaikea päättää” -rastien prosentuaalinen määrä analyysiin valituista opiskelijoiden vastauslomakkeista (40 % kaikista lomakkeista), tehtävä 2.

Näyte	Kommenttoijat yht.	Kommentit yht.	Vaikea päättää yht.	%	Taitotaso	Muita huomioita
1		44	2	4,5	1	Yksimielisyys
2		32	8	25,0	1	
3		33	16	48,4	3	
4		26	13	50,0	4	
5		22	7	31,8	2	
6		31	9	29,0	3	
7		17	13	76,4	3	
8		24	5	20,8	5	
9		28	12	42,8	2	
10		44	3	6,8	1	
11		26	17	65,3	3	
12		32	12	37,5	3	
13		21	12	57,1	2	
14		31	4	12,9	2	
15		17	16	72,7	3	
16		26	2	7,6	2	
17		43	19	44,1	3	
18		50	5	10,0	6	
19		40	3	7,5	2	
20		24	13	54,1	2	
21		33	8	24,2	5	
22		28	16	57,1	3	
23		33	15	45,4	3	

Tehtävässä 2 eniten ”vaikea päättää” -rasteja keräsivät näytteet 7 (76,4 %), 15 (72,7 %) ja 11 (65,3 %). Kun tarkastellaan näitä kolmea näytettä, voidaan tehdä sama havainto kuin tehtävässä 1 eli että kaikki kolme näytettä edustavat taitotasoa B1. Siten näihin kolmeen näytteeseen liittyvät kommentit ovat niin ikään samansuuntaisia kuin tehtävässä 1. Tutkimukseen osallistujilla tuntuu siis olleen vaikeuksia verrata omaa suoritustaan suunnilleen samantasoisin suorituksiin.

Vähiten ”vaikea päättää” -rasteja keräsivät näytteet 1 (4,5 %), 10 (6,8 %) ja 19 (7,5 %). Nämä kolme näytettä sijoittuvat taitotasoasteikon alapäähän siten, että näytteet 1 ja 10 edustavat alinta tasoa A1 ja näyte 19 taitotasoa A2. Tutki-

mukseen osallistujilla ei siten näytä olleen vaikeuksia myöskään tehtävässä 2 näiden kolmen taitotasoasteikon alempaan ääripäähän sijoittuvan näytteen kanssa.

Mittarin luotettavuutta käsitelleessä osioissa esille nousivat aiemmin huonosti toimineina näytteinä näytteet 1, 18 ja 21. Tehtävässä 2 näiden näytteiden saamat ”vaikea päättää” -rastien prosenttiosuudet ovat hieman suuremmat kuin tehtävässä 1 vaihdellen 4,5 %:n ja 24,2 %:n välillä. Nämä kolme näytettä edustavat jälleen, samoin kuin tehtävässä 1, taitotasoasteikon ääripäitä siten, että näyte 1 on A1-tason näyte, näyte 18 C2- ja näyte 21 C1-tason näyte. Tässä yhteydessä voidaan mainita, että vaikka näiden kolmen näytteen vp-rastien prosenttiosuudet eivät sijoitu lähellekään eniten vp-rasteja saaneiden näytteiden prosenttiosuuksia, niin erityisesti näytteessä 21 osallistujilla on ollut jonkin verran vaikeuksia verrata omaa suoritustaan tähän näytteeseen, mistä kertoo melkein 25 %:n vp-rastien prosenttiosuus.

*Eniten ”vaikea päättää” -rasteja keränneet näytteet*

Yleishuomiona voidaan aluksi todeta, että tehtävän 2 eniten ”vaikea päättää” -rasteja keränneiden kolmen näytteen prosenttiosuudet ovat huomattavasti korkeammat kuin vastaavien kolmen näytteen tehtävässä 1. ”Vaikea päättää” -rastien prosenttiosuudet tehtävässä 2 vaihtelevat 65,3–76,4 %:n välillä, kun taas tehtävän 1 eniten ”vaikea päättää” -rasteja keränneiden kolmen näytteen prosenttiosuudet asettuvat välille 51,7–65,2 %. Osallistujat ovat siten selvästi pitäneet huomattavasti vaikeampana oman suorituksensa vertaamista tiettyihin tehtävän 2 kuin tehtävän 1 näytteisiin.

Tarkastelemme nyt vuorotellen kaikkia kolmea tehtävän 2 eniten ”vaikea päättää” -rastia kerännyttä näytettä. Ensimmäisenä tarkasteluvuorossa on näyte 7.

### **Näyte 7 (B1)**

Et \_ voilà \_ euh \_ monsieur Julien Fabre-Perrault \_ et voilà \_ euh ma collègue Kristiina Saukkonen \_ \_ et \_ ma collègue Kristiina Saukkonen \_ elle parle très bien français <16 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, hieman jäykkä esittely, puhetuotos melko hidas ja vähän katkonainen, kollegan nimi väärin esittelyssä, ääntäminen melko selkeää.

Näyte keräsi jopa 38 ”vaikea päättää” -rastia, joista suurinta osaa oppijat eivät selitä millään tavalla. Sosiolingvistiseen puoleen liittyvät kommentit olivat enemmistönä tässä näytteessä (10 kpl), ja ne liittyivät muun muassa kahden henkilön esittelyjärjestykseen. Toiseksi eniten kommentteja tuli ääntämiskategoriaan (6 kpl). Kolmanneksi eniten annettiin sujuvuuteen liittyviä kommentte-

ja, joita oli kuitenkin vain 5 kappaletta. Yleisesti ottaen näyte 7 ei kerännyt paljoakaan ”vaikea päättää” -rastien runsasta määrää selittäviä kommentteja, sillä kommenttien määrä vaihtelee eri kategorioissa yhdestä kommentista sosiolingvistisen osaamisen 10 kommenttiin. Silti on selvää, että näyte on ollut ongelmallinen oppijoille vertailutilanteessa.

### Näyte 15 (B1)

Euh \_ voici \_ euh ma collègue madame Kristiina Saukkonen-Riipinen \_ euh \_ voici \_ c'est monsieur Julien \_ Fabre-Perrault \_ ma collègue parle français très bien <14 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, melko sujuva ja napakka näyte, kuitenkin hieman taukoja, ääntäminen hyvä.

Näyte 15 keräsi paljon ”vaikea päättää” -rasteja (38 kpl) sekä jonkin verran kommentteja muihin kategorioihin. Sujuvuuteen liittyvät kommentit, jotka yleensä totesivat näytteen olevan sujuva, olivat jälleen enemmistönä (14 kpl). Seuraavaksi eniten kommentteja kertyi ääntämiskategoriaan (10 kpl); ääntäminen sai pelkästään myönteistä palautetta oppijoilta. Puhujan sosiolingvistiseen selviytymiseen liittyviä kommentteja kerääntyi kolmanneksi eniten (8 kpl). Näytteen kommentoijat totesivat puhujan esittelyn olleen selkeä.

### Näyte 11 (B1)

Et voilà est-ce que je peux vous \_ présenter \_ mon \_ \_copine\_ \_ madame Saukkonen-Riipinen \_ \_et Kristiina \_ voilà \_ monsieur Julien Fabre-Perrault \_ de \_ de France \_ \_ \_ monsieur Fabre-Perrault \_ madame Saukkonen parle très bien français \_ \_ \_ mieux que moi <30 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, luonnolliselta kuulostava esittely, copine-sanan (po. collègue) käyttö häiritsee muuten melko hyvää kokonaisuutta, miellyttävä puheääni ja hyvä puhetyyli.

Kielioppi-kategoriaan liittyvät kommentit ovat enemmistönä tässä näytteessä (17 kpl). Valtaosa kommentteista keskittyy väärän sanavalinnan ympärille (*copine*). Seuraavaksi eniten kommentteja näyte keräsi sujuvuudesta (12 kpl). Kolmanneksi eniten kommentoitiin sosiolingvistiseen puoleen liittyviä seikkoja (11 kpl). Nämä kommentit jakaantuivat niihin, joiden mielestä puhuja oli esittelysään kohtelias ja selkeä sekä niihin, jotka epäilivät, oliko puheen tyyli tai rekisteri sittenkään oikeanlainen kielenkäyttötilanteeseen nähden.

Vähiten "vaikea päättää" -rasteja keränneet näytteet

Tehtävän 2 kolme vähiten "vaikea päättää" -rasteja kerännyttä näytettä sijoituvat kaikki taitotasoasteikon alatasoille (taso A1 = näytteet 1 ja 10 ja taso A2 = näyte 19). Tehtävässä 1 kaksi kolmesta vähiten "vaikea päättää" rasteja keränneestä näytteestä edusti taitotasoa C1 ja vain yksi näyte tasoa A1. Tarkastelemme aluksi näytettä 1.

**Näyte 1 (A1)**

Bonjour \_euh\_ c'est ma collègue Kristiina Saukk \_Saukkonen-Riipinen\_ il est \_arrivé \_ \_ \_euh\_ \_ \_ \_ \_ quelxxx ≤epäselvä≥\_ \_ \_ pour visiter \_ le \_euh\_ <20 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, epäröivä ja hidas tuotos, puhuja takeltelee eikä saa tehtävää suoritettua tehtävänannon mukaisesti.

Näyte 1 keräsi vain kaksi "vaikea päättää" -kommenttia eikä se muutenkaan kerännyt määrällisesti kovin paljo kommentteja (44 kommenttia). Sujuvuuskategoriaan sijoittuvat kommentit ovat hallitsevia, sillä niitä on eniten (14 kpl). Tätä aloittelevan kielenoppijan puhetuotosta kommentoitiin mm. siksi, että siinä ei saatu tehtyä/sanottua sitä, mitä olisi pitänyt ja lisäksi tuotoksen sujumattomuutta luonnehdittiin mm. adjektiiveilla tankkaava, epäröivä ja katkonainen. Toiseksi eniten kommentteja (10 kpl) keräsi näytteen sociolingvistinen ulottuvuus, joka tässä tehtävässä liittyy kahden henkilön, joista toinen on tärkeässä asemassa, toisilleen esittelyyn ammattikontekstissa. Oppijat kommentoivat mm. sitä, että esittelyä ei saatu tehtyä ja että tilannetta ei saatu hoidettua loppuun asti. Myös näytteen sisältö sai osakseen 6 kommenttia, joissa viitattiin tehtävän suorittamatta jättämiseen ja siihen, että näytteen puhuja ei tiennyt, mitä sanoi tai että hän ei sanonut mitään järkevää.

**Näyte 10 (A1)**

Euh \_ \_ je \_ vous invite \_euh\_ mon collègue \_ ma collègue Kristiina Saukkonen \_euh\_ et \_euh\_ ici \_ monsieur Julien Fabre-Perrault \_ \_euh\_ \_ \_ Kristiina Saukkonen \_ parle français \_ bien <23 s>

Yleisluonnehdinta puhetuotoksesta:

naisääni, hidas, katkonainen ja epäröivä tuotos, ääntämisessä vieraan kielen vaikutusta, käyttää esitellä-verbinä kutsua-verbiä (*inviter*), mikä hankaloittaa kielenkäyttötilanteen hahmotusta.

Näyte keräsi eniten (17 kpl) sujuvuuskategoriaan kuuluvia kommentteja. Kommentit keskittyivät tuotoksen hitauteen ja ei-sujuvuuteen. Seuraavaksi eni-

ten (14 kpl) kommentteja keräsi kielioppi-kategoria; kommentoijista noin puolet kiinnitti huomiota väärän verbin käyttämiseen esittelytilanteessa. Myös ääntämisen liittyviä kommentteja saatiin melkein yhtä paljon kuin kielioppikategoriaan luokiteltuja kommentteja (13 kpl). Näistä kommentteista suurin osa koski ääntämisen heikkoutta, mutta myös englannin kielen vaikutus puhujan ääntämisessä mainittiin.

### Näyte 19 (A2)

Je voudrais [vudrie:]\_ \_ \_ \_ je voudrais \_ présenter vous\_ euh mon collègue  
\_Kristiina Saukkonen \_ \_ euh voilà \_ Julien Fabre-Perrault [perro:t] \_ \_ \_ \_mon  
collègue parle [parle]\_français \_ excellente <25 s>

Yleisluonnehdinta puhetuotoksesta:

miesääni, ei kovin sujuva, hitaasti etenevä, joutuu miettimään paljon tuottaessaan, ääntämisongelmia.

Sujuvuuteen liittyvät kommentit ovat tässäkin näytteessä enemmistönä (26 kpl). Kommenteissa todetaan, että puhuja etenee hitaasti, pitää taukoja ja on epäröivä. Toiseksi suurin kommenttikategoria on ääntämiskategoria, johon kommentteja kertyi 17 kappaletta. Niistä suurimmassa osassa todetaan puhujan ääntämisen olevan puutteellista tai huonoa. Kielioppiin liittyvät seikat saivat kolmanneksi eniten kommentteja (7 kpl). Kommentoijien mukaan puhujan tuotos sisälsi kieliopillisia virheitä.

Yhteenvedona tehtävään 2 liittyvistä eniten tai vähiten vp-rasteja keränneistä näytteistä voidaan todeta seuraavaa: B1-tasoiset eniten vp-rasteja keränneet näytteet 7, 15 ja 11 sisältävät kaikki sosiolingvistiseen puoleen liittyviä kommentteja (kolme suurinta kommentoitua kategoriaa). Vähiten vp-rasteja sisältävistä näytteistä vain näyte numero 1 sisälsi niin paljon tehtävälle 2 ominaisia sosiolingvistisiin seikkoihin liittyviä kommentteja, että tämä kategoria pääsi kolmen suurimman kommenttikategorian joukkoon. Lisäksi on syytä mainita, että sujuvuuteen liittyvät kommentit esiintyvät kolmen suurimman kommenttikategorian joukossa kaikissa kuudessa tarkastellussa näytteessä.

#### 5.4.2.3 Yhteenvedo "vaikea päättää" -rasteja keränneiden puhenäytteiden ominaisuuksista

Tutkimukseen osallistuneet oppijat kiinnittivät kaikissa tarkastelluissa näytteissä tasaisesti huomiota puhetuotoksen sujuvuuteen. Tästä seurasi se, että sujuvuus-kategoria nousi jokaisen kuuden tarkastellun näytteen osalta kolmen eniten kommentteja keränneen kategorian joukkoon. Sujuvuus oli suurin kategoria kaikissa tehtävää 2 koskevien vähiten vp-rasteja keränneiden taitotasosteikon alapäähän sijoittuvien näytteiden kohdalla.

Ei ole todennäköisesti mitenkään sattumaa, että tutkimukseen osallistujat kommentoivat niin runsaasti sujuvuuteen liittyviä seikkoja. Sujuvuus on



yksi suullisen kielitaidon käytetyimmistä arviointikriteereistä ja ratkaisevimista tekijöistä (Huhta 1993; Elsinen & Juurakko-Paavola 2006; Iwashita ym. 2008), ja siten voidaan olettaa, että valtaosa ihmisistä yhdistää hyvään suulliseen kielitaitoon sujuvuuden vaikutelman. Sujuvuutta koskevien kommenttien suuresta määrästä huolimatta on hyvin todennäköistä, että jos oppijoita pyydetäisiin erikseen määrittelemään, mitä heidän mielestään sujuvuus tarkoittaa kielitaidon osalta, asiasta saataisiin hyvin monenlaisia tulkintoja. Oppijoidenhan ei tarvinnut tässä tutkimuksessa keskittyä määrittelemään sitä, miten heistä kukin sujuvuuden ymmärtää. Näkemys sujuvuuden vaikutelmasta puhe- tuotoksessa tuntuuikin olevan yhteydessä hyvin moniin kielen ja sen käyttämiseen liittyviin tekijöihin. Yhtäjaksoinen ja epäröimätön puhetuotos liitetään usein sujuvuuteen, ja siten puheessa esiintyviin katkoihin, hidastuksiin tai ”miettimiseen” saatetaan suhtautua kielteisesti. Kuitenkin on syytä muistaa, että puhutulle kielelle on ominaista esimerkiksi miettimisajan saaminen käytämällä puhutulle kielelle tyypillisiä täytesanoja tai lausumisen pitkittämistä, joiden seurausta juuri ovat puheessa esiintyvät katkot ja epäröinnit.

Onko sujuvuutta koskevien kommenttien suuren lukumäärän kohdalla siis kyse siitä, että sujuvuutta ei olekaan helppo määritellä yksiselitteisesti, ja siten jokaisella kielenoppijalla on oma sisäinen käsityksensä siitä, mikä kunkin mielestä on sujuvaa puhetta ja mikä ei? Vai onko kenties kyse siitä, että sujuvuuden yksiselitteisen määrittelyn hankaluus saa aikaan sen, että jokainen oppija tuntee hallitsevansa sen, mikä on sujuvuutta – ilman, että kukaan voisi välttämättä tulla kumoamaan heidän sujuvuutta koskevaa määrittelyään? Vai onko sittenkin mahdollisesti kyse siitä, että sujuvuus on käsitteenä oppijoiden mielissä sen verran ”epämääräinen” (verrattuna esimerkiksi johonkin kieliopin hallintaan liittyvään kielen piirteeseen), että kuunneltavat tuotokset on ikään kuin helppo jakaa sujuviin ja ei-sujuviin ilman sen suurempia perusteluita? Tämän tutkimuksen tehtävänä ei ole löytää vastauksia sujuvuuden määrittelyyn eikä sen ensisijaisuuteen oppijoiden kommentteissa, joten yllä olevat pohtivat kysymykset jäävät siten tarkempaa vastausta vaille tämän työn osalta.

Tehtävään 2 liittyvä sosiolingvistiseen kielenhallintaan kuuluva kategoria oli kolmen suurimman kommenttikategorian joukossa silloin, kun oppijat olivat antaneet paljon ”vaikea päättää” -rasteja B1-tasoisille kolmelle näytteelle. Kielen sosiolingvistiseen hallintaan liittyvät seikat näyttävät siten tuottavan päänvaihavaa oppijoille; tämä toteamus pätee eritoten silloin, kun oppijat vertaavat omaa suoritustaan suunnilleen omaa tasoaan vastaavan tasoiseen suoritukseen.

Puhenäytteiden sisältöön liittyviä kommentteja saatiin oppijoilta hyvin niukasti. Tämä saattaa johtua tehtävien luonteesta mutta myös siitä, että puhe-suoritusta arvioitaessa arvioidaan usein nähtävästi ikään kuin vanhaan tottumuksen pohjautuen vain suorituksen kieliopillisia komponentteja (kuten ääntäminen ja sanasto) ja/tai kielellisiä ominaisuuksia (esim. sujuvuus). 2000-luvulla tehdyissä tutkimuksissa on kuitenkin selvitetty, että suullisten puhe-suoritusten arvioijat kiinnittävät arvioidessaan paljon huomiota testin suorittajan puheen sisältöön (esim. Brown 2007; Eckes 2009). Muun muassa Sato (2011, 224) onkin kritisoinut sitä, että jo puhumisen arviointia koskevien arviointias-

teikkojen laadinnassa suorituksen sisältö jätetään usein huomioimatta. Täten toimimalla ikään kuin sanattomasti vihjataan, että suullinen suoritus toimii näytteenä ainoastaan testin suorittajan/puhujan kieliopillisesta kyvystä.

Yleisenä toteamuksena molempien tehtävien osalta voidaan tuoda esille, että sujuvuus-kategorian lisäksi ääntämis- ja kielioppikategoriat yltyvät yleensä kommenttien määrän perusteella kolmen suurimman kommenttikategorian joukkoon sekä tehtävässä 1 että tehtävässä 2.

## 6 LUOTETTAVUUSTARKASTELU JA JOHTOPÄÄTÖKSET

### 6.1 Tutkimusprosessin tarkastelua

Tässä tutkimuksen osassa edetään siten, että ensin käydään läpi tutkimuksen tulokset, jotka antavat vastauksia työn alussa esitettyihin tutkimuskysymyksiin. Tutkimuskysymysten vastausten käsittelyn jälkeen punnitaan tutkimuksen ja siinä käytettyjen itsarviointivälineiden luotettavuutta kokonaisuutena. Tarkastelun teoreettisena pohjana toimii Bachmanin ja Palmerin testin laadun ja hyödyllisyyden tarkasteluun kehitetty malli (Bachman & Palmer, 1996). Koska tutkimus on empiirinen ja sisältää monia erilaisia käytännön työvaiheita, Bachmanin ja Palmerin mallin pohjalta mietitään tutkimuksen toteuttamiseen liittyviä valintoja ja toimenpiteitä sekä sitä, kuinka onnistuneita tietyt valinnat olivat sen suhteen, mitä alun perin tutkimuksessa oli tarkoitus tehdä. Tällä tarkastelulla pyritään arvioimaan myös sitä, minkälaisia mahdollisia vaikutuksia tietyillä toimenpiteillä on saattanut olla tutkimustuloksiin. Tutkimuksessa käytettyjen mittainstrumenttien luotettavuus ja toimivuus eivät näin ollen ole ainoita seikkoja, joilla voi olla merkitystä tutkimuksesta saataviin tuloksiin. Luvussa 6.4 keskitytään tutkimuksen rajoitusten tarkasteluun, minkä jälkeen ovat vuorossa ehdotukset mahdollista jatkotutkimusaiheista.

Tutkimuksen pääpaino oli sen selvittämisessä, kuinka luotettava ja käytännössä toimiva on uudenlainen itsearviointimenetelmä. Tätä itsearviointimenetelmää ja sen tuloksia verrattiin sekä perinteisempien itsearviointimenetelmien että ulkoisen arvioinnin avulla saatuihin tuloksiin. Tutkimuksen empiirinen osa alkoi käytettyjen mittavälineiden analysoinnilla tilastollisia mittalukuja käyttäen.

Tämän lisäksi tutkimukseen sisältyi myös kvalitatiivinen osa. Kvalitatiivisen osan tarkoituksena oli tuoda lisävalaistusta tilastollisista analyyseistä saatujen tulosten esiin tuomiin seikkoihin. Tarkoituksena oli myös paljastaa sellaisia oppijoiden itsearviointiin liittyviä näkemyksiä, tunteita ja asenteita, joita ei

olisi saatu tarkasteluun tilastollisten menetelmien avulla. Tutkimuksessa panostettiin siten, niin kuin empiirisessä tutkimuksissa kuuluukin, sen todentamiseen, että tutkimusprosessin aikana käytetyt mittavälineet ovat riittävän laadukkaita, jotta niiden antamia tuloksia voidaan pitää luotettavina.

## 6.2 Vastauksia tutkimuskysymyksiin

Tässä osassa esittelen tutkimuksen antamia vastauksia tutkimuskysymyksiin. Tutkimuskysymyksistä kolme ensimmäistä (1, 2 ja 3) liittyvät itsearviointimenetelmien toimivuuteen sekä itsearviointien ja ulkoisten arviointien ja itsearviointien keskinäiseen vastaavuuteen. Tutkimuskysymykset 4, 5 ja 6 puolestaan koskevat oppijoiden käsityksiä kirjallisista itsearviointiväittämistä ja niiden toimivuudesta sekä tutkijan havaintoja siitä, millaiset kieleen liittyvät tendenssit ja piirteet nousivat esiin, kun oppijat suorittivat itsearviointia kuunneltavien puhenäytteiden perusteella.

### Tutkimuskysymys 1

*Kuinka toimivia ovat oppijoiden käyttämät erilaiset itsearviointimenetelmät?*

Tutkimuksessa selvisi, että kaikkien kolmen käytetyn itsearviointimittarin reliabiliteetti oli hyvä, ja siten kunkin mittarin osioiden voidaan todeta mitan neen samantyyppisiä asioita (mittarien sisäinen yhtenäisyys oli siis korkea). Tehtävä 1 erottui jälleen tehtävästä 2 korkeamman reliabiliteettia mittaavan Cronbachin alfa -kertoimensa vuoksi (Cronbachin alfa tehtävässä 1 ja paperilla tehdyssä itsearvioinnissa oli molemmissa .94, kun taas tehtävässä 2 vastaava reliabiliteettikerroin oli .89). Kun kaikkien kolmen itsearviointimittarin laadukkuutta tarkasteltiin perinteisen osioanalyysin näkökulmasta keskittyen tarkastelemaan arvon .30 ylittävien osioiden osio-testi-korrelaatioiden prosentuaalista määrää, hätiköityjen johtopäätösten vaara mittareiden keskinäisen paremmuuden suhteen oli suuri. Kuten edellä työssä jo kävi ilmi, prosentuaalinen .30 korrelaatiokertoimien ylittävien osioiden määrä on 85 % sekä paperilla tehdyssä itsearvioinnissa että tehtävään 1 liittyvässä puhenäytteisiin pohjautuvassa itsearvioinnissa. Mittareiden laadun arviointia ei ole kuitenkaan syytä tehdä vain .30 osio-testi-korrelaation ylittävien osioiden prosenttiosuutta tarkastelemalla, vaan tarkasteluun on otettava lisäksi .30 raja-arvon ylittäneiden osioiden korrelaatioiden suuruusluokat. Tässä tarkastelussa vasta saadaan todellinen käsitys muun muassa siitä, että tehtävän 1 puhenäytteiden kuunteluun perustuvan itsearviointimittarin laatu oli huomattavasti korkeampi kuin paperilla tehdyn itsearvioinnin tai tehtävään 2 liittyvän itsearviointimittarin laatu. Tämä toteamus rakentuu sille tosiasialle, että tehtävän 1 osioista noin 65 % ylittää .60 korrelaatioarvon, kun vastaava arvo samassa suuruusluokassa paperilla tehdyssä itsearviointilomakkeessa on 28 % ja tehtävän 2 osalta vain noin 21 %.

Rasch-analyysistä saadut itsearviointimittareiden laadukkuutta koskevat tulokset olivat samansuuntaisia kuin perinteisen osioanalyysin tulokset. Infit mean square -arvon, jonka tulisi sijaita arvojen .70 ja 1.30 välillä, prosentuaaliset osuudet olivat hyvin samansuuruisia kaikissa kolmessa itsearviointimittarissa (88,4–92,5 % osioiden Infit-arvoista sijoittui näiden raja-arvojen välille). Jos taas tarkastelu tässäkin kohdassa ulotetaan hieman pidemmälle kuin vain mittareiden osioiden kokonaisprosenttiosuuksiin, voidaan huomata, että tehtävää 1 koskeva itsearviointimittari sisälsi kaikista kolmesta mittarista eniten niitä osioita, joiden Infit-arvo oli paras mahdollinen eli 1.0 (27 %). Vastaavaan suuruusluokkaan sijoittuvat prosenttiosuudet paperilla tehdyssä itsearviointimittarissa (12,5 %) ja tehtävää 2 koskevassa itsearviointimittarissa (17,3 %) olivat jälleen huomattavasti pienempiä.

Mittainstrumenttien laadukkuuteen liittyen on syytä muistuttaa jo aiemmin työssä mainitusta seikasta, joka vaikuttaa kaikkiin kolmeen instrumenttiin ja joka liittyy perinteisen osioanalyysin esiin tuomiin taitotasosteikon ylä- tai alapäähän sijoittuviin ongelmallisiin osioihin. Perinteisen osioanalyysin lähtökohtana näet on ajatus, jonka mukaan puolet esimerkiksi johonkin kokeeseen osallistuvista oppijoista vastaa kysymyksiin oikein toisen puolen taas vastatessa kysymyksiin väärin. Tämänkaltainen lähtökohta ei kuitenkaan sovi sellaisiin aineistoihin, joissa osallistujien valtaosan kielen osaaminen keskittyy taitotasosteikon keskivaiheille ja joissa on siis hyvin vähän todellisia taitotasosteikon ääripäitä edustavia oppijoita. Ongelmaksi tällaisessa tapauksessa muodostuu se, että suurin osa keskitasolle sijoittuvista oppijoista vastaa jompaankumpaan taitotasosteikon ääripäähän kuuluvaan väittämään tai näytteeseen joko ”kyllä” (asteikon alatasoille sijoittuvat osiot) tai ”ei” (asteikon ylätasoille sijoittuvat osiot), ja siten vastauksiin ei saada riittävästi hajontaa. Tästä seuraa variaation vähyys vastauksissa. Variaation vähyys taas puolestaan aiheuttaa sen, että huonosti toimineen osion ja mittarin muiden osioiden välillä ei ole yhteisvaihtelua. Taitotasosteikon ääripäihin sijoittuvien näytteiden huonoon toimimiseen tuleekin suhtautua varauksella, sillä perinteisten tilastollisten analyysien esille tuoma huono toimivuus saattaa olla täysin tekninen vähäiseen variaatioon liittyvä ominaisuus.

Tutkimuksen aikana ilmeni myös, että kahdelle korkeimmalle taitotasolle (C1 ja C2) sijoittuvat väittämät ja puhenäytteet sekoittuvat keskenään. Tutkimus vahvistaa siten lukuisissa yhteyksissä todetun tendenssin, jonka mukanaan C1- ja C2-tasoa on vaikea määritellä niin, että ne erottuisivat toisalta toisistaan ja toisaalta alempana asteikossa sijaitsevasta B2-tasosta.

Tehtävän 2 osalta on vielä korostettava, että tehtävän puhenäytteet olivat keskimäärin vain noin 17 sekunnin pituisia (tehtävän 1 näytteiden keskimääräinen pituus oli noin 35 sekuntia). Tehtävän 2 näytteiden lyhyys saattoi vaikuttaa kielteisesti niiden toimivuuteen. Näytteiden alkuperäisen tason arvioijilla oli jonkin verran vaikeuksia määritellä näytteiden taitotaso ja ainakin joillakin tutkimukseen osallistuneista oppijoista saattoi olla vaikeuksia oman suorituksensa vertaamisessa lyhyeen näytteeseen. Tämänkaltaisissa tutkimuksissa kannattaisikin varmasti tulevaisuudessa määritellä puhenäytteille jokin mini-

mipituus, jotta tutkimuksen toteuttaminen ei vaikeutuisi tämänkaltaisen yksityiskohdan vuoksi. Sopiva minimipituus voisi olla esimerkiksi 30–35 sekuntia.

Yhteenvedonomaaisesti voidaan todeta, että tutkimuksessa käytettyjen itsearviointimittareiden osiot olivat kokonaisuudessaan hyvin toimivia ja laadukkaita etenkin, kun otetaan huomioon niiden määrä (42 kpl / kirjalliset väittämät + 26 kpl / tehtävän 1 puhenäytteet + 23 kpl / tehtävän 2 puhenäytteet). Ainoastaan 42 itsearviointiväittämää sisältävästä itsearviointilomakkeesta oli poistettava yksi väittämä sen huonon toimivuuden vuoksi. Uuteen itsearviointimenetelmään liittyvien kuunneltavien puhenäytteiden joukosta ei sen sijaan tarvinnut poistaa niiden heikon laadun vuoksi yhtään puhenäytettä.

Lopuksi on syytä todeta joitakin huomioita tutkimuksessa käytetyistä analyysimenetelmistä. Tämän tutkimuksen väittämien toimivuutta tarkasteltaessa lähtökohtana voitiin pitää oletusta, jonka mukaan tutkimuksessa käytetyt kirjalliset itsearviointiväittämät perustuvat laajaan tutkimukseen ja muodostavat siten pääosin toimivan asteikon/mittarin. Kuitenkin tässä tutkimuksessa ainoastaan sellaiset selvästi ongelmalliset osiot kuin osio 29, jonka korrelaatioarvo oli negatiivinen, nousi esille asteikon keskivaiheilta perinteisessä osioanalyysissä. Näyttäisikin siltä, että ainakin tutkimuksessa käytetyn kaltaisella aineistolla perinteinen osio-testi-korrelaatio pystyy nostamaan esille mahdollisia ongelmallisia osioita vain asteikon ääripäistä (tässä tosin edellä selostettu vastausten epätasainen jakauma saattaa selittää osan tuloksista). Sitä vastoin Rasch-analyysi vaikuttaisi pystyvän osoittamaan mahdollisia ongelmaosioita myös muualta kuin asteikon ääripäistä. Kuten edellä on jo eri yhteyksissä mainittu, huono osio-testi-korrelaatio helpoissa tai vaikeissa osioissa saattaa usein olla tekninen tai keinotekoinen ilmiö. Rasch-analyysi vaikuttaisi siis olevan sekä herkempi havaitsemaan mahdollisia ongelmakohtia asteikon keskitason osioissa että vahvistamaan sen, onko asteikon ääripäihin sijoituvissa helpoissa tai vaikeissa osioissa todellisia toimimattomuusongelmia vai ei. Kuten kaikissa analyyseissä, joskus esille voi nousta myös ongelmia, jotka sitten yksityiskohtaisemmassa tarkastelussa eivät muodostukaan ongelmallisiksi mittarin kannalta. Rasch-analyysi kuitenkin näyttää identifioivan suuremman joukon mahdollisia ongelmallisia osioita, jotta ne voidaan tarvittaessa ottaa lähempään tarkasteluun.

Yllä olevaa perinteistä osioanalyysiä ja Rasch-analyysiä koskevan pohdinnan pohjalta voidaan todeta, että perinteinen osio-testi-korrelaatioihin perustuva analyysi yksin käytettynä tai sen yhdistäminen laadulliseen aineistoon ei näytä riittävän tehokkaalta tavalta saada tietoa mittarin mahdollisesti ongelmallisista osioista koko mittarin tasolla. Perinteinen osioanalyysihän ei näyttänyt identifioivan mittarin keskivaiheille sijoituvia mahdollisesti ongelmallisia osioita. Sen vuoksi tämänkaltaisten itsearviointivälineiden laadun tai toimivuuden analysoinnissa suositeltavin lähestymistapa on juuri modernin testiteorian mukaisen osioanalyysin ja laadullisen aineiston tarkastelun yhdistäminen.

## Tutkimuskysymys 2

*Missä määrin erilaiset itsearviointit vastaavat toisiaan?*

Tutkimuksessa käytetyt eri itsearviointit osoittivat jokseenkin hyviä keskinäisiä vastaavuuksia. DIALANG-hankkeesta lainattu 42-kohtainen itsearviointilomake oli toisena osapuolena kaikissa kolmessa suurimpia korrelaatioita saaneissa yhdistelmissä. Korkein korrelaatiokerroin (.902) oli itsearvioinnissa, jossa verrattiin toisiinsa DIALANG-hankkeesta lainattuja suullisen kielitaidon yleistasokuvauksia (A1–C2) ja 42-kohtaista tilannekohtaista itsearviointiväittämää sisältävää mittaria. Seuraavaksi korkein korrelaatiokerroin (.763) liittyi tutkimuksessa kokeiltuun uuteen itsearviointimenetelmään, eli kyseessä oli 42-kohtaisen itsearviointilomakkeen ja tehtävää 1 koskevan 26 puhenäytteen välinen suhde. Kolmanneksi korkein korrelaatiokerroin (.748) löytyi 42-kohtaisen itsearviointilomakkeen ja tehtävien 1 ja 2 yhdistettyjen puhenäytteiden kohdalla. Lopuksi todettakoon, että itsearviointivälineiden ”pituus” eli mahdollisten vastausvaihtoehtojen kokonaislukumäärä mittarissa voi myös vaikuttaa havaittujen korrelaatioiden suuruuteen. 6-portaisella asteikolla tehdyn yleisitsearviointin hiukan matalammat korrelaatiot muihin itsearviointivälineisiin saattavat osittain selittyä myös sillä, että siinä on vain kuusi mahdollista vastausvaihtoehtoa, kun taas väittämiin perustuvassa itsearvioinnissa saavutetun tuloksen (pistemäärän) vaihteluväli voi periaatteessa olla 0–42 (sitä vastaavassa Rasch-asteikossa on myös suurin piirtein sama määrä pistemäärävaihtoehtoja). Puhe-suorituksiin perustuvissa itsearvioinneissakin arvioitavia näytteitä - ja siis mahdollisia pisteitä - oli reilusti yli 20 kummassakin tehtävässä (kuten myös niistä lasketuissa Rasch-arvoissa). Jos kaikkien mittarien laatu on muuten samaa tasoa, niin pelkästään tuloksissa esiintyvä suurempi vaihtelu (jos se on oikean suuntaista) tuottanee hiukan parempia korrelaatiota ”pitempien” asteikkojen / mittareiden välille verrattuna vain 6-portaiseen asteikkoon.

## Tutkimuskysymys 3

*Missä määrin oppijoiden tekemät itsearviointit vastaavat ulkoista arviointia?*

Kolmanteen tutkimuskysymykseen saadaan vastauksia tarkastelemalla tutkimuksen aikana tehtyjen tilastollisten analyysien tuloksia, joissa on selvitetty sitä, missä määrin oppijoiden itsearviointi ja ulkoinen arviointi vastaavat toisiaan. Jotta voidaan tarkastella, miten luotettavaa oli tutkimukseen osallistuneiden oppijoiden itsearviointi erilaisilla itsearviointimenetelmillä, palataan tarkastelemaan kvantitatiivisten analyysien tuloksia, joissa verrattiin oppijoiden itsearviointeja ulkoiseen arviointiin sekä niihin tuloksiin, joissa verrattiin eri itsearviointeja toisiinsa.

Vastauksia ensimmäiseen tutkimuskysymykseen saatiin selvittämällä itsearviointien ja ulkoisten arviointien välisiä korrelaatioita. Samassa yhteydessä analysoitiin myös eri itsearviointityyppien välisiä korrelaatioita. Itsearviointeja

verrattiin ensin yhdistettyihin ulkoisiin arviointeihin kokonaisvaltaisesti siten, että ulkoisten arvioijien arvioinnit sisälsivät yhdistetyn oppijan puhumistason molemmissa tehtävissä (tehtävät 1 ja 2). Tämän jälkeen oppijoiden itsearviointeja verrattiin ulkoiseen arviointiin niin, että ulkoista arviointia edusti arvioijien antama oppijan puhumistaso ensin tehtävässä 1 ja sen jälkeen tehtävässä 2. Itsearviointitapoja oli siis käytössä kolmenlaisia:

- 6-portainen yleisitsearviointi (A1-C2)
- tilannekohtainen 42-kohtainen paperilla tehty itsearviointi ja
- kuunneltaviin puhenäytteisiin perustuva itsearviointi tehtävissä 1 ja 2.

Ulkoisen arvioinnin ja itsearvioinnin väliset korrelaatiokertoimet osoittivat, että voimakkaimmat yhteydet itsearvioinnin ja ulkoisen arvioinnin väliltä löytyivät, kun itsearviointia edusti DIALANG-hankkeesta lainattu 42 yksityiskohtaista väittämää sisältävä itsearviointilomake. Itsearviointien ja ulkoisten arviointien välinen korrelaatio oli korkeimmillaan (.792) 42-kohtaisen itsearviointilomakkeen ja ulkoisen arvioinnin välillä silloin, kun ulkoinen arviointi muodostui oppijan yhdistetystä puhumistasosta tehtävissä 1 ja 2. Toiseksi korkein korrelaatio (.768) saatiin, kun 42-kohtaisen itsearviointilomakkeen avulla tehtyä itsearviointia verrattiin ulkoiseen arviointiin silloin, kun ulkoista arviointia edusti oppijan puhumistaso tehtävässä 1. Kolmanneksi korkein itsearviointien ja ulkoisten arviointien välinen korrelaatio (.753) saatiin, kun oppijan tehtävää 1 koskevaa puhumistaitoa verrattiin tehtävän 1 puhenäytteisiin (26 kpl). Tehtävää 2 koskevat ulkoisten arviointien ja itsearviointien väliset korrelaatiokertoimet olivat yleisesti huomattavasti matalampia kuin tehtävän 1 tai molempien tehtävien yhdistetyt korrelaatiokertoimet.

DIALANG-järjestelmän suullisen kielitaidon itsearviointiosasta lainattu empiirisesti validoitu 42 väittämää sisältävä tilannekohtainen itsearviointilomake kuului tutkimuksen hyvin toimiviin mittavälineisiin, kun toimivuuden mittapuuna pidetään korkeaa korrelaatiota suhteessa ulkoiseen arviointiin. Tämä oli hyvin odutuksenmukaista, koska lomake on empiirisesti validoitu DIALANG-hankkeen aikana. Myös kuunneltaviin puhenäytteisiin perustuva uudentyyppinen itsearviointi tehtävässä 1 korreloi varsin hyvin ulkoiseen arviointiin. Sen sijaan tehtävässä 2 vastaava suhde arviointien välillä jäi heikommaksi; korrelaatiokerroin tässä tapauksessa oli .648. Kun otetaan huomioon, että tehtävä 1:n kuunneltaviin puhenäytteisiin perustuva itsearviointi suhteessa ulkoiseen arviointiin saavutti analyysissä melko hyvän korrelaatiokertoimen (.753), joka ei ole kovinkaan kaukana toiseksi parhaasta korrelaatiokertoimesta (.768; ulkoinen arviointi ja 42 väittämää sisältävä tilannekohtainen itsearviointi), on paikallaan todeta, että tehtävä 2 sisältää mitä todennäköisimmin sellaisia ominaisuuksia, jotka vaikuttivat siihen, että se ei toiminut yhtä hyvin mittainstrumenttina kuin tehtävä 1. Tehtävän 2 ominaisuuksia ja mahdollisia syitä sen heikompaan toimimiseen pohditaan tarkemmin jäljempänä kohdassa, jossa tarkastellaan tutkimuksen autenttisuutta Bachmanin ja Palmerin mallin pohjalta.



Lopuksi on syytä todeta, että vaikka ulkoisen ja itsearvioinnin välillä esiintyi varsin korkeitakin korrelaatioita, arvioinnit näyttivät kuitenkin perustuvan myös osin erilaisiin asioihin. Esimerkiksi korrelaatio .8 tarkoittaa, että kyseiset kaksi arviointia perustuvat noin 2/3 mittareiden yhteiselle vaihtelulle eli jollekin samalle taidolle (ranskan puhumistaito). Kuitenkin noin 1/3 tai jopa suurempikin osa tämän tutkimuksen mittareiden yhteisvaihtelusta perustuu johonkin muuhun seikkaan kuten esimerkiksi mittausvirheeseen. Todennäköistä edelleen on, että mittareista kumpikin mittaa jotakin sellaista, mitä toinen mittari ei tavoita.

## Tutkimuskysymys 4

*Mitä käsityksiä oppijat esittävät paperilla esitettyjen itsearviointiväittämien toimivuudesta?*

Ennen tilannekohtaisia itsearviointiväittämiä koskevien kommenttien tarkastelua mainitsen lyhyesti puhumisen itsearviointilomakkeen etusivulla olevan yleistasoarvion (A1-C2) valintaan liittyviä oppijoiden kommentteja. Yleistasoarvion valintaan liittyviä kommentteja ei ollut kovin paljoa (10 kpl), mutta kommentteissa toistuvat tietyntyyppiset oppijoiden kokemat vaikeudet. Vaikeuksia tuotti muun muassa se, että yleistasoarvion valinta kuudesta taitotasosta saattoi tuntua oppijasta hankalalta tai jopa sattumanvaraiselta. Osa oppijoista totesi kommentteissaan, ettei ollut varma omasta yleistasostaan, ja ainakin yhdellä yleistason valintaa kommentoineella oppijalla tutkimukseen osallistumishetkellä tehty yleistason valinta tuntui olevan riippuvainen oppijan senhetkisestä mielentilasta tai asenteesta (positiivisuus vs. negatiivisuus, hetkellinen itseluottamus vs. itseluottamuksen puute). Lisäksi osa oppijoista koki, että kuuden yleistaitotason sanalliset kuvaukset eivät olleet riittävän kattavia kaikenlaisen kielitaidon oikealle tasolle sijoittamiseen (esimerkiksi kahden tason välimaastoon sijoittuva kielitaito).

Yleishuomiona paperilla tehdystä 42-kohtaisesta itsearviointilomakkeesta voidaan todeta, että oppijat antoivat vähemmän kommentteja niistä lomakkeen väittämistä, jotka heistä tuntuivat toimivilta kuin niistä, jotka tuntuivat ongelmallisilta. Toisaalta on huomattava, että lomakkeen muoto ohjasi paljolti siihen, että kommentteja tuli enemmän niissä tapauksissa, kun jokin väittäjä ei tuntunut oppijoista toimivalta. Puhumisen itsearviointilomakkeen viimeisellä sivullahan kysyttiin, oliko oppijoilla mitään erityisiä vaikeuksia vastata lomakkeen kysymyksiin ja samalla pyydettiin tarkennusta siitä, millaisia mahdolliset vaikeudet olivat. Lisäksi oppijoita pyydettiin mainitsemaan mahdollisen tavallista hankalamman väittämän numero kommentteissaan. Siitä huolimatta, että oppijat kommentoivat enemmän ongelmallisiksi kokemiaan väittämiä kuin muita väittämiä, saatiin väittämistä kerättyä myös positiivisia kommentteja.

Taulukossa 29 esitellään yhteenvetona allekkain tutkimuksessa tilastollisten analyysien mukaan hyvin tai huonosti toimineille väittämille ominaisia piirteitä ja annetaan kyseiseen ominaisuuteen liittyviä esimerkkejä (sulkeissa väittämän perässä oleva numero viittaa väittämän numeroon lomakkeessa).

TAULUKKO 29 Tutkimuksessa hyvin tai huonosti toimineiden itsearviointiväittämien ominaisuuksia.

Toimivan väittämän ominaisuus	Esimerkki hyvin toimineesta väittämästä (yllä) Esimerkki huonosti toimineesta väittämästä (alla)
Teksti lyhyt tai melko lyhyt	Osaan puhua ranskaa luontevasti, sujuvasti ja tehokkaasti tutuissa tilanteissa. (2)
	Osaan vastata haastattelussa esitettyihin, yksinkertaisiin ja suoriin henkilökohtaisia asioita koskeviin kysymyksiin, kun ne esitetään hitaasti ja selkeästi. (9)
Mieluiten yksiosainen (ei konjuktiolla yhdistettyjä lauseita)	Osaan kysyä toisten ihmisten työstä ja vapaa-ajasta sekä vastata toisten vastaaviin kysymyksiin. (14)
	Osaan ylläpitää keskustelua, mutta minua voi joskus olla vaikea ymmärtää, kun yritän sanoa tarkasti, mitä haluan. (23) → miten vastata väittämän kahteen eri osaan, jos henkilö ei osaa ylläpitää keskustelua (=ei) ja häntä voi olla vaikea ymmärtää (=kyllä)?
Sisältö konkreettinen, henkilöön itseensä tai hänen elämänsä tai kiinnostuksen kohteisiinsa liittyvä	Osaan kertoa melko yksityiskohtaisesti kokemuksistani ja kuvata tunteita ja reaktioita. (40)
	-
Väittämässä annetaan esimerkki siitä, mitä jollakin väittämän sanalla tai ilmauksella tarkoitetaan	Osaan kuvailla ja määritellä konkreettisia asioita, joille en juuri sillä hetkellä satu muistamaan sopivaa sanaa (esimerkiksi pysäköintisakko, luottokortti tai vakuutuskirja). (42)
	Osaan palata takaisin vaikeisiin kohtiin ja muotoilla uudestaan sanottavani niin, että toiset tuskin huomaavat sitä. (6) → mitä tarkoittaa vaikeisiin kohtiin palaaminen?
Väittämässä viitataan vain rajalliseen määrään (1-2 kpl) kielellistä toimintaa	Pystyn ilmaisemaan itseäni niin hyvin, että jopa syntyperäiset kielenpuhujat luulevat minun asuneen maassa kauan aikaa. (28)
	Osaan esittäytyä ja esitellä muita ihmisiä ja käyttää tavanomaisia tervehdys- ja hyvästelyilmauksia. (41) → kolme kielellistä aktiviteettia samassa väittämässä
Väittämän sisältö vastaa tai on lähellä oppijan oman kulttuurin käytänteitä	-
	Osaan ilmaista yksinkertaisia numeroita, määriä, hintoja ja aikoja. (19) → ranskan kielen suomen kielestä poikkeava tapa muodostaa kymmenlukuja

Suurin osa oppijoiden puhumisen itsearviointilomaketta koskevista kommentteista liittyi siis jonkinlaisiin vaikeuksiin vastata kirjallisiin itsearviointiväittämiin. Väittämien muotoilut, jotka eivät oppijoiden mielestä olleet tarpeeksi yk-

siselitteisiä, saivat aikaan paljon kommentteja. Tästä esimerkkinä voidaan ottaa tiettyjen sanojen tai ilmausten tulkintaa koskevat vaikeudet. Esimerkiksi käsitteet *tavallinen keskustelu* (väittäjä 22), *vaikea asia* (väittäjä 15) ja *yksinkertainen kysymys* (väittäjä 32) saivat oppijat pohtimaan, mitä käsitteillä tarkoitetaan. Sanomattakin on selvää, että tutkimuksessa tai itsearviointissa yleensäkin käytettyjen itsearviointiväittämien tulisi olla niin selkeitä ja yksitulkintaisia, ettei oppija joutuisi tuntemaan itseään epävarmaksi tulkintansa oikeellisuudesta. Väittämien monitulkintaisuudella saattaa olla odottamattomia seurauksia sekä oppijoiden vastauksiin että tutkimuksen tuloksiin.

Paljon kommentteja keräsi myös vastausvaihtoehtojen rajoittuminen dikotomisiin ”kyllä”- ja ”ei”-vaihtoehtoihin. Jotkut oppijat pitivät yksinkertaisesti tämänkaltaisia vastausvaihtoehtoja liian karkeina ja mustavalkoisina. Tähän samaan eli vaihtoehtojen liialliseen rajoittuneisuuteen viitataan myös Floropouloun DIALANG-järjestelmää koskevassa tutkimuksessa, jossa oppijat toivoivat, että ”kyllä”- ja ”ei”-kategorioiden välillä olisi ollut vastaamista helpottavia välikategorioita, jollaiseksi ehdotuksena oli ”joskus”-vastauskategoria (Floropoulou 2002a).

Kolmas ison määrän kommentteja keränneistä aiheista oli se, että oppijat pitivät väittämiä toisiaan muistuttavina ja samantyyppisinä ja kertoivat tämän asian vaikeuttaneen joissakin tapauksissa väittämiin vastaamista. Hankalinta tutkimuksen teon näkökulmasta on se, että osa oppijoista näytti epäilevän, että kysymysten samankaltaisuus olisi ollut harkittua ja että sillä olisi pyritty testaamaan jotain muuta kuin mitä tutkimuksen tavoitteeksi oli selostettu. Selvää luonnollisesti on, että ei ole hyvä asia, jos oppijoiden mieleen tutkimuksen aikana nousee epäilyksiä, että tutkimuksella yritetään selvittää jotain muuta kuin sitä, mistä osallistujille on kerrottu. Toisaalta kun otetaan huomioon, että paperilla tehdyssä itsearviointissa käytetty lomake oli aika pitkä ja sisälsi 42 itsearviointiväittämiä, on selvää, että osa väittämistä saattaa muodoltaan tai sisällöltään muistuttaa toisiaan. Tämän vuoksi tutkimuksen alussa kuitenkin voisi olla hyödyllistä mainita, että vaikka osa käytetyistä itsearviointiväittämistä saattaa vaikuttaa sisällöltään melko samanlaisilta, niillä on jokaisella kuitenkin oma perusteltu paikkansa itsearviointimittarin osana. Ennen itsearviointiväittämien käyttöä voisi olla perusteltua selostaa oppijoille, että väittämät on pyritty laatimaan niin, että niissä kysytään vain yhtä asiaa kerrallaan. Tästä johtuen monet väittämistä ovat melkein identtisiä ja vain niiden avaintermi vaihtuu.

Eräs paljon kommentteja kerännyt seikka oli se, että monien oppijoiden tuntui olevan vaikeaa kuvitella ja sijoittaa mielessään itsensä sellaiseen suulliseen kielenkäyttötilanteeseen, jossa he eivät todellisuudessa olleet koskaan olleet. Kyseessä onkin oppijan edeltävän kokemuksen kannalta joko roolileikki tai simulaatio (Huhta & Suontausta 1993). Jos tilanne oli oppijalle tuttu, hän voi simulaation keinoin asettaa itsensä väittämän kuvaamaan tilanteeseen. Jos taas väittämässä esitetty tilanne oli tuntematon oppijalle, hänen oli ikään kuin asetettava sellaiseen rooliin, joka mahdollisti kyseisen kielenkäyttötilanteen hahmottamisen. Tutkimuskäyttöön lomakkeita laadittaessa on otettava huomioon, että lomakkeessa tulee olla eritasoisia väittämiä, jotta saadaan eritasoisilta kie-

lenoppijoilta vastauksia väittämiin ja riittävää variaatiota vastauksiin. Selvää on myös, että itsearviointilomakkeita, joilla on tarkoitus arvioida oppijan kielitaidon sijoittumista taitotasosteikolle, ei voida laatia siten, että ne sisältäisivät esimerkiksi vain tietyille taitotasolle kuuluvia väittämiä. Itsearviointin tarkoituksenahan on muun muassa se, että oppija oppii tarkastelemaan omaa kielitaitoaan, sen vahvuuksia ja heikkouksia. Monentasoisia väittämiä sisältävä itsearviointilomake mahdollistaa sen, että oppija oppii vähitellen suhteuttamaan kielitaitonsa tason realistisesti taitotasosteikolle huomattessaan, että monet väittämät kuvaavat kielenkäyttötilanteita, joissa vaadittava kielitaidon taso on vielä toistaiseksi ylempänä kuin hänen oma taitotasonsa.

Itsearviointi kirjallisten itsearviointiväittämien avulla aiheutti monenlaisia pohdintoja siitä, tunsivatko tutkimukseen osallistuneet oppijat olevansa kelpoisia arvioimaan omaa suullisen kielitaitonsa tasoa vai eivät. Näissä kommentteissa oppijoita tuntui huolestuttavan se, että he joko yli- tai aliarvioivat omaa taitoaan, mitä tietenkin tapahtuikin joissain tapauksissa. Tämä ilmiselvästi kertoi siitä, että osa oppijoista epäili yleisellä tasolla omaa itsearviointitaitoaan. Todettakoon siis, että vaikka tämän tutkimuksen tulokset itsearviointin ja ulkoisen arviointin vastaavuuden osalta ovat hyvin myönteisiä, yksittäisillä oppijoilla saattaa olla tunne, että he eivät hallitse itsearviointia ja tarvitsevat siihen lisää harjoitusta, jotta he voisivat tuntea osaavansa verrata omaa osaamistaan suhteessa esimerkiksi kirjallisiin itsearviointiväittämiin.

Tutkimukseen osallistujat kommentoivat kirjallisia itsearviointiväittämiä usein myös siten, että kommentteista näkyi oppijoiden olevan hyvin vakuuttuneita siitä, että tänään väittämään vastaukseksi valittu "ei"-vaihtoehto voisi vastauspäivästä esimerkiksi viikon päähän sijoittuvana uutena itsearviointipäivänä hyvinkin muuttua "kyllä"-vastaukseksi. Todellisuudessa kenenkään suullisen kielitaidon taso ei ratkaisevasti muutu viikossa, mutta nämä kommentit liittyivätkin suoraan todellisiin ranskankielisiin kielenkäyttötilanteisiin, joista saamansa kokemuksen turvin erityisesti vähän korkeammalle taitotasolle edenneet oppijat olivat varmoja siitä, että heidän suullisen kielitaitonsa käyttövalmius paranisi huomattavasti hyvin lyhyessäkin ajassa, kunhan he vain pääsisivät käyttämään ranskaa suullisesti.

Joiltakin ylimmän tason oppijoilta saatiin palautetta, jonka pohjalta voitiin havaita, etteivät he olleet aina ymmärtäneet taitotasoarviointin periaatteita, joiden mukaan ylimpien tasojen kielenpuhujien voidaan olettaa selviytyvän useimmista tasoaan alemmista kielenkäyttötilanteista. Tämä asia tuli tutkimuksessa esille hieman yllätyksenä, ja muistutti tutkimuksen tekijää siitä, että mitään asiaa ei tule ottaa itsestäänselvyytenä tutkimusaineistoa ja sitä koskevaa ohjeistusta suunniteltaessa. Kaikille tutkimukseen osallistujille olisi siten ollut hyvä selostaa taitotasoarviointin perusasiat ennen tutkimukseen osallistumista vaikkapa etukäteen lähetetyssä viestissä.

Yleisesti ottaen on todettava, että puhumisen itsearviointilomakkeen yleis-tasoarvion valinnasta ja sen tilannekohtaisista itsearviointiväittämisistä saadut kommentit olivat erittäin valaisevia ja antoivat mahdollisuuden ymmärtää huomattavasti paremmin sitä, miten oppijat kokivat tämänkaltaisten kirjallisten itsearviointiväittämien käytön itsearviointinissa. Ilman puhumisen itsearviointi-

lomakkeesta kerättyä palautetta tutkimuksen anti olisi jäänyt huomattavasti vaisummaksi huolimatta tilastollisten analyysien kattavista tuloksista.

## Tutkimuskysymys 5

*Mikä on tyypillistä niille puhenäytteille, joissa oppijoilla oli eniten vaikeuksia verrata omaa puhetuotostaan kuunneltuihin näytteisiin?*

*Entä mikä on tyypillistä niille puhenäytteille, joissa esiintyi vähiten vaikeuksia?*

Perimmäinen syy siihen, että tässä tutkimuksessa kerättiin oppijoiden kommentteja puhenäytteistä on se, että kommenttien avulla yritettiin laajentaa ymmärrystä siitä, mitä itsearviointitilanteessa silloin tapahtuu, kun itsearviointia tehdään vertaamalla omaa suoritusta kuunneltaviin puhenäytteisiin. Sisällöllisesti oltiin kiinnostuneita siitä, mihin oppijat kiinnittivät huomiota verratessaan omaa puhe-suoritustaan ja kuunneltavia puhenäytteitä.

Tutkimuskysymys numero 5:tä varten tutkimusaineisto käytiin aluksi läpi 40 %:sesti ja merkittiin ylös kaikkien näytteiden saama ”vaikea päättää”-rastien kokonaismäärä. Kun molempien tehtävien näytteiden keräämien ”vaikea päättää”-rastien kokonaismäärä oli selvillä, kummastakin tehtävästä otettiin yksityiskohtaisempaan tarkasteluun kolme eniten ja kolme vähiten ”vaikea päättää”-rasteja kerännyttä näytettä (= yhteensä 6 näytettä). Tämän jälkeen koko aineiston kommentit käytiin läpi tarkasteluun valittujen kuuden näytteen osalta.

Tämä tutkimus vahvistaa osittain jo aiemmin joissain muissakin yhteyksissä todetun tendenssin, jonka mukaan taitotasosteikon keskivaiheille sijoituvan kielitaidon arviointi voi olla hankalaa sekä verratessa oman kielitaidon tasoa että arvioitaessa muiden kielitaidon tasoa (= ulkoinen arviointi) asteikon keskivaiheilla. Tämä tendenssi ilmeni tutkimuksessa siten, että oppijoilla oli eniten ”vaikea päättää”-rasteja niiden puhenäytteiden kohdalla, jotka sijoittuivat mittarin keskivaiheille. Toisaalta tämän tendenssin vastakkainen ilmiö tuli tutkimuksessa esille silloin, kun todettiin, että itsearvioinnin ja ulkoisen arvioinnin välinen vastaavuus oli parasta taitotasolla B1 (myös tasolla A2 vastaavuus oli kohtuullista). Mielenkiintoista on se, että kyseinen havainto on täysin vastakkainen sille yleiselle oletukselle, että oman suorituksen vertailu olisi vaikeaa kielen oppimisen alkuvaiheessa oleville kielen oppijoille.

Tutkimukseen osallistuneiden oppijoiden valtaosan kielitaito sijoittuu taitotasolle B1, ja eniten ”vaikea päättää”-rasteja keränneet kolme näytettä sekä tehtävästä 1 että tehtävästä 2 edustavatkin kaikki B1-tason näytteitä. Oppijoiden kokemiin vastausvaikeuksiin saattaa olla osasyynä jo useaan kertaan tutkimuksessa esille otettu dikotominen ”kyllä” ja ”ei”-vastausvaihtoehtojen rajallisuus. On selvää, että jos mietitään vastausvaihtoehtojen valintaa oppijan näkökulmasta, voidaan näiden kahden vaihtoehdon todeta rajoittavan tietyllä tapaa väittämiin vastaamista erityisesti niissä tapauksissa, joissa oppija mahdollisesti tunnisti näytteen puhujan kielitaidon olevan hyvin samantasoinen kuin

hänen oma kielitaitonsa. Toisaalta on muistettava, että tämänkaltaiset vastausvaihtoehdot olivat välttämättömiä tässä tutkimuksessa käytettyjen tilastollisten analyysien tarpeita silmällä pitäen.

Mitä tulee niihin näytteisiin, jotka keräsivät osakseen vähiten ”vaikea päättää” -rasteja sekä tehtävässä 1 että tehtävässä 2, on todettava, että kaikki 6 näytettä edustavat taitotasoasteikon ääripäitä eli sijoittuvat joko tasolle A1 / A2 tai C1. Nämä näytteet vaikuttavat siten olevan jollakin tasolla selvempiä ja ”helpompia” vertailun kannalta kuin muut näytteet.

Oppijoiden toimintatapa kommentoinnin suhteen ansaitsee muutaman toteamuksen tässä kohdassa. Osalla tutkimukseen osallistuneista oppijoista ei näet näyttänyt olleen vaikeuksia päättää missään tutkimukseen osallistumisensa vaiheessa, kumpi vastausvaihtoehto kahdesta valita. Esimerkkinä tällaisesta oppijasta olkoon eräs taitotaso B2 edustava oppija, joka oli kirjoittanut kommentteja jokaisesta tehtävän 1 ja 2 näytteestä, mutta jolla ei ollut yhtään ”vaikea päättää” -rastia vastauslomakkeissaan. Tällaisten havaintojen kohdalla tulee mieleen, että osa oppijoista näyttää tulkitsevan annettuja ohjeita oman tyyliinsä mukaisesti. On näet varsin vaikea uskoa, että jollakulla oppijalla ei olisi missään vaiheessa kuunnellussa 49 puhenäytettä (26 / tehtävä 1 + 23 / tehtävä 2) minkäänlaisia vaikeuksia verrata omaa puhe-suoritustaan kuultuihin näytteisiin. Mitä luultavimmin kyse on siitä, että kyseinen henkilö on joko tietoisesti päättänyt keskittyä vain lomakkeen pakollisten osien täyttämisen lisäksi antamaan mahdollisimman paljon kommentteja kuunnelluista näytteistä tai sitten kyse on yksinkertaisesti siitä, että kyseiseltä henkilöltä on jäänyt huomaamatta mahdollisuus kommentoida väittämiä myös ”vaikea päättää” -rastien avulla.

Samoin huomio kiinnittyi aineiston läpikäynnin aikana siihen, että joillakin osallistujilla, joiden kielitaito sijoittui melko alas taitotasoasteikossa, oli suhteellisen paljon ”vaikea päättää” -rasteja (ja ”ei”-valintoja) mutta ei minkäänlaisia selityksiä tai perusteluja sille, mikä teki valinnan vaikeaksi tietyn näytteen kohdalla. Ehkä tällaisessa tapauksessa kyseessä oli puutteellisesta kielen hallinnasta johtuva kykenemättömyys arvioida yksityiskohtaisesti, mikä kielen piirre aiheutti vaikeuksia valintatilanteessa.

Molempien tehtävien puhenäytteistä ja niiden kommentoinnista voidaan siten todeta yhteenvedonomaisesti seuraavaa:

- Suurin osa tutkimukseen osallistuneista oppijoista oli kielitaidoltaan tasoa A2–B1, ja heille tuntui olevan vaikeinta verrata omaa suoritustaan nauhalta kuunneltaviin näytteisiin, jotka olivat likimain samantasoisia kuin heidän oma suullinen tuotoksensa.
- Niillä näytteillä, jotka aiemmin työssä nousivat toimimattomuutensa vuoksi esille mittarin laatua koskevassa osassa, ei näyttänyt olevan yhteyttä niihin näytteisiin, joiden kohdalla oppijoilla oli vaikeuksia päättää, selviytyivätkö he tehtävästä paremmin vai huonommin kuin näytteen puhuja.

- Oppijoilla ei ollut vaikeuksia verrata omaa puhe-suoritustaan taito-tasoasteikon ylimpiä tai alimpia ääripäitä edustaviin näytteisiin kum-mankaan tehtävän kohdalla.
- Oppijoilla näytti olevan taipumus toistaa tiettyjä itselleen ominaisia toi-mintapoja läpi koko lomakkeen täytön.
- Kielitaidon alimmille tasoille sijoittuvien oppijoiden näyttäisi olleen jon-kin verran vaikeaa perustella, miksi he vastasivat selviytyneensä tehtä-västä heikommin tai samantasoisesti kuin puhenäytteen puhuja.

Yllä olevan listan kaksi ensimmäistä kohtaa liittyvät toisiinsa, ja niitä onkin paikallaan vielä pohtia lyhyesti. On mielenkiintoista havaita, että ne näytteet, jotka oppijat kokivat vaikeiksi arvioinnin tai vertailun kannalta (= vaikea päät-tää -rastit), toimivat tilastollisessa mielessä erittäin hyvin. Vaikuttaa siis siltä, että hankalaltakin tuntuva päätöksenteko/arviointi tuotti hyvin toimivia tulok-sia. Keskitasolle sijoittuvista osioista voidaan todeta yleisesti, että niiden ”keski-tasoisuus” on varmasti arviointia hankaloittava tekijä: niissä on vähän välittö-mästi havaittavia mittarin ääripäihin sijoittuvien osioiden ominaisuuksia, kuten esimerkiksi katkonaisuutta, tuottamiseen tai ääntämiseen liittyviä ongelmia (alatasot) tai äärimmäistä sujuvuutta ja kohdekielenomaista idiomaattista kie-lenkäyttöä (ylätasot).

## Tutkimuskysymys 6

*Mihin kielen piirteisiin oppijat kiinnittävät huomiota arvioidessaan kielitaitoaan uuden itsearviointimenetelmän avulla?*

Tarkasteltaessa, mihin kielen piirteisiin liittyviä seikkoja oppijat kommentoivat eniten tai vähiten ”vaikea päättää” -kommentteja keränneiden näytteiden kohdalla, nousee puheen sujuvuus-kategoria esille yhtenä kolmesta eniten kom-mentteja keränneestä kategoriasta molemmissa tehtävissä ja kaikkien tarkastel-tujen näytteiden kohdalla. Kaikentasoisilla oppijoilla tuntui olevan sanansa sanottavana siitä, onko jokin puhe sujuvaa vai ei. Kuten tunnettua, sujuvuuden määrittely on vaativa tehtävä (Elsinen & Juurakko-Paavola 2006, 62, Huhta 1993, 158; Luoma 2004, 87–90), joten tässä mielessä on hieman yllättävää, että sujuvuus painottuu niinkin suuresti oppijoiden kommentteissa. Oppijoiden sujuvuuteen liittämiä piirteitä tai tekijöitä näyttävät olevan ainakin seuraavat: puheen yhtäjaksoisuus (ei taukoja), luontevuus, luonnolliselta kuulostava puheen tuottaminen, epäröimättömyys puheen tuottamisessa ja riittävä puheen tuottamisnopeus. Sujuvuutta koskevia kommentteja oli paljon sekä eniten että vähiten ”vaikea päättää” -rasteja keränneissä näytteissä. Puheen sujuvuus tai sen vastakohta puheen sujumattomuus näyttääkin muodostavan yhden puhu-tun kielen peruspiirteistä, johon viittaamalla ainakin ei-kieliammatillaiset, ku-ten tähän tutkimukseen osallistuneet oppijat, määrittelevät puhetuotoksen hy-vää tai huonoa laatua. Sujuvuuden lisäksi kielen muut piirteet, joihin oppijat kiinnittivät suurissa määrin huomiota näytteitä kuunnellessaan, olivat puhujan ääntäminen ja hänen kieliopilliseen osaamiseensa liittyvät seikat.

Ääntäminen on tietenkin hyvin odotuksenmukaisesti kielen piirre, joka nousee esiin puhutun kielen kohdalla. Ääntämiseen liittyviä kommentteja olikin paljon kaikissa tarkastelluissa näytteissä. Yleisestä normista poikkeavaan ääntämiseen tai erityisen hyvänä pidettävään ääntämiseen suhtauduttiin osittain hyvin tunteenomaisesti, mistä kertovat tietyntyyppiset kommentit (esimerkiksi *"hirveä ääntäminen, mutta sujuva"* (A2); *"Oudosti lausuttu"* (B1)).

Kielioppikategoriaan kerääntyi kommentteja erityisesti niiden näytteiden kohdalla, joissa puhuja käytti väärää sanaa toisen sanan tilalla. On todettava, että väärän verbin tai sanan käyttäminen hyvin lyhyessä tuotoksessa saattaa vaikuttaa suorituksen arvioijien arviointeihin epäedullisella tavalla. Sen vuoksi tämältyyppisissä tutkimuksissa kannattaisi ehkä harkita, voisiko jonkin tehtävään liittyvän sanan antaa valmiina jo tehtävänannossa, jotta päästäisiin tämälkaltaisesta ongelmasta eroon. Tehtävässä 2 esimerkiksi on paljon tärkeämpää se, että puhuja osaa esitellä tärkeän vieraan riittävää kohteliaisuutta ja kielen muodollisuusastetta noudattaen kuin se, käyttääkö hän vahingossa ranskankielisessä tuotoksessaan esimerkiksi jonkin toisen kielen esitellä-verbiä muistuttavaa sanaa. Kyseessä on kuitenkin vain yhden sanan tuntemus, joka vaikuttaa hyvin paljon osallistujan puhe-suorituksesta saatavaan yleiskuvaan.

Kuunneltaviin puhenäytteisiin liittyvän osallistujan vastauslomakkeen avulla oli siis tarkoitus kerätä pääasiassa tietoa siitä, missä näytteissä oppijoilla oli vaikeuksia päättää, valitsevatko he *"kyllä"* vai *"ei"*-vaihtoehdon kunkin puhenäytteen kohdalla. Lisäksi lomakkeen *"vaikea päättää"* -kohdan avulla pyrittiin selvittämään niitä kielen piirteitä, joilla oppijat perustelivat vastausvalintaansa. Näiden tietojen lisäksi lomakkeista on toki mahdollista saada irti satunnaisesti muitakin tietoja kuin mitä edellä mainitaan kuten esimerkiksi lomakkeeseen merkittyä oppijan antamaa yleistä palautetta. Lisäksi lomakkeita läpi käydessä voi tehdä muitakin hyödyllisiä havaintoja oppijoiden vastauksiin liittyen.

Tällaisten ylimääräisten hyödyllisten havaintojen pohjalta voitiin havaita esimerkiksi sellainen seikka, että osa oppijoista tuntui pystyvän helposti perustelevaan, mitä olivat ne näytteissä esiintyvät kielen piirteet, joiden pohjalta he pitivät kuunneltua näytettä omaa tuotostaan huonompana. Sen sijaan kun kyse oli joistakin taitotasoasteikon yläpäähän sijoittuvista näytteistä, joita kyseiset oppijat pitivät parempina kuin omaa suoritustaan, he eivät jostain syystä viittaneet näytteessä mihinkään kielen piirteeseen, jonka perusteella heidän tekemänsä *"ei"*-valinta olisi ollut perusteltavissa. Näissä tapauksissa siis suurimmassa osassa näytteitä (=joita oppija piti huonompina kuin omaa suoritustaan) oli kaikissa viittaukset siihen kielen piirteeseen, jonka pohjalta oppija oli vastauksensa valinnut, kun taas niissä näytteissä, joita oppija piti omaa suoritustaan parempina, ei ollut yhdessäkään minkäänlaisia kommentteja vastausvalintaa valaisemassa. Voitaishiinko tästä kenties vetää ainakin sellainen johtopäätös, että on helpompi huomata suorituksessa olevat puutteet kuin sen ansiot? Kyseessä näyttäisi siten olevan samantyylinen ilmiö kuin mihin kiinnitettiin huomiota kirjallisten itsearviointiväittämien kohdalla: kommentteja saatiin itsearviointiväittämistäkin enemmän silloin, kun väittämät sisälsivät joitain vastaamista vaikeuttavia tekijöitä.



Lopuksi on syytä mainita, että osallistujan vastauslomake, johon tutkimukseen osallistujat siis merkitsivät valintansa ”kyllä”- ja ”ei”-vaihtoehtojen välillä näytteitä kuunnellessaan, oli hyvin tiivis lomake, jossa oli paljon tekstiä. Kuten tutkimuksessa on jo mainittu, oppijoilta kului tutkimuksen suorittamiseen aikaa yhteensä kaksi tuntia, ja siten kaikenlaisten kommenttien kirjoittaminen tutkimuksen pakollisten ”kyllä”- ja ”ei”-vaihtoehtojen valinnan lisäksi oli täysin oppijoiden hyvän tahdon ja aktiivisuuden varassa. Näistä syistä ei puhe-suoritusten kuunteluun liittyvään vastauslomakkeeseen enää lisätty sellaisia näytteiden kuuntelun tai niiden avulla tehdyn itsearviointin onnistumiseen liittyviä avokysymyksiä kuin mitä paperilla tehdyssä itsearviointilomakkeessa oli. Näin jälkepäin tutkimuksen osallistujien palautetta tarkastellessa tulee mieleen, että olisi ollut mielenkiintoista kerätä palautetta siitä, millaiseksi oppijat kokivat kuunneltujen näytteiden avulla tehdyn itsearviointin. Sellaista systemaattista palautetta tai tietoa ei näet yleensä saada tarkasteltavaksi, mitä ei tutkimuksissa erikseen kysytä.

Kuunneltujen näytteiden kohdalla on kuitenkin syytä tuoda esille osallistujien vähitellen tutkimuksen edistymisen kuluessa edennyt väsähtäminen tutkimuksen osallistumisajan loppupuolella. Keskittymisen herpaantuminen ja tietynlainen väsähtäminen näkyivät muutamissa papereissa vapaina kommentteina, joissa todettiin muun muassa seuraavaa:

*”Liian monta näytettä! Kerkeää unohtaa oman. Vertaa lopussa vain mielikuvaan omasta tasosta” (3).*

Tämänkaltainen kommentti onkin varsin ymmärrettävä kaksituntisen kielistudioistunnon loppupuolella tai sen jälkeen. Väsähtäminen näkyi joissakin vastauslomakkeissa siten, että tehtävästä 1 oli kommentoitu melkein jokaista näytettä, mutta siirryttäessä tehtävään 2 vain muutama ensimmäistä näytettä sisälsivät kommentteja ja lomakkeen lopussa ei siten ollut enää minkäänlaisia kommentteja tai perusteluja pakollisten ”kyllä”- ja ”ei”-valintojen lisäksi. Jos oppija kokee tehtävän liian väsyttäväksi tai esimerkiksi monotoniseksi, hän pidättäytyy tekemästä mitään ylimääräistä varsinaisen pakollisen osuuden lisäksi, ja pois jäävä tieto on tietenkin aina tutkimuksen kannalta harmillinen asia. Tutkimuksesta saatujen kommenttien ja omien havaintojen perusteella onkin selvää, että tämänkaltaisessa tutkimuksessa raja oppijan keskittymiskyvyn kannalta liian pitkän mittarin ja tilastollisten analyysien edellyttämän riittävän pitkän mittarin välillä on hiuksenhieno.

### 6.3 Tutkimuksessa kehitettyjen itsearviointivälineiden laadun ja hyödyllisyyden tarkastelua Bachmanin ja Palmerin mallin pohjalta

Tutkimuksen tavoite oli kehittää uudentyyppinen kielitaidon itsearviointiväline, jossa oppijat suorittavat annetun puhumistehtävän, kuuntelevat oman suorituksensa ja lopuksi vertaavat suoritustaan kuulemiinsa muiden oppijoiden suorituksiin samassa tehtävässä. Koska kyseisen kaltaista itsearviointia ei ole aikaisemmin kovinkaan paljon tutkittu, kyse on eräänlaisesta perustutkimuksesta, jossa pyritään selvittämään arviointitavan yleistä toimivuutta. Tästä johdun tavoitteena oli kehittää mahdollisimman laadukas arviointiväline, joka olisi samalla riittävän reliabeeli ja validi. Välineen tuli myös olla tarpeeksi käytännöllinen tutkimustarkoituksiin mutta ei välttämättä vielä muun tyyppiseen käyttöön esimerkiksi kielenopetuksessa. Jotta edellä kuvattuun päämäärään päästäisiin, tutkimuksessa verrattiin uutta arviointivälinettä tilastollisin keinoin sekä toisiin, perinteisempiin itsearviointivälineisiin että ulkopuoliseen arvioon oppijoiden kielitaidosta. Oppijoilta kerättiin myös laadullisempaa aineistoa, jotta arviointivälineiden toimivuudesta saataisiin rikkaampi kuva. Sellaiset näytteet ja väittämät, jotka tilastollisten analyysien perusteella näyttivät epäilyttäviltä tai joita oppijat olivat kommentoineet poikkeuksellisen runsaasti tai joiden avulla tapahtuvaa itsearviointia he pitivät hankalana, otettiin yksityiskohtaisempaan tarkasteluun, jotta voitiin selvittää, tulisiko ne säilyttää osana arviointivälineitä vai ei.

Kun tutkimuksen ja siinä kehitettyjen itsearviointivälineiden laatua tarkastellaan alla Bachmanin ja Palmerin kriteerien avulla, kannattaa siis pitää mielessä edellä esitetty kuvaus tutkimuksen ja sen tuloksena syntyneiden instrumenttien tarkoituksesta. Arviointivälineillä on omat vahvuutensa ja heikkoutensa, ja tuskin mikään väline täyttää täysin kaikkia laatuvaatimuksia. Oleellisinta onkin tarkastella, onko testi tai muu arviointiväline laadukas juuri niiden ominaisuuksien suhteen, jotka ovat tärkeimpiä nimenomaan sen käyttötarkoituksen kannalta.

Seuraavaksi siirrytään tarkastelemaan tutkimuksen ja siinä käytettyjen itsearviointivälineiden laatua kokonaisuutena. Tarkastelussa käytetään pohjana Bachmanin ja Palmerin kehittämää teoreettista mallia (*Test Usefulness*), joka on tarkoitettu testien laadun tarkasteluun (Bachman & Palmer 1996). Peruste sille, että tässä käsitellään arviointivälineiden lisäksi myös koko tutkimusprosessia, on se, että tutkimuksessa oli tässä tapauksessa samalla kyse arviointivälineiden laadinta- ja validointiprosessista. Tutkimusprosessin eri vaiheilla on siis suora vaikutus laadittujen välineiden laatuun, ja sen avulla voidaan selittää sekä välineiden vahvuuksia että niissä mahdollisesti esiintyviä ongelmia ja heikkouksia. Tutkimuksen ja arviointivälineiden analyysin lisäksi esitellään, mitä tutkimuksesta opittiin mahdollisesti myöhemmin tehtäviä samanlaisia tai jatkotutkimuksia varten.

Bachman ja Palmer näkevät testin laatua ja yleishyödyllisyyttä tarkastelevan mallin olevan keskeinen perusta testin laadunvalvonnalle testin koko elinkaaren ajan. Mallin avainkysymys on ”kuinka hyödyllinen tietty testi on siihen tarkoitukseen, johon se on laadittu?”. Testin laatu / hyödyllisyys on Bachmanin ja Palmerin mallissa nivottu suoraan testin tarkoitukseen, ja siksi ensin määritelläänkin tutkimuksessa kehitettyjen arviointivälineiden tarkoitus. Mallin lähtöajatus on se, että kaikki testin ominaisuudet vaikuttavat testin hyödyllisyyteen ja että mikään niistä ei välttämättä ole tärkeämpi kuin jokin toinen ominaisuus. Siten eri ominaisuuksien välisen tasapainon löytäminen kulloiseenkin tilanteeseen on oleellista (Bachman & Palmer 1996, 17–38.)

Bachmanin ja Palmerin malli sisältää kuusi erilaista toisiinsa linkittyvää testin ominaisuutta, jotka ovat reliaabelius, käsitevalidius, autenttisuus, vuorovaikutteisuus, (testin) vaikutus ja käytännöllisyys. Nämä ominaisuudet on listattu taulukkoon 30, jossa myös kuvataan lyhyesti omin sanoin, mitä kukin ominaisuus Bachmanin ja Palmerin mallissa tarkoittaa.

TAULUKKO 30 Bachmanin ja Palmerin (1996) testin luotettavuutta (*test usefulness*) tarkasteleva malli.

Testin ominaisuus	Pääidea	Mitä ominaisuus tarkoittaa Bachmanin ja Palmerin mallissa
<b>Reliaabelius</b>	Ensisijaisia mittaominaisuuksia testeissä	Mittaamisen konsistenssi (tulos kerrasta toiseen sama)
<b>Käsitevalidius</b>		Tulosten tulkinnan merkitsevyys: testin tulos indikoi sitä taitoa, jota halutaan mitata. Liittyy tulosten yleistettävyyteen testitilanteen ulkopuolelle
<b>Autenttisuus</b>	Aina suhteellisia; ei täydellistä autenttisuutta eikä	Testin tulos vastaa kielenkäyttöä muussa kuin testitilanteessa; testitehtävä kohdekielen kielenkäyttötilanne; yhteydessä testin suorittajan kuvaan testistä
<b>Vuorovaikutteisuus</b>	interaktiivisuutta	Testin suorittajan yksilöllisten kognitiivisten ja affektiivisten ominaisuuksien ja tehtävän piirteiden vuorovaikutus
<b>Testin/tutkimuksen vaikutus</b>	Mikro- ja makrotasot	Testin käytön vaikutus yksilöön tai yhteiskuntaan/koulutusjärjestelmiin
<b>Käytännöllisyys</b>	Resurssit	Käytössä olevat resurssit / Tarvittavat resurssit

### 6.3.1 Reliaabelius

Reliaabelius tai reliabiliteetti eli luotettavuus yhdessä käsitevalidiuden kanssa on mittaominaisuus, jonka pääasiallisena tarkoituksena on toimia riittävän evidenssin osoittajana sille, että testin pisteitä tai tuloksia voidaan käyttää johtopäätösten tai muiden päätösten tekemiseen. Reliabiliteetti liittyy siten mittauksen luotettavuuteen, sen tulosten tarkkuuteen ja tutkimuksen toistettavuuteen (Heikkilä 1998, 30). Luotettavuutta voidaan pitää koko tutkimuksen perustana. Käyn nyt läpi niitä tutkimuksen työvaiheita ja tuloksia, joiden pohjalle tämän tutkimuksen tulosten ja tutkimuksen aikana kehitettyjen itsearviointivälineiden luotettavuus rakentuu.

#### *Yki-korpuksen puhesuoritusten arviointi*

50 alkuperäissuorituksen arviointi, jonka tekivät viisi arvioijaa, osoittautui melko haastavaksi työksi. Arvioijien välillä oli huomattavan suuria eroja joidenkin arvioitavien Yki-suoritusten välillä. Jokainen arvioija arvioi 50 suoritusta / tehtävä, joten kaikkiaan jokaisella arvioijalla oli yhteensä 100 arvioitavaa suoritusta.

Tehtävän 2 arviointi oli tuotosten lyhyiden vuoksi hankalaa. Olisikin ollut varmasti paikallaan ohjeistaa arviointityötä jollakin tavalla samaan tapaan kuin myöhemmin tutkimuksessa tehtiin tutkimukseen osallistuneiden suorittajien puhesuoritusten kohdalla, jotta arvioijat olisivat osanneet kiinnittää huomionsa tiettyihin seikkoihin tai kielen piirteisiin arviointityönsä aikana. Kirjallisten ohjeiden puuttuessa kukin arvioija tulkitse luultavasti tehtävään 2 liittyviä arviointiasteikkoja omalla tavallaan. Arviointityön jälkeinen keskustelu arvioijien kanssa osoitti, että arvioijat olivat painottaneet eri piirteitä tehtävän 2 kohdalla: osa arvioijista korosti omien sanojensa mukaan tuotoksen ymmärrettävyyttä sen sosiolingvististen vaatimusten kustannuksella (ks. liite 8). Näissä tapauksissa puhesuoritusten taitotasoarviot olivat suhteellisen korkeita. Sen sijaan ne arvioijat, jotka tähdensivät sosiolingvististen taitojen hallinnan tärkeyttä, näyttivät arvioivan suorituksia hieman ankarammin ja antavan systemaattisesti alhaisempia taitotasoarvioita.

Tarkempi ohjeistus olisi todennäköisesti yhdenmukaistanut arviointeja kokonaisuutena, vaikka se ei tietenkään olisi täysin poistanut kaikkeen taitotasoarviointiin liittyvää henkilökohtaista kriteereiden tulkintaa. Onkin todettava, että jos tässä vaiheessa tutkimusta olisi ollut mahdollista käyttää Facets-analyysiä arvioijien toiminnan tarkistamiseksi, näyttöiden valintaprosessi olisi helpottunut ja valittujen näyttöiden laatu parantunut ainakin jossain määrin.

#### *Puhumisen itsearviointilomake*

Kuten tutkimuksen kuluessa on jo selvinnyt, DIALANG-hankkeessa laadittu puhumisen itsearviointilomake toimi tutkimuksessa hyvin. Erityisesti sen sisältämät 42 tilannekohtaista itsearviointiväittämää korreloivat hyvin myös oppijoiden muiden itsearviointien kanssa (ks. tarkemmin tutkimuskysymys 2b).

Lomakkeen luotettavuudesta voisi kertauksenomaisesti mainita, että monet tilastollisissa analyyseissä huonosti toimineista väittämistä olivat taitotasoasteikon ääripäihin sijoittuneita väittämiä, joiden huono toimivuus saattoi olla teknistä ja vähäisestä variaatiosta johtuvaa. Lisäksi huomionarvoista on se, että perinteinen osioanalyysi nosti helposti esille taitotasoasteikkojen ääripäihin sijoittuvat ongelmallisilta vaikuttavat osiot, kun taas Rasch-analyysin osoittamat toiminnaltaan ongelmalliset lomakkeen osiot eivät painottuneet asteikon ääripäihin.

Lomakkeen täytön aikana kielistudiossa oppijat esittivät säännönmukaisesti joitakin lomakkeen täyttämiseen liittyviä kysymyksiä. Kysymykset koskivat jo tutkimuksessa esiteltyä tilannetta, jossa oppijalla oli vaikeuksia vastata lomakkeessa oleviin sellaisiin väittämiin, joista heillä ei ollut kokemusta todellisessa kielenkäyttötilanteesta. Kielistudioistuntojen kuluessa kävi selväksi, että tilanteen tuntemattomuus häiritsi monia tutkimukseen osallistujia. Asia ohjeistettiin suullisesti erikseen jokaisen ryhmän kohdalla, kun kävi selväksi, että vastaaminen näissä tilanteissa tuotti selviä vaikeuksia osalle oppijoista. Vastaavallisissa tutkimuksissa olisikin ehkä paikallaan liittää lomakkeeseen kirjallinen ohjeistus siitä, miten suhtautua väittämään, jonka kuvaama tilanne ei ole oppijalle todellisesta elämästä tuttu. Tämä asia on syytä huomioida, sillä tutkimusten mukaan tämänkaltaisella kokemukseen liittyvällä seikalla (tai itse asiassa sen puutteella) voi olla suuri merkitys itsearviointissa ilmenevään variaatioon (Ross 1998, 13).

Toinen jo tutkimuksessa esille otettu seikka, joka liittyy siihen, että oppijoiden vastaukset ovat totuudenmukaisia, liittyy vaikeuksiin valita kahden vaihtoehdon ("kyllä" tai "ei") välillä. Henkilöt, joilla oli tämänkaltaisia vastausvaikeuksia, alkoivat useimmiten tulkita väittämän kuvaamaa kielenkäyttötilannetta liian monitahoisesti ottaen huomioon kaiken mahdollisen aineksen, mitä pystyivät mielessään liittämään kyseiseen väittämään. Usein tämä monipolvinen väittämän kuvaaman tilanteen tulkinta johti siihen, että lopullisen "kyllä"-tai "ei"-vastauksen antaminen tuntui olevan mahdotonta. Kun tämänkaltaisen väittämien "ylitulkinta" huomattiin, oppijoita ohjeistettiin suullisesti vastaamaan väittämien kuvaamaan tilanteeseen spontaanisti. Heitä kehoitettiin vastaamaan sen mielikuvan pohjalta, joka heille muodostui, kun he lukivat väittämän ensimmäistä kertaa.

Itsearviointilomakkeen 42 tilannekohtaista itsearviointiväittämää ja 6 Viitekehysten taitotason kirjallista kuvausta tasoilta A1-C2 oli käännetty DIA-LANG-hankkeen aikana englannin kielestä suomen kieleen. Tässä tutkimuksessa onkin otettava huomioon, että tehtyjen käännösten tasolla ja tarkkuudella voi olla jonkinasteista merkitystä tämän tutkimuksen tuloksiin.

*Kuunneltaviin puhenäytteisiin perustuva itsearviointi: tehtävät 1 ja 2*

Kuunneltavia näytteitä koskevan kielistudioistunnon lopussa esille tuli muutamana kerran yllättävä ongelma, kun jotkut tutkimukseen osallistuneet oppijat yllättäen huomasivatkin vastanneensa koko lomakkeen täytön ajan eri kysymykseen kuin mihin oli tarkoitus vastata. Tässä vaiheessa tarkastettiin yhdessä osallistujan kanssa, oliko kysymys, johon osallistuja oli vastannut, samanmuotoinen kuin mihin hänen olisi pitänyt vastata eli tarkoittivatko hänen valitsemansa ”kyllä”-vastaukset todellakin ”kyllä”-vastauksia vai oliko kaikki lomakkeen rasti käännettävä toisin päin, jotta saatiin esille ne vastaukset, joita osallistuja oli halunnut antaa. Tätä hankaluutta yritettiin alun perin estää sillä, että kysymys, johon oppijan tuli vastata verratessaan omaa suoritustaan nauhalta kuunneltaviin näytteisiin, oli kirjoitettu kaksisivuisen lomakkeen molemmille sivuille näkyviin. Lisäksi kielistudioistunnon aluksi muistutettiin vielä osallistujia siitä, että jos jostain syystä kuuntelun aikana kysymyksen muoto unohtuu, kysymyksen voi helposti palauttaa mieleen vilkaisemalla lomakkeen yläreunaan, mihin vastattava kysymys oli kirjoitettu tummennetulla tekstillä.

Kuunneltaviin puhenäytteisiin perustuvassa itsearviointilomakkeessa oli ”vaikea päättää” -sarake, jonka avulla oli tarkoitus kerätä varsinaisten ”kyllä-” tai ”ei”-vastausvaihtoehtojen lisäksi ylimääräistä tietoa oppijalta niistä näytteistä, joiden kohdalla vastausvalinnan tekeminen tuotti vaikeuksia oppijoille. ”Vaikea päättää” -rastin merkitsemistä koskevan ohjeistuksen yhteydessä korostettiin, että vaikka ”vaikea päättää” -kohtaan merkittäisiin rasti, tuli kyseisen näytteen kohdalla silti ehdottomasti valita jompikumpi lomakkeen pakollisista ”kyllä”- tai ”ei” -vaihtoehtoista. Ohjeistuksesta huolimatta osa osallistujista merkitsi kuitenkin vain ”vaikea päättää” -rastin tiettyjen näytteiden kohdalle jättäen tekemättä tilastollisten analyysien kannalta tärkeän ”kyllä” tai ”ei” -valinnan. Siten tilastollisista analyyseistä puuttuu joidenkin osallistujien kohdalla joitakin vastauksia. Tämänkaltaista ongelmaa voi olla mahdoton välttää täysin, mutta ehkä lomakkeeseen voisi ajatella ”pakollinen”-merkintää muistutukseksi niiden sarakkeiden kohdalle, jotka halutaan jokaisen osallistujan täyttävän.

Tutkimuksessa käytetyt mittarit olivat varsin pitkiä (yli 40 kirjallista väitettä tai yli 20 kuunneltavaa puhenäytettä / tehtävä), mikä osaltaan auttoi hyvän reliabiliteetin saavuttamisessa. Mittavälineiden pituutta olisi todennäköisesti mahdollista lyhentää ilman, että niiden reliabiliteetti kärsisi kovinkaan paljon. Mittareiden lyhentäminen olisi varmasti tarpeen, jos tämänkaltaista itsearviointia haluttaisiin käyttää käytännön kielenopetustilanteessa. Mittavälineiden laatuksiteereiden painoarvo on suoraan yhteydessä testin tarkoitukseen. Jos testiä on tarkoitus käyttää itsearviointitaidon kehittämiseen luokkatilanteessa, on selvää, että mittarin reliabiliteetin ei tarvitse olla samaa luokkaa kuin tämänkaltaisessa tutkimuksessa, jossa kokeiltiin uudenlaisen itsearviointimenetelmän toimivuutta.

### 6.3.2 Käsitemvalidius

Käsitemvalidius kuuluu Bachmanin ja Palmerin mukaan testitulosten pohjalta tehtyjen tulkintojen merkityksellisyyteen ja sopivuuteen. Heidän mukaansa jokaisen testijärjestelmän on pystyttävä puolustamaan niitä tulkintoja, joita tehdään sen tuottamien tulosten pohjalta. Puolustaminen liittyy läpinäkyvyyteen siinä mielessä, että on oltava konkreettista näyttöä tai evidenssiä siitä, että testin tulos heijastaa sitä osaamista, mitä kulloinkin halutaan testata. Bachman ja Palmer muistuttavat lisäksi, että testin validointi on jatkuva prosessi ja että testitulosten ei voida milloinkaan väittää olevan täysin valideja (Bachman & Palmer 1996, 21 – 23.)

#### *Tutkimuksessa käytettyjen näytteiden tasoista ja näytteiden valinnasta*

Suurin osa studiossa kuunneltavista näytteistä oli Yleisten kielitutkintojen perus- tai keskitason näytteitä. Näytteistä vain muutama erikseen nauhoitettu edusti ylimmän tason osaamista. Tutkimuksen toteuttamisen jälkeen on selvää, että kuunneltavien näytteiden joukossa olisi ollut hyvä olla enemmän ylimmän tason puhenäytteitä, joista osa olisi varsin hyvin voinut olla syntyperäisten tuotoksia. Tutkimukseen osallistujien joukosta ne osallistujat, joiden suullinen kielitaito sijoittui ylimmille tasoille, olisivat mitä luultavimmin kokeneet näytteiden kuuntelun ja oman suorituksensa niihin vertaamisen motivoivammaksi, mikäli näytteissä olisi ollut enemmän haastetta myös heille.

Tutkimuksessa käytettyjen näytteiden valinnasta (26 kpl / tehtävä 1 ja 23 kpl / tehtävä 2) on paikallaan todeta sen verran, että miesten osuus Ykin ranskan testin suorittajissa on hyvin vähäinen kaikilla kolmella tutkintotasolla. Sen vuoksi Yki-korpuksen 50 näytteestä vain murto-osa oli miesten puhetuotoksia. Kun koottiin yhteen oppijoiden kuunneltavaksi tarkoitettuja näytteitä, yritettiin äänitallenteelle nauhoittaa vaihtelevuuden vuoksi vuorotellen mies- ja naisäänä. Miesäänien määrä olikin niin vähäinen, että ne harvat miesten tuottamat puhetuotokset, joita Yki-korpuksessa oli, kelpuutettiin kuunneltavien näytteiden joukkoon melkein automaattisesti. Onkin mahdollista, että miesten puhetuotosten vaatimustason kanssa tehtiin näytteiden valintavaiheessa enemmän kompromisseja kuin naisten puhetuotosten kanssa. Kuunneltavien näytteiden joukossa oli jo edellä mainittu tehtävän 1 näyte 15, jonka taitotasosta arvioijat olivat hyvin erimielisiä. Näytteiden valintavaiheessa olisi ehkä ollut hyväksi, jos olisi voitu nauhoittaa lisää ainakin muutama miesten tuottama puhenäyte / tehtävä. Täten olisi voitu välttää ottamasta tutkimuksen kuunneltaviin näytteisiin mukaan esimerkiksi näytteen 15 kaltainen näyte, jonka taitotason arviointi selvästikään ei ollut ongelmatonta.

*Puhumisen itsearviointilomakkeen väittämistä*

Lomakkeen 42:sta väittämästä väittäjä numero 29 nousi analyyseissä esille väittämänä, jonka korrelaatiokerroin oli tutkimuksen ainoana negatiivinen ja jota siten voidaan pitää osiona, joka ei toiminut lainkaan mittarin osiona:

Väittäjä 29 (taso A2; korrelaatiokerroin -.5071)

Selviän lyhyistä sosiaalisista tilanteista, mutta pystyn harvoin pitämään keskustelua yllä.

Väittämän 29 ominaisuuksia ja analyyseissä saatuja arvoja on analysoitu tarkemmin kvantitatiivisten analyysien yhteydessä aiemmin työssä. Siksi tässä kohdassa tyydynkin vain toteamaan, että kaksiosaisen väittämän pilkun molemmilla puolilla olevat osat ovat ristiriitaisia loogisesti tarkastellessa. Oppijoiden vastaukset kuvastavat sitä, kuinka hankalaa väittämään on ollut vastata: pidemmälle kieliopinnoissaan ehtineet oppijat ovat vastanneet tähän A2-tasoiseen väittämään "ei", kun taas kieliopintojensa alkutaipaleella olevat oppijat ovat vastanneet väittämään "kyllä". Toisin sanoen vastaukset ovat perustuneet täysin väärille tulkinnoille. Osaselityksenä oppijoiden ristiriitaisilta tuntuviin vastausvalintoihin voi olla se, että eri tasoilla olevat oppijat osallistuvat tyypillisesti vaativuudeltaan jonkin verran erilaisiin sosiaalisiin tilanteisiin. Sitä eri kielitaidon tasoilla olevat oppijat miettivät tiettyjä omalle taitotasolleen tyypillisiä sosiaalisia tilanteita ja valitsevat vastauksensa sen mukaan. Tämän väittämän kohdalla nouseekin aiheelliseksi käsitevaliuden-käsite, ts. mitä väittämän sisältö itse asiassa mittaa – kielitaitoa vai esimerkiksi sosiokulttuurista osaamista yleensäkin, kuten eräs tutkimukseen osallistunut oppija kommentoi.

Kirjalliset itsearviointiväittämät saattavat olla hankalia käyttää ja tulkita siitä johtuen, että ne ovat kontekstistaan erotettuja. Voidaankin miettiä, olisiko tämänkaltaisia lomakkeita käytettäessä parempi määrittää, mitä jollakin käsitteellä tarkasti ottaen tarkoitetaan (esim. väittäjä 39: Osaan ilmaista tunnetiloja kuten *hämmästyminen*, *onnellisuus*, *suru*, *mielenkiinto* ja *välinpitämättömyys* ja reagoida niihin). Tämänkaltaista käsitteiden määrittelyä on tehnyt mm. Myles (1997) antaessaan tarkat määritelmät käyttämänsä itsearviointilomakkeen käsitteistä. Tässä tutkimuksessa käytetty itsearviointilomake oli tosin melko pitkä ilman määritelmiäkin, mutta yleisesti ottaen käsitteiden määrittäminen voisi kyllä olla hyödyllistä väittämien tulkinnan kannalta.

Kirjallisia itsearviointilomakkeita laadittaessa tulisi mahdollisuuksien mukaan ottaa huomioon käytettävyyteen ja käytännöllisyyteen liittyvien seikkojen (kuten lomakkeen pituus) lisäksi myös kielitaidon eri komponenttien esiintyminen lomakkeiden kuvaimissa tai väittämässä. Jotta oppijan kielitaidosta voitaisiin saada mahdollisimman laaja kuva, suullista kielitaitoa koskevien itsearviointiväittämien tulisi sisältää ainakin kommunikatiivisen kompetenssin elementtejä eli kielellisten komponenttien lisäksi myös sosiolingvistisiä ja pragmaattisia komponentteja. Näiden eri komponenttien määrällinen jakautuminen riippuu paljolti siitä, mitä lomakkeella halutaan selvittää tai millaisiin



tehtäviin tai tilanteisiin suoritettava itsearviointi mahdollisesti liittyy. Tietynlainen kielitaidon eri osakomponenttien tasapainottaminen lomakkeissa käytävissä itsearviointiväittämässä takaisi varmasti monipuolisemman ja tarkemman kuvan oppijan kielitaidosta. Tässä tutkimuksessa käytetty itsearviointilomake oli kuitenkin lainattu suoraan DIALANG-hankkeesta, ja tämän tutkimuksen välillisenä tarkoituksena haluttiin selvittää, kuinka toimiva on DIALANG-hankkeessa kehitetty suullisen kielitaidon itsearviointilomake. Tästä syystä tutkimuksessa käytettiin edellä mainitussa hankkeessa kehitettyä itsearviointilomaketta muuttamatta lainkaan sen sisältöä tai rakennetta.

Oppijoiden antamien vastausten yleisen luotettavuuden kannalta on todettava, että joidenkin oppijoiden kommentteista kävi aineiston analyysin aikana ilmi, että he eivät olleet sisäistäneet sitä, että jos oppijan kielitaito sijoittui esimerkiksi taitotasolle 4, hänen voitiin odottaa pääosin hallitsevan taitotasolle 1, 2 ja 3 sijoittuvat kielenkäyttötilanteet. Taitotasoarviointiin liittyvä yleinen kielitaidon progressio olisi syytä selostaa hyvinkin tarkkaan oppijoille mahdollisissa tulevaisissa kielitaitoon tai kielitaidon itsearviointiin liittyvissä tutkimuksissa.

Yleisesti ottaen voidaan todeta, että on haastava tehtävä laatia sellaisia kirjallisia itsearviointilomakkeita, jotka olisivat riittävän yleisellä tasolla mutta kuitenkin tarpeeksi yksityiskohtaisia ja yksitulkintaisia, jotta niitä voitaisiin käyttää monissa eri itsearviointitilanteissa. Paperilla esitetyt itsearviointiväittämät näyttävät joissakin oppijoissa saavan aikaan monenlaista persoonakohtaista ylitulkintaa ja väittämässä esiintyvien yksittäisten sanojen tulkitsemiseen liittyvää takertumista. Tutkimuksesta saadun kokemuksen pohjalta tuntuukin tärkeältä, että itsearviointilomakkeita täyttävillä henkilöillä on välitön mahdollisuus kysyä neuvoa esimerkiksi jonkin käsitteen tulkinnasta läsnä olevalta lomakkeen sisällön hyvin tuntevalta henkilöltä.

#### *Tulosten yleistettävyys*

Tämä tutkimus on kohdistunut ranskan kieleen siten, että ranska on ollut vieraana kielenä, ja tutkimuksen kohderyhmä ovat olleet aikuiset kielen oppijat. Tulosten yleistettävyyttä muihin kieliin pohdittaessa onkin syytä muistaa, että osa tutkimuksen tuloksista voi olla hyvinkin kielikohtaisia. On mahdollista, että vähemmän opiskeltujen kielten oppijoilla on taipumusta arvioida oma kielitaitonsa paremmaksi kuin esimerkiksi englannin kieltä opiskelleet arvioivat oman kielitaitonsa. Lisäksi ranskan kielellä on Suomessa hyvin erilainen asema kuin esimerkiksi juuri englannilla, jota lähestulkoon kaikki osaavat ainakin jossain määrin. Ranska on harvinaisempi kieli Suomessa sekä opiskelijamääriltään että todellisilta taitajamääriltään.

Tutkimukseen osallistuneiden sukupuolijakauma on myös tässä tutkimuksessa epätasainen, koska valtaosa osallistuneista oppijoista on naisia. Toisaalta kyseinen jakauma edustaa hyvin kielen oppijoiden enemmistöä suomalaisissa korkeakouluissa. On vaikea ottaa kantaa siihen, kuinka yleistettäviä tämän tutkimuksen tulokset ovat suhteessa eri kieliin tai kohderyhmiin. Asiaa tulisi tutkia erikseen, jos tästä aiheesta haluttaisiin lisävarmennusta.

### 6.3.3 Autenttisuus

Bachmanin ja Palmerin (1996, 23–24) mukaan autenttisuus määrittää testitehtävän ja kohdekielen todellisen käytön välistä yhteyttä. Kyseessä on heidän mukaansa tärkeä testin laatuun liittyvä ominaisuus, koska sillä on yhteytensä siihen, kuinka testituloksia voidaan yleistää testin ulkopuolelle (mikä taas viittaa edellä käsiteltyyn käsitevaliditeettiin), ja siihen, miten relevanttina testattava tehtävän kokee. Jos testattava kokee, että tehtävä on relevantti, tällä on vaikutusta hänen testisuoritukseensa, mikä tekeekin autenttisuuden painoarvosta tuloksiin nähden merkittävän. Tässä tutkimuksessa ei ole kerätty tutkimukseen osallistuneilta oppijoilta kommentteja käytetyistä tehtävistä, joten kaikki alla olevat tehtäviä koskevat pohdinnat ovat tutkimuksen tekijän omia. Pohdin siis alla lyhyesti tutkimuksessa käytettyjen kahden tehtävän ominaisuuksia Bachmanin ja Palmerin mallin autenttisuuden näkökulmasta.

#### *Tehtävä 1*

Tehtävä 1, jossa oppijan tuli kuvailla nuorta istuvaa miestä, osoittautui tutkimuksen kuluessa onnistuneeksi valinnaksi. Sen avulla saatiin tehtyä se, mitä alun perin tutkimusta suunniteltaessa haluttiinkin saada aikaiseksi, eli kerättyä pidempi yhtäjaksoinen puhenäyte kielenoppijoilta. Tehtävän 1 toimiminen tuli esille myös tilastollisissa analyyseissä, sillä sen toimivuuteen mittainstrumenttina liittyvät monet edellä esitellyt myönteiset tulokset. Tehtävän 1 kielteisiä puolia ovat sen staattisuus ja tietynlainen monotonisuus, joka ei välttämättä innosta kaikkia tutkimukseen osallistuneita oppijoita käyttämään mielikuvitustaan kovin luovasti heidän tuottaessaan puhetta ja siten keskittymään siihen, että he tekisivät kaikkensa antaakseen mahdollisimman hyvän kuvan omasta kielitaidostaan. Kokonaisuutena tehtävä 1 oli kuitenkin hyvä valinta varsinkin, kun valtaosa tutkimukseen osallistuneista oppijoista edusti B1-tasoisia kielitaitoa. Tehtävässä 1 pyydetty henkilön kuvaaminen oli niin tavanomainen tehtävä, että jokainen tutkimukseen osallistuja pystyi kertomaan kuvasta sen verran, että suullista tuotosta saatiin arvioinnin näkökulmasta riittävästi. Jos kuvasta kertomistehtävän tilalle olisi valittu vaativampi kertomistehtävä, kuten esimerkiksi sarjakuvan selostaminen, olisivat oppijoiden tuotokset ehkä jääneet lyhyemmiksi, kun otetaan huomioon, että enemmistön kielitaito edusti kuitenkin kynnystaitotasoa. Vaativamman tehtävän valinta olisi varmasti voinut motivoida ylimmän tason oppijoita enemmän, mutta samalla tämänkaltainen valinta olisi suosinut vain hyvin pientä vähemmistöä tutkimukseen osallistuneista oppijoista valtaosan oppijoista tuottaessa vähemmän puhetta tehtävän vaikeutumisen vuoksi.

Tehtävä 1 osoittautui siis toimivaksi tämän tutkimuksen tarpeisiin. On kuitenkin syytä mainita, että tehtävä 1:n kaltainen kuvasta kertomistehtävä ei juuri tässä muodossaan kovin hyvin vastaa todellisen elämän kielenkäyttötilanteita. Tehtävän 1 avulla pystyttiin kuitenkin tekemään se, mitä sen avulla haluttiinkin tehdä eli keräämään tutkimuksessa tarvittavia pidempiä puhe-suorituksia.

sia. Tästä huolimatta on vaikea kovin tarkasti määritellä kyseisen tehtävän relevanssia aikuisopiskelijalle.

Tässä vaiheessa on syytä tuoda esille myös, että Bachmanin ja Palmerin (1996) mallin mukaan käsitellyt asiat ovat usein yhteydessä myös johonkin toiseen mallin käsitteeseen kuin ainoastaan siihen käsitteeseen, jonka otsikon alla ne esitellään. Siten esimerkiksi tässä kohdassa esitellyllä tehtävällä 1 avulla saatavalla tuotoksella on suora yhteys myös edelliseen käsitevalidiuskohtaan.

### *Tehtävä 2*

Tehtävään 2 liittyvä puhetuotos osoittautui niin lyhyeksi, että suoritusten arvioijilla oli vaikeuksia arvioida näytteitä ja osallistujien suorituksia. Mitä todennäköisimmin oppijat kokivat ainakin osittain samankaltaisia vaikeuksia verratessaan omaa tuotostaan kovin niukkoihin puhenäytteisiin. Lisäksi tutkimukseen osallistuneilla lukiolaisilla oli vähäisen elämäkokemuksensa vuoksi suuria vaikeuksia täyttää tehtävän edellyttämiä kohteliaisuus- ja muodollisuusvaatimuksia. Tehtävä 2 valittiin siksi, että tutkimuksessa olisi kaksi erityyppistä ja eripituista tehtävää. Tehtävän 2 piti edustaa ammattimaiseen kontekstiin liittyvää vuorovaikutustilannetta ja antaa siten oppijoille mahdollisuus tuottaa täysin erilaista kieltä kuin kuvasta kertomistehtävässä 1. Tehtävä 2 osoittautui kuitenkin heikommin toimivaksi kuin tehtävä 1, sillä sitä koskevat tilastollisten analyysien tunnusluvut olivat lähestulkoon poikkeuksetta huomontia verrattuna tehtävään 1 (tai paperilla tehtyyn itsearviointiin). Toisaalta tehtävän 2 avulla saatiin kerättyä oppijoilta erityyppistä puhetuotosta kuin tehtävässä 1. Siten tehtävä 2 toimi hyvin tehtävää 1 täydentävänä tehtävänä. Vaikka tehtävän 2 ominaisuudet ja sen yleinen toimivuus eivät yltäneet tehtävän 1 tasolle, tehtävä 2 ei kuitenkaan osoittautunut niin ongelmalliseksi, että se olisi esimerkiksi pitänyt jättää kokonaan huomiotta tutkimuksessa.

Tutkimuksessa käytettävistä tehtävistä voidaan lopuksi todeta, että tutkimuksen tulokset näyttävät olevan jonkin verran kontekstisidonnaisia: tutkimuksessa harjoitettu kuunneltaviin puhenäytteisiin perustuva itsearviointi toimi hyvin tehtävän 1 kaltaisessa kuvasta kertomistehtävässä, kun taas professionaalisessa kohteliaisuuden ja muodollisuuden hallintaa edellyttävässä esitelytehtävässä 2 sama itsearviointi ei toiminut yhtä hyvin. Kun tehtävää 2 tarkastellaan sen sijaan autenttisuuden näkökulmasta, voidaan sen todeta olevan hyvinkin autenttinen erityisesti tutkimukseen osallistuneiden aikuisopiskelijoiden näkökulmasta.

#### **6.3.4 Vuorovaikutteisuus**

Bachmanin ja Palmerin mallissa (1996) interaktiivisuus liittyy siihen, kuinka testisuorittajan henkilökohtaiset ominaisuudet saadaan otettua käyttöön testitilanteessa siten, että suorittajalta saadaan arvioitavaksi mahdollisimman kattava näyte hänen kielitaidostaan. Mallin mukaan tärkeimmät henkilön ominaisuudet kielitaidon arvioinnin kannalta ovat kielitaito (kielitieto, strateginen kompetenssi ja metakognitiiviset strategiat), alakohtainen tieto ja affektiiviset tekijät.

*Tutkimukseen osallistuneet oppijat*

Aluksi on syytä todeta, että tutkimusasetelman avulla ei ole mahdollista saada suoraa tietoa siitä, missä määrin oppijan kognitiiviset ominaisuudet / kognitiivinen prosessointi itsearviointinissa vastaa prosessointia todellisessa kielenkäyttötilanteessa. On kuitenkin todennäköistä, että näytteisiin liittyvässä itsearviointinissa, jossa oppija on itse tehnyt puhetehtävät eikä vain lukenut paperilta tietäntyyppiseen tilanteeseen liittyviä itsearviointiväittämiä, uusi itsarviointiväline vastannee paremmin todellista kielenkäyttöä vuorovaikutteisyyden suhteen. Voidaan siis olettaa, että kun oppija käyttää kieltä konkreettisesti, tilanteessa on mukana paljon samanlaista kielen kognitiivista prosessointia kuin vastaavassa oikeassa puhetilanteessa. Oppijoiden affektiivisista piirteistä kuten esimerkiksi motivaatiosta kerron jäljempänä lyhyesti.

Tässä tutkimuksessa oli kahdentyyppisiä oppijoita. Enemmistö oli yliopisto-opiskelijoita, jotka osallistuivat pääosin tutkimukseen omasta vapaasta tahdostaan ja vähemmistö oli lukiolaisia, jotka osallistuivat tutkimukseen opettajansa päätöksestä. Yleisenä kommenttina voidaan todeta, että osallistujat olivat hyvin sitoutuneita tutkimuksessa vaadittavien tehtävien suorittamiseen.

Ottaen huomioon, että tutkimukseen osallistui kahden lukioluokan oppilaita, on syytä korostaa opettajan asiaan sitoutumisen ja asennoitumisen tärkeyttä. Kahden lukioluokan oppilaiden asenne tutkimukseen osallistumista kohtaan oli näet täysin päinvastainen ja riippui siitä, oliko luokan ranskan opettaja paikalla tutkimukseen osallistuttaessa vai ei. Oppilaat, joiden opettaja oli henkilökohtaisesti paikalla kielistudiossa, osoittivat erittäin suurta keskittymistä ja motivaatiota tehtävien suorittamiseen, kun taas kielistudiossa ilman opettajaansa olleet oppilaat eivät kaikilta osin onnistuneet ottamaan tutkimukseen osallistumista vakavasti. Riittävällä vakavuudella asiaan suhtautuminen on tärkeää empiirisen tutkimuksen aineiston keräämisen kannalta, koska tutkimukseen tarvitaan tietty määrä osallistujien suorituksia. Olisikin ensiarvoisen tärkeää, ettei yhtään suoritusta jouduttaisi hylkäämään tutkimusaineistosta sen takia, että tehtäviä ei ole suoritettu asianmukaisesti tai että niistä on suoritettu vain osa. Tässäkin tutkimuksessa, jossa aineiston keruu yleisesti ottaen onnistui hyvin, varsinaisia hylättyjä lukiolaisten suorituksia oli 5 kappaletta, ja tämän lisäksi vaillinaisesti täytettyjä lomakkeita oli muutamia. Tähän tutkimukseen liittyvän kokemuksen perusteella voidaan siten suositella, että mikäli osallistujien joukossa on nuoria henkilöitä, heidän opettajansa motivoiminen asialle on enemmän kuin tarpeellista.

### 6.3.5 Vaikuttavuus

Tutkimuksen vaikuttavuudella tarkoitetaan sitä, miten tietyn testin käyttö vaikuttaa yksilöön (mikrotaso) tai laajemmin yhteiskuntaan ja sen koulutusjärjestelmään (makrotaso; Bachman & Palmer 1996, 29–30). Tämänkaltaisen tutkimuksen vaikutukset ovat tietenkin täysin eri tasolla kuin esimerkiksi Suomen valtion virallisen kielitutkintojärjestelmän Yleisten kielitutkintojen testien vai-

kutukset erityisesti silloin, kun järjestelmän todistuksia käytetään Suomen kansalaisuuden hakemiseen edellytettävän suomen tai ruotsin kielen taidon osoittamiseen. Ylioppilastutkinto voidaan mainita toisena valtakunnallisena tutkintojärjestelmänä, jolla on vaikutusta sekä yksilöiden elämään että yhteiskuntaan ja koulutusjärjestelmään yleisemminkin.

Tutkimuksen vaikuttavuuden osalta voidaan pohtia sitä, mitä vaikutuksia tutkimukseen osallistuneille oppijoille on siitä, että he saivat tutkimukseen osallistumisensa johdosta kirjallista palautetta omasta itsearviointitaidostaan kahden suorittamansa tehtävän pohjalta. Voidaan ajatella, että jos oppijan vastaanottama palaute vastasi oppijan käsitystä hänen omasta taitotasostaan, asialla saattoi olla myönteisiä vaikutuksia sekä oppijan suhtautumiseen ranskan kielen suulliseen taitoon että oppijan motivaatioon kielen opiskelua ja itsearviointia kohtaan. Jos taas oppija koki, että ulkoisten arvioijien antama arvio ei lainkaan vastannut hänen omaa käsitystään kielitaidosta ja itsearvioinnista tai jos annettu palaute ei jostain syystä ollut riittävän selvää, voidaan vain arvailla, millaisia vaikutuksia tällaisella palautteella voi olla oppijalle. Luultavasti tällaisessa tapauksessa oppijoiden reaktiot vaihtelisivat suuresti heidän persoonansa mukaan: joku saattaa ottaa ei-odotuksenmukaisen palautteen vastaan masentuen siitä ja luovuttaen, kun taas joku toinen saattaa päinvastoin sisuuntua tekemään asian eteen enemmän töitä. Olipa oppijan reaktio ja siitä seuraava toiminta mikä hyvänsä, annetulla palautteella on merkitystä jo senkin vuoksi, että tutkimus on oppijan silmissä luotettavampi silloin, kun toimitaan niin kuin on sovitettu (= palautetta luvattu tutkimuksen rekrytointivaiheessa). Kyse on siis myös tutkimuksen uskottavuuteen liittyvästä mielikuvasta oppijoiden silmissä.

Joka tapauksessa oppijan äänen kuuluville saaminen on tärkeää oppijoiden motivaation kasvattamisen kannalta. Varsinkin opetustehtävissä on keskeistä huomata, että oppijoiden näkemysten kerääminen testistä ja käytetyistä tehtävistä vievät aktiivista oppijuutta eteenpäin ja edistävät siten myös itsearvioinnin asiaa. Kun oppijalle jää myönteinen kuva siitä, että häntä kuunnellaan, tämä vaikuttaa hänen suorituksensa paranemiseen (Bachman & Palmer 1996, 32).

Tässä tutkimuksessa kerättiin oppijoiden näkemyksiä ja kokemuksia sekä itsearvioinnin yleisestä sujumisesta, yleistaitotason valinnasta, tilannekohtaisista itsearviointiväittämistä että kuunnelluista puhenäytteistä kahdessa tehtävässä. Oppijan itsearvioinnin luotettavuutta koskevien uusien myönteisten tutkimustulosten julkistamisella voi olla vaikutusta esimerkiksi kielen opetuksen käytänteisiin keskipitkällä aikavälillä. Käytännön opetustyön kannalta merkitystä saattaa olla myös sillä, että uusia itsearviointimenetelmiä kokeillaan ja että ne osoittautuvat toimiviksi, kuten tässä tutkimuksessa ilmeni. Itsearviointiin kuuluva oman oppimisen suunnittelu ja arviointi, itseohjautuvuus ja aktiivinen osallistuminen toteutuisivat kaikki mainiosti opetustilanteessa siten, että oppijat otettaisiin yhä enenevässä määrin mukaan testin suunnittelu- ja kehittämisvaiheeseen (Bachman & Palmer 1996, 32).

### 6.3.6 Käytännöllisyys

Bachmanin ja Palmerin mallissa (1996) testin käytännöllisyys on yksinkertaisesti sitä, että testin (tai tutkimuksen) toteuttamiseen tarvittavien henkilö-, materiaali- ja aikaresurssien tulisi olla tasapainossa. Jos kyse on esimerkiksi uuden testin käyttöönotosta, testin käytännöllisyyttä tarkastellaan kaikkien kolmen edellä mainitun resurssiryhmän näkökulmasta. Tarkastelen alla lyhyesti tutkimuksen resurssointiin liittyviä seikkoja sekä sitä, kuinka käytännöllisiä tutkimuksessa käytetyt itsearviointivälineet olivat.

#### *Aineiston keruu*

Kuten jo edempänä työssä on käynyt ilmi, tutkimuksen aineiston keruu oli melko hidas prosessi siitä johtuen, että kovinkaan moni opiskelija Suomessa ei tunne itseään "valmiiksi" ilmaisemaan itseään ranskan kielellä suullisesti. Monille osallistujille oli tärkeää saada tietää, että tutkimukseen haluttiin rekrytoida kaikentasoisia opiskelijoita, eikä esimerkiksi pelkästään jo hyvin ranskaa osavia. Monet oppijat halusivat myös varmistaa ennen osallistumishalukkuutensa ilmaisua, että he eivät joutuisi puhumaan ranskan kieltä kenenkään toisen opiskelijan kuunnellessa. Edelleen osa opiskelijoista halusi tietää ennen tutkimukseen osallistumista, ketkä toimisivat heidän suorituksensa arvioijina ja millä tavalla arviointi suurin piirtein tehtäisiin. Tämänkaltaisen aineiston kerääminen vaati siten aikaa, vaivaa, suostuttelua ja rohkaisua, jotta oppijat saatiin ensinnäkin kiinnostumaan tutkimukseen osallistumisesta ja vieläpä tulemaan paikalle kielistudioon usein omalla ajallaan illansuussa. Jälkikäteen tilannetta analysoiden tutkimusaineiston keruu sujui erinomaisesti. Tutkimuksen aineiston keruuta ja siihen liittyviä yksityiskohtia eritellään taulukossa 31.

TAULUKKO 31 Aineiston keruun yksityiskohtien erittelyä.<sup>24</sup>

	<b>Kommentit</b>
Tutkimuksen toteuttajat/ Henkilöresurssit	<ul style="list-style-type: none"> <li>- vastuhenkilö kielistudioistunnoissa (=tutkimuksen tekijä)</li> <li>- työn ohjaajat ja tilastollisten analyysien toteuttamista koskevat neuvonantajat</li> <li>- arvioijat</li> <li>- tutkimukseen osallistujat</li> <li>- muita henkilöitä : ranskan laitosten ja kielikeskusten johtajat ja sihteeri, ranskan kielen opettajat, ranskan kielen ainejärjestöjen yhteyshenkilöt, kielistudioiden tekniset vastuuhenkilöt, toimivuuskokeiluun osallistuneet opiskelijat, ylintä tasoa edustaneet opiskelijat (ylimmän tason lisänäytteiden äänittäminen)</li> </ul>
Tutkimuksen käytännön toteuttajan vaikutus	- suuri : tutkimusaineistoa keräävän tutkimuksen tekijän ohjeistus toteuttamisprosessista kielistudioistunnon alussa
Materiaaliresurssit (materiaalit, tilat, varusteet, rahoitus)	<p>1. Materiaalit:</p> <ul style="list-style-type: none"> <li>- Yki-suoritukset</li> <li>- lomakkeet: itsearviointilomakkeet, arvioinnissa käytetyt lomakkeet, palautelomakkeet (arvioijat ja oppijat), arviointiasteikot, arviointiohjeistus</li> <li>- maamerkkikasetit</li> <li>- kielistudiossa kuunneltavaksi tarkoitetut kasetit</li> </ul> <p>2. Tilat:</p> <ul style="list-style-type: none"> <li>- kielistudiot: kasettien koostaminen, tutkimusaineiston keruu</li> </ul> <p>3. Varusteet</p> <ul style="list-style-type: none"> <li>- kielistudion laitteisto</li> </ul> <p>4. Rahoitus:</p> <ul style="list-style-type: none"> <li>- matka-avustukset aineistonkeruuta varten</li> <li>- apurahat väitöskirjatyöskentelyä varten</li> </ul>
Aikaresurssit	<ul style="list-style-type: none"> <li>- tehtävien valinta Yki-suoritusten perus- ja keskitason suoritusten joukosta hidasta vähäisen suorittajamäärän vuoksi</li> <li>- osallistujien rekrytointi hidasta ranskan kielessä</li> <li>- aineiston keruu kielistudiossa (noin 2h / kerta)</li> <li>- aineiston arviointi aikaa vievää näytteiden ja oppijoiden suoritusten määrän vuoksi</li> </ul>
Vastauslomakkeiden käyttö	<ul style="list-style-type: none"> <li>- strukturoitu samanaikainen ryhmäkäyttö</li> <li>- samanlainen lomakkeiden esittely kaikkien osallistujaryhmien kohdalla</li> <li>- lomakkeiden lukumäärä: kolmannen täytettävän lomakkeen täytön yhteydessä havaittavissa väsymystä</li> </ul>

<sup>24</sup> Mukaelma Heikkilän teoksessa Tilastollinen tutkimus olevasta taulukosta (Heikkilä 2001, 20)

Saatujen vastausten määrä	<ul style="list-style-type: none"> <li>- suuri</li> <li>- hävikki:</li> <li>- 1 osallistuja, joka ei palauttanut itsearviointilomakkeita ;</li> <li>- 5 kpl lukiolaisia, joiden suoritukset oli hylättävä</li> <li>- muutama osallistuja, jotka tulkinneet väärin "vaikea päättää" -sarakkeen yhteyden lomakkeen täyttämiseen (= "vaikea päättää" -rastin yhteydessä eivät olleet tehneet valintaa "kyllä"- ja "ei"-vaihtoehtojen välillä näytteitä kuunnellessaan)</li> </ul>
Vastausten saannin nopeus	<ul style="list-style-type: none"> <li>- nopea</li> <li>- valmiit jokaisen kielistudioistunnon lopussa</li> </ul>
Vastausten tarkkuus	<ul style="list-style-type: none"> <li>- hyvä ja systemaattinen</li> <li>- väsymisen ja kyllästymisen vaikutus lomakkeiden täyttämiseen / täyttämättä jättämiseen ja vastauksiin?</li> </ul>
Väärinkäsitysten mahdollisuus	<ul style="list-style-type: none"> <li>- melko pieni, koska kielistudioistuntoa johti tutkimuksen tekijä, joka oli paikalla kielistudiossa jokaisessa studioistunnossa</li> <li>- joitakin väärin tulkittuja ohjeita</li> </ul>

Tutkimuksessa käytettyjen itsearviointivälineiden voidaan todeta olleen käytännöllisiä tutkimustarkoitukseen, sillä oppijat suorittivat tehtävät niille varatussa ajassa.

## 6.4 Tutkimuksen rajoituksia

Tutkimusprosessin edistyessä havaitaan tavallisesti sellaisia asioita, jotka eivät välttämättä toimineet juuri niin, kuin alun perin oli suunniteltu. Luvussa 6.4 käsitellään tutkimukseen liittyviä rajoituksia, joiden olemassaolo on syytä ottaa huomioon esimerkiksi tutkimuksen tulosten yleistettävyyttä tarkasteltaessa. Näiden seikkojen huomioon ottaminen on erityisen tärkeää myös, jos joku toinen haluaa esimerkiksi toteuttaa myöhemmin samantyyppisen tutkimuksen. Tämän tutkimuksen rajoitukset, joita käsitellään seuraavaksi, koskevat tutkimusaineiston hankintatapaa, väsymyksen vaikutusta oppijoiden suorituksiin, oppijoiden puhe-suoritusten arvioijia, kuunneltavien näytteiden tasoja ja mahdollisesti myös tutkimukseen osallistuneiden oppijoiden motivoituneisuutta.

### 6.4.1 Tutkimusaineiston hankintatapa

Tutkimuksen itsearviointilomakkeissa käytettiin dikotomisista "kyllä-" ja "ei"-vastausvaihtoehtoja siitä syystä, että tämä tutkimus liittyy suoraan DIALANG-hankkeeseen, jossa on käytetty samanlaisia dikotomisista vastausvaihtoehtoja. Pitkin matkaa tutkimusprosessin aikana kävi kuitenkin selväksi, että mahdollisuus valita vain jommankumman edellä mainitun vastausvaihtoehdon välillä osoittautui erityisesti joidenkin tutkimukseen osallistuneiden oppijoiden mielestä liian karkeksi. Kuten edellä tekstissä on jo esitetty, jotkut oppijat kokivat, että kielitaitoa on vaikea arvioida vain "kyllä" tai "ei" -vastausvaihtoehtojen



avulla. Tätä asiaa kommentoineet oppijat ehdottivat, että lomakkeissa olisi hyvä olla käytössä myös "lievempiä" vastausvaihtoehtoja kuin "kyllä" ja "ei"-vaihtoehdot. Ehdotetut lisävastausvaihtoehdot olivat "jotain siltä väliltä", "en osaa sanoa" tai "ehkä" -tyyppisiä vaihtoehtoja.

Tutkimukseen osallistuneen oppijan näkökulmasta tämänkaltaiset kommentit ja ehdotukset ovat tietenkin täysin ymmärrettäviä. Tutkimusta tehtäessä tehdään kuitenkin samalla valintoja, joiden tarkoituksena on päästä mahdollisimman hyvään lopputulokseen sen suhteen, mitä tutkimuksessa halutaan tehdä. Tämänkaltaisessa empiirisessä tutkimuksessa, jonka aineiston analyysi perustuu pitkälti tilastomenetelmien käyttöön, yksiselitteiset vastaukset tutkimukseen osallistuneilta voivat helpottaa joidenkin analyysien tekemistä. Lisäksi "kyllä"/"ei" -tyyppistä päätöksentekoa käytettiin hyvin yleisesti Viitekehyyksen taustalla olevassa kehittämistyössä ja siten myös DIALANG-järjestelmään liittyvissä tutkimuksissa. Tämän tutkimuksen rakenteen suunnittelussa on siten noudatettu tutkimusasetelman laatimisajankohtana 2000-luvun alussa vallalla olleita tutkimuskäytäntöjä.

Käsittelen nyt vielä pohtien muutaman esimerkin avulla, millaisia vaikutuksia tutkimuksessa käytetyillä dikotomisilla vastausvaihtoehdoilla voi olla tämän tutkimuksen tuloksiin. Luvussa 5.1.2.5 ilmeni, että itsearviointilomakkeiden väittämien tai kuunneltujen puhenäytteiden kohdalla taitotasoasteikon ääripäihin sijoittuneiden väittämien tai näytteiden toimivuus oli selkeästi riippuvainen toisaalta siitä, että vastausvaihtoehtoja oli vain kaksi ja toisaalta siitä, että taitotasoasteikon ääripäitä edustavia tutkimukseen osallistujia oli hyvin rajallinen määrä. Esimerkkinä tästä ilmiöstä tarkasteltiin C1-tasoista tehtävä 1:n näytettä numero 12, jossa puheentuottajan kielitaidon taso ja tuotos on monesakin suhteessa erinomainen. Itsearviointimittarin osana näyte 12 ei kuitenkaan erottele riittävästi, sillä vain kuusi oppijaa kaikista tutkimukseen osallistuneista on vastannut suoriutuneensa tehtävästä paremmin kuin nauhalta kuultu henkilö (= kyllä). Tarkasteltaessa näytteen 12 osio-testi-korrelaatiota voidaan sen todeta olevan hieman alle suositellun rajan .30. Sitä ei voida pitää erityisen huonona, kun otetaan huomioon, että vain kuusi osallistujaa valitsi "kyllä"-vaihtoehdon kyseisen näytteen kohdalla. Korrelaatiokertoimen arvosta voidaan siis päätellä, että kaikki kuusi "kyllä"-vaihtoehdon valinnutta vastasivat aivan "oikein" tämän näytteen kohdalla, koska kaikkien "kyllä"-vastauksen valinneiden kielitaidon taso todellakin sijaitsi taitotasoasteikon ylimmillä tasoilla (huomattakoon, että nämä henkilöt onnistuivat jokainen arvioimaan oman yleistaitotasonsa erinomaisesti, sillä ainakin tehtävän 1 suhteen ulkoiset arvioijat olivat arvioineet kyseisten henkilöiden taitotason täsmälleen samaksi kuin oppijat itse).

Ongelmalliseksi tutkimuksessa käytettyjen dikotomisten vastausten käytön tekee se, että osa niistä oppijoista, joiden kielitaito oli hyvää ja sijaitsi siten taitotasoasteikon ylimmillä tasoilla, saattoi kokea, että näyte oli hyvin samantasoinen kuin heidän oma suorituksensa, minkä jälkeen he valitsivat tutkimuksessa määriteltyjen vastausohjeiden mukaisesti "ei"-vaihtoehdon. Näyte 12 kuuluu niiden kuuden näytteen joukkoon, joita tarkasteltiin tutkimuksen eniten

tai vähiten "vaikea päättää" -rasteja keränneinä näytteinä. Näyte keräsi 12 luokittelematonta kommenttia, joista 7 on saatu oppijoilta, joiden itsearviointiyleistaitotaso sijoittuu taitotasolle B2–C2. Näitä merkintöjä tarkasteltaessa voidaan havaita, että C1-tasolle oman tasonsa arvioinneista oppijoista 4 on vastannut tämän näytteen kohdalla siten, etteivät he suoriutuneet omasta mielestään paremmin tehtävästä kuin näytteen puhuja (=ei). Näistä neljästä yksi piti kommenttiansa mukaisesti omaa tasoaan samana kuin näytteen puhujan taso, yksi kommentoi näytteen olleen täysin virheetön ja kaksi eivät kommentoineet mitenkään "ei"-valintaansa. Yksi C1-tasoinen oppija vastasi näytteen 12 kohdalla "kyllä" ilman muita kommentteja, mutta hänkin siis lisäsi rastin "vaikea päättää" -sarakeeseen. Taitotasolla C2 olevan oppijan vastaus oli niin ikään taitotaso C1 edustavan näytteen kohdalla se, ettei hän omasta mielestään suoriutunut tehtävästä paremmin kuin näytteen puhuja (= ei). Viimeinen seitsemästä tarkastellusta oppijasta oli tason B2 oppija, joka vastasi oletuksenmukaisesti "ei" C1-tasoiseen näytteeseen 12. Edellä kuvatut ongelmat liittyvät aikaisemmin kuvatun Bachmanin ja Palmerin analyysimallin reliabiliteetti ja käsitevalidius -kohtiin. Tutkimuksessa ja siinä kehitetyissä itsearviointivälineissä on siis mahdollisesti näissä laatupiirteissä ongelmia, jotka johtuvat käytetystä "kyllä" / "ei"-vastaustavasta.

Edellä tarkastellut esimerkit selventävät osaltaan sitä tosiasiaa, että jos tutkimuksen osallistujissa on suhteellisen vähän aivan ylimpiä taitotasoja edustavia oppijoita ja jos he kokevat ylimmille tasolle sijoittuvien näytteiden kohdalla kielitaitonsa olevan samalla tasolla kuin näytteen puhujan, ei-vastausten kokonaismäärä muodostuu suhteettoman suureksi, mikä taas saa aikaan sen, että oppijoiden vastauksissa ei ole riittävästi variaatiota.

Tässä yhteydessä voidaan siis samalla pohtia, olisiko tämänkaltaisissa tutkimuksissa syytä jollakin tavalla kiinnittää huomiota niihin oppijoiden itsearviointeihin, joissa oppija on halunnut ilmaista mielipiteenään, että hän selviytyi tehtävästä samantasoisesti kuin kuunnellun näytteen puhuja. Nythän nämä "samaa tasoa" -kommentteja saaneet näytteet jäävät kokonaisuuden kannalta ikään kuin piiloon tässä tutkimuksessa, jossa samaa tasoa ilmaistaessa valitaan "ei"-vaihtoehto. Tässä tutkimuksen käytännön toteutuksen jälkeisessä vaiheessa voidaan todeta, että "vaikea päättää" -rastien yhteydessä olevista kommentteista osa on "samaa tasoa" -tyyppisiä kommentteja. Kuitenkaan mitään tarkkaa tietoa siitä, kuinka paljon näitä samaa tasoa ilmaisevia kommentteja tutkimusaineistossa on tai kuinka monet niistä esiintyvät "vaikea päättää" -rastien yhteydessä, ei ole käytettävissä.

Näiden pohdintojen jälkeen onkin syytä todeta, että tässä tutkimuksessa käytetty aineistonhankintatapa näyttäisi vaikuttavan jonkin verran tutkimuksesta saatuihin tuloksiin erityisesti hyvää ja erittäin heikkoa kielitaitoa edustavien väittämien/näytteiden kohdalla.

#### 6.4.2 Väsymyksen vaikutus oppijoiden suorituksiin

Kuunneltaviin puhenäytteisiin perustuvan itsearvioinnin tuloksia tarkasteltaessa herää kysymys, mikä vaikutus erityisesti tutkimuksen tuloksiin on voinut

olla sillä, että osa osallistujista vaikutti selvästi väsyneeltä tutkimuksen suorittamisen loppuvaiheessa. Viimeinen tutkimuksen kielistudiossa toteutettu osa oli itsearviointi, jossa omaa puhe-suoritusta verrattiin tehtävän 2 puhenäytteisiin. Olisiko siis ehkä mahdollista, että osa tehtävää 2 koskevista heikoimmista tuloksista tässä tutkimuksessa voisivat selittyä sillä, että oppijat olivat väsyneitä eivätkä keskittyneet suorittamaan tehtävään 2 liittyvää itsearviointia yhtä tarkasti kuin aiemmin toteutettua tehtävää 1 koskevaa itsearviointia? Mahdollisen väsymyksen vaikutusta tuloksiin ei tietenkään voida todistaa millään tavalla näin jälkikäteen, mutta epäilyskin tämänkaltaisen seikan vaikutuksista tutkimustuloksiin on syytä ottaa huomioon.

Tulevaisuudessa siten tämänkaltaisissa tutkimuksissa, jotka koostuvat kerralla suoritettavista monista eri aktiviteeteista ja joissa tutkimukseen osallistumisaika on verrattain pitkä, kannattaisi varmasti harkita tutkimuksessa käytettävien lomakkeiden esittämistä oppijoille eri järjestyksessä. Tässä tutkimuksessa vastauslomakkeiden järjestyksenvaihto olisi voitu tehdä niin, että puhenäytteisiin perustuvan itsearvioinnin kohdalla olisi vuoroteltu sen suhteen kumman tehtävän, tehtävän 1 tai 2, itsearviointi olisi jäänyt viimeisenä suoritettavaksi aktiviteetiksi. Tällä tavoin oltaisiin voitu välttyä siltä, että tehtävää 2 koskeva puhenäytteiden kuunteluun perustuva itsearviointi olisi kärsinyt mahdollisesta oppijoiden väsymyksen vaikutuksesta. Oppijoiden väsymisestä johtuvat ongelmat aiheuttavat epävarmuutta erityisesti arviointivälineiden käsitevaliditeetin ja vuorovaikutukseen Bachmanin ja Palmerin (1996) mallissa, koska väsymys saattaa vähentää mittarin kykyä mitata haluttuja taitoja ja vääristää sitä tapaa, jolla oppijan henkilökohtaiset kognitiiviset ja affektiiviset ominaisuudet ovat mukana tehtävän suorittamisessa.

### 6.4.3 Oppijoiden puhe-suoritusten arvioijat

Osallistujien puhe-suoritusten arviointiin käytettiin kolmea arvioijaa, joista yksi oli vaihtuva arvioija siten, että yhdeksi arvioijaksi oli yhdistetty viiden eri arvioijan työ. Arvioijien johdonmukaisuutta ja luotettavuutta selvitetessä paljastui kuitenkin, että kyseinen yhdistämisratkaisu ei osoittautunut toimivaksi. Viidestä arvioijasta koostuvan vaihtuvan arvioijan arvioinnit päätettiin siis poistaa lopullisesta aineistosta. Arviointien epäjohdonmukaisuus johtui osittain siitä, että arvioijien työmäärä yritettiin pitää pienenä siten, että jokainen "vaihtuva" arvioija sai arvioitavakseen vain rajallisen määrän suorituksia, mikä ei välttämättä helpottanut arviointityötä kokonaisuudessaan. Oppijoiden suorituksia arvioitavaksi jaettaessa saattoi käydä myös niin, että lukiolaisten kanssa työskennellyt arvioija saattoikin saada arvioitavakseen esimerkiksi vain aikuisten yliopistolla opiskelevien henkilöiden suorituksia. Eri arvioijien tottumus tietynlaisen ja tietynlaisen oppijaryhmän kielitaidon arvioimiseen yhdistettynä suhteellisen pienen arvioitavien suoritusten määrään voi selittää johdonmukaisuuden puutteen vaihtuvien arvioijien työn kohdalla. Selvää siis joka tapauksessa on, että kolmanneksi arvioijaksi tarkoitettu vaihtuva arvioija ei toiminut tässä tutkimuksessa siten kuin tutkimuksen suunnitteluvaiheessa toivottiin.

On kuitenkin syytä tähdentää, että vaikka vaihtuvien arvioijien arvioinnit päätettiin poistaa lopullisesta aineistosta, kenenkään osallistujien puhe-suoritus-tien arviointiin osallistuneen arvioijan ammattitaitoa ei ole syytä epäillä arvioin-tien aineistosta poistamisen johdosta. Olisi näet epäoikeudenmukaista ja epä-eettistä todeta kaikkien vaihtuvien arvioijien, jotka olivat kielitaidon arviointiin koulutettuja ja arviointityöstä kokemusta omaavia, olevan arviointityössään epäjohdonmukaisia. Arviointien poistaminen lopullisesta aineistosta liittyy mi-tä suurimmissa määrin siihen tosiasiaan, että kokonaiskuvan luominen koko arviotavasta aineistosta ohjaa ja helpottaa arviointityön sujumista, vaikka osal-listujien suorituksia ei taitotasoarvioinnissa verratakaan keskenään. Edellä ku-vattu liittyy Bachmanin ja Palmerin (1996) mallin kohtiin reliabiliteetti ja mah-dollisesti myös käsitevalidius (riippuen siitä, mihin piirteisiin eri arvioijat kiin-nittivät huomiota).

#### 6.4.4 Kuunneltavien näytteiden tasot

Tutkimuksessa käytettiin kahta tehtävää, joista toisesta oppijat kuuntelivat 26 puhenäytettä (tehtävä 1) ja toisesta 23 puhenäytettä (tehtävä 2). Muun muassa näytteiden toimivuutta tarkasteltaessa todettiin, että eri taitotasoja edustavien puhenäytteiden lukumäärä oli osittain epätasainen. Erityisesti taitotasosteikon ylätasolle sijoitettavia näytteitä ei ollut kovin montaa; esimerkiksi tehtävässä 1 oli 2 taitotason B2 näytettä, 3 taitotason C1 näytettä ja ei yhtään taitotason C2 näytettä. Tehtävässä 2 kolmelta ylimmältä taitotasolta on näyte, mutta edelleen niiden määrä on melko rajallinen: 1 näyte (B2), 2 näytettä (C1) ja 1 näyte (C2).

Molempiin tutkimuksessa käytettyihin tehtäviin liittyvät kuunneltavat näytteet painottuvat siten taitotasosteikon alapäähän, mikä onkin luonnollista sitä taustaa vasten, että näytteet on valittu Yki-testien perus- ja keskitason suorituksista. Yki-testeistä valittujen näytteiden lisäksi tutkimuksen käyttöön nauhoitettiin erikseen muutamia ylimmän tason suorituksia. Näin jälkikäteen on kuitenkin myönnettävä, että olisi ollut paikallaan hankkia vielä lisää muutamia ylimmän tason suorituksia, jotta itsearviointimittarit olisivat olleet tasapainoi-sempia eri taitotasojen näytteiden osalta.

On silti muistettava, että monesti tutkimuksen näytteiden lopullinen taso määräytyi viiden eri arvioijan arviointityön pohjalta siten, että näyte sijoitettiin sille tasolle, mille enemmistö arvioijista näytteen sijoitti. Tutkimuksen näytteis-sä oli siis jonkin verran näytteitä, joiden tasoista arvioijien välillä vallitsi erimie-lisyys (muun muassa juuri tehtävän 1 edellä keskusteltu näyte 12, jonka tasoksi määräytyi C1, koska kolme arvioijaa sijoitti näytteen tälle tasolle, vaikka 2 muu-ta arvioijaa viidestä sijoittivatkin kyseisen näytteen tasolle C2). Esimerkiksi tä-män näytteen kohdalla olisi riittänyt, että vielä yksi arvioija olisi ollut sitä miel-tä, että näyte 12 on C2-tasolle kuuluva näyte, jotta tehtävässä 1 olisi ollut mukana myös C2-tasoinen näyte. Näytteiden lopulliset tasot ovat siten enemmistö-päätökseen nojaavan kompromissin tulosta, ja tässä mielessä joissakin tapa-uksissa näytteiden lopullisten tasojen voidaankin ajatella olevan suhteellisia.

Lisäksi on vielä korostettava tutkimukseen liittyvänä yleishavaintona sitä tosiasiaa, että jos jonkin näytteen taitotason määrittelyn suhteen näyttää olevan

ongelmia, näytteen käyttö mittarissa olisi parasta hylätä. Tällaisesta tapauksesta esimerkkinä voidaan mainita tehtävän 1 näyte 15, jonka tasosta arvioijien mielipide jakautui kolmelle eri taitotasolle (A2, B1 ja B2). On näet hyvin todennäköistä, että jos arvioijien on ollut vaikea päästä yksimielisyyteen jonkin näytteen tasosta, niin myös oppijoilla tulee olemaan vaikeuksia verrata omaa tuotostaan kyseiseen näytteeseen.

Tämän tutkimuksen aineisto on kerätty ja arvioitu vuonna 2000 ja siihen aikaan ei valitettavasti vielä ollut mahdollista tehdä arvioijien työtä yksityiskohtaisesti tarkastelevia Facets-analyysejä. Mikään yllä olevista pohdinnoista ei kuitenkaan poista sitä tosiasiaa, että molempia tehtäviä koskevien itsearviointimittareiden olisi ollut hyvä sisältää enemmän ylimpien taitotasojen puhenäytteitä. Näin erityisesti mittareiden käsitevalidius olisi entisestään parantunut.

#### 6.4.5 Tutkimukseen osallistuneet oppijat

Kun pohditaan tämän tutkimuksen tulosten yleistettävyyttä muihin kieliin ja muiden kielten opiskelijoihin, on syytä ottaa huomioon muutama seikka. Tutkimukseen osallistujien enemmistö oli 21–25-vuotiaita naisyliopisto-opiskelijoita, joista suurin osa (53,8 %) oli opiskellut ranskan kieltä kolme vuotta lukiossa. Niiden osuus, jotka olivat opiskelleet ranskan kieltä pitkään (yli viisi vuotta) yliopistossa, oli vain 6,7 % kaikista tutkimukseen osallistuneista oppijoista. Enemmistö tutkimukseen osallistuneista oppijoista ei siis ollut erikoistunut ranskan kieleen yliopistollisten kieliopintojensa kautta, ja voidaan siis olettaa, että suurin osa tutkimukseen osallistuneista oppijoista oli muita kuin ranskan pääaineopiskelijoita. Tässä mielessä heidän ei voida tulkita edustavan mitenkään erityisemmin valikoitunutta joukkoa. Suurin osa tutkimukseen osallistuneista oppijoista ilmoittautui itse halukkaiksi osallistumaan tutkimukseen, mikä osoittaa suurehkoa motivoituneisuutta oppijoiden puolelta. Tästä näkökulmasta asiaa tarkastellessa voidaan kuitenkin kysyä, edustavatko tutkimukseen osallistuneet oppijat keskimääräistä kieltenopiskelijaa ja ovatko heidän itsearviointejaan koskevat tuloksensa suoraan yleistettävissä keneen tahansa ja minkä tahansa opiskeltavan kielen opiskelijaan. On näet mahdollista, että tutkimukseen osallistuneet oppijat (kahta lukioluokkaa lukuun ottamatta) edustavat tavanomaista kieltenopiskelijaa motivoituneempaa joukkoa, koska he ovat omatoimisesti hakeutuneet osallistumaan tutkimukseen, johon osallistuminen edellytti oman ajan käyttöä pääsääntöisesti illalla. Tältä pohjalta voidaankin kysyä, voiko tämänkaltaisella motivoituneisuudella olla ainakin jossain määrin vaikutusta itsearviointin tuloksiin ja sitä kautta tämän tutkimuksen tulosten yleistettävyyteen.

### 6.5 Mahdollisia jatkotutkimusaiheita

Käsillä oleva tutkimus sisältää monenlaisia työvaiheita ja toimenpiteitä. Tutkimuksen aineisto on melko laaja, mikä on mahdollistanut sen, että tarkasteltavia

tutkimuskysymyksiä on työssä useita. Tutkimuksen tuloksia, antia ja rajoituksia pohdittaessa eteen on avautunut joitakin ideoita, joissa voisi ajatella olevan potentiaalia aiheeseen liittyviin jatkotutkimuksiin.

Ensimmäinen mahdollinen jatkotutkimusaihe olisi ilman muuta se, minkälainen vaikutus tämänkaltaisen tutkimuksen tuloksiin olisi sillä, jos oppijoiden itsearvioinneista otettaisiin erikseen tarkasteluun "kyllä"- ja "ei"-vastausvaihtoehtojen rinnalle "samaa tasoa" -vaihtoehto. Tämänkaltaisen tutkimusasetelma edellyttäisi luonnollisesti tutkimusasetelmaan liittyvien tilastollisten analyysien uudelleen määrittelyä ja tutkimusta koskevien lomakkeiden muokkausta. Tällä tavalla luultavasti voitaisiin olla enemmän uskollisia tutkimukseen osallistuneiden oppijoiden itsearviointia koskeville näkemyksille ja valinnoille sekä saada erilaista tietoa itsearvioinnista.

Toinen kysymys, joka voisi olla mielenkiintoinen selvitettävä tämän tutkimuksen kaltaisissa tutkimuksissa, olisi se, mitä mieltä tutkimukseen osallistuneet oppijat ovat tutkimukseen valituista tehtävistä. Jos oppija on motivoitunut ja kokee käytetyn tehtävän relevantiksi, tällä asialla on merkitystä siihen, miten hän onnistuu omassa tuotoksessaan, kuten edellä todettiin. Jos ajatellaan, että jotkut oppijat eivät ole pitäneet jompaakumpaa tai pahimmassa tapauksessa kumpaakaan tutkimukseen valituista tehtävistä itselleen mielenkiintoisena ja/tai relevanttina, oppijan oma puhe-suoritus on saattanut olla hyvinkin hänen yleistä taitotasoaan alempana. Tässä tutkimuksessa on verrattu oppijan oman puhetuotoksen, kuunnellun puhenäytteen tason ja oppijan itsearvioinnin välisiä yhteyksiä tai vastaavuuksia. Siten olisikin toivottavaa jo tilastollisten analyysien totuudenmukaisuuden kannalta, että oppijan puhetuotoksen taso vastaisi tutkimuksessa sitä tasoa, jolle hänen suullinen kielitaitonsa keskimäärin kuuluu.

Kolmas aihe voisi olla oppijoiden itsearvioinnista saatujen laadullisten vastausten tarkempi analyysi.

Neljäntenä hankkeena voisi olla se, että tämän tutkimuksen mittareiden pohjalta laadittaisiin lyhennetyt opetuskäyttöön tarkoitetut versiot tai useita versioita eritasoisille oppijaryhmille.

Viides mieleen tuleva aihe olisi sen selvittäminen, millaisia tuloksia saataisiin, jos tämän tutkimuksen kaltainen tutkimus toteutettaisiin jonkin toisen kielen kuin ranskan oppijoiden parissa. Toisen kielen lisäksi mielenkiintoista olisi myös selvittää, millaisia eroja tuloksiin saataisiin, jos tutkimus toteutettaisiin eri-ikäisten ja eri taitotasolla olevien oppijoiden kanssa.

## 6.6 Tutkimuksen tulosten hyödynnettävyys opetuksessa

Tutkimuksen loppumetreillä on syytä pysähtyä pohtimaan, mitä tämänkaltaisella tutkimuksella on annettavaa eli onko tutkimuksesta hyötyä vain tiedeyhteisölle siinä mielessä, että se omalta osaltaan auttaa laajentamaan kuvaa tietyn tyyppisen itsearvioinnin toimivuudesta tietyssä kontekstissa.

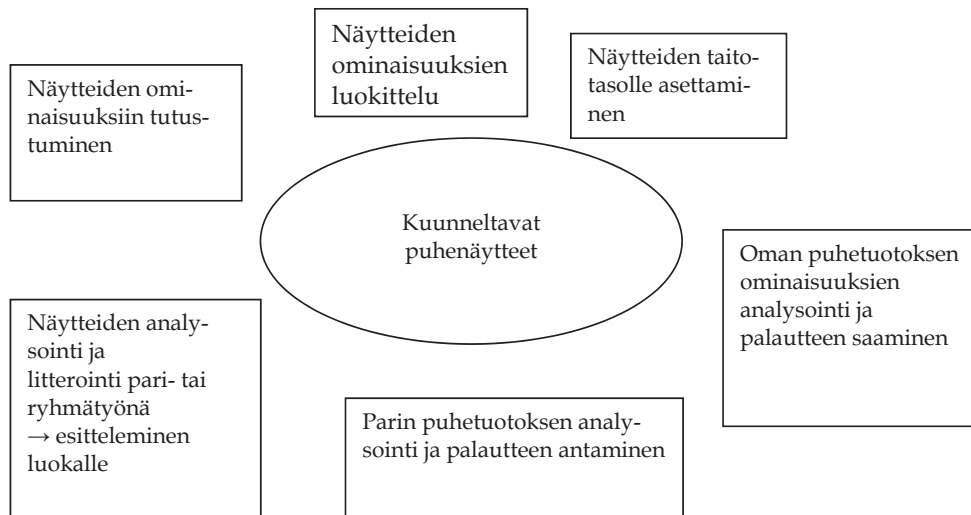
Todettakoon aluksi, että tutkimuksesta saadut tulokset liittyvät siihen enemmistönä olevaan itsearviointitutkimusten joukkoon, joka osoittaa positiivista vastaavuutta itsearvioinnin ja ulkoisen arvioinnin välillä. Tämä tutkimus osoitti sekä perinteisen paperilla tehtävän että uuden työssä kokeillun kuunneltaviin puhenäytteisiin pohjautuvan itsearvioinnin olevan hyvin samansuuntaista kuin vastaava ulkoinen arviointi. Kun otetaan huomioon, että tutkimukseen osallistuneita oppijoita ei millään tavalla koulutettu oman suullisen kielitaitonsa arvioimiseen, itsearvioinnin ja ulkoisen arvioinnin verraten hyvää vastaavuutta voidaan pitää lupaavana tuloksena itsearvioinnin käytön kannalta.

Niinikään voidaan miettiä, voisiko tämänkaltaisesta tutkimuksesta olla myös käytännön hyötyä esimerkiksi opetuskäytössä. Tutkimuksen tekijänä toivoisin, että työllä olisi konkreettista annettavaa myös luokkahuoneessa tapahtuvalle opetukselle. Tämänkaltaisen empiirisen tutkimuksen vahvuuksia on ilman muuta melko laaja kerätty aineisto, mutta samalla on muistettava, että aineiston kerääminen ainakin ranskan kielessä vaati paljon työtä ja aikaa. Tuskin voidaankaan ajatella, että tässä tutkimuksessa harjoitettu itsearviointimenetelmä voitaisiin ikään kuin suoraan siirtää esimerkiksi luokkahuoneeseen, koska pelkkä aineistonkeruu on vaatinut paljon resursseja. Toisaalta tämänkaltaisen itsearviointiprosessin siirtäminen suoraan opetustilanteeseen ei olekaan perusteltua, koska opetuksen tarpeet eivät välttämättä ole samat kuin tutkimuksen. Kielenopetukseen sisältyvän itsearviointimenetelmän käyttöönottamisen kannalta tärkeintä on varmasti se, että tietty itsearviointimenetelmä on todettu esimerkiksi tutkimuksen avulla toimivaksi, minkä jälkeen sitä halutaan kokeilla käytännössä. Jos tutkimuksessa esiteltyä puhenäytteisiin perustuvaa itsearviointia haluttaisiin kokeilla luokassa, olisi varmasti syytä muokata itsearviointitapaa siten, että siitä tulisi taloudellisempi. Kuten Bachmanin ja Palmerin (1996) edellä esitellyssä testin luotettavuutta tarkastelevassa mallissa tuli esille, testien yksi tärkeä ominaisuus on niiden käytännöllisyys eli se, että käytettävissä olevat resurssit ovat tasapainossa suhteessa siihen, millaisia resursseja tietyn testin tai opetusmenetelmän käyttöönotto luokassa edellyttää. Opetusmenetelmän taloudellisuutta mietittäessä olisi siis saatava riittävän paljon hyötyä aineistosta, johon on uhrattu tietty määrä resursseja.

On myös syytä huomioida, että tämä tutkimus edustaa tavallaan perustutkimusta uuden itsearviointivälineen toimivuudesta; tutkimus toteutettiin tietyllä tavalla "ihanneolosuhteissa" eli aineistoa oli paljon, tutkimusasetelman suunnitteluun käytettiin runsaasti aikaa ja käytetyt itsearviointimittarit olivat riittävän pitkiä. Seuraava askel olisikin se, että tässä tutkimuksessa kokeillusta menetelmästä kehitettäisiin sovelluksia opetuksen ja oppimisen eri tarpeisiin.

Tutkimuksessa käytettyä puhenäytteisiin perustuvaa itsearviointiaineiston käyttöä voitaisiinkin hyvin laajentaa. Esimerkiksi tässä tutkimuksessa harjoitetun itsearviointitavan lisäksi (tehtävän suorittaminen, oman suorituksen kuuntelu, oman suorituksen vertaaminen kuunneltaviin puhenäytteisiin) aineistoa voitaisiin käyttää hyödyksi esimerkiksi seuraavilla tavoilla luokkahuoneessa:





KUVIO 8 Tutkimuksessa käytetyn kuunneltaviin puhenäytteisiin perustuvan itsearviointimenetelmän käyttämahdollisuuksia kielten opetuksessa.

Tässä vaiheessa tutkimusprosessia ja uutta itsearviointimenetelmää pohdittaessa on myönnettävä, että uuden menetelmän käyttöönotto esimerkiksi opetustilanteissa edellyttää varmasti enemmän vaivannäköä ja aikaa opettajalta kuin perinteisen paperilla tehdyn itsearvioinnin käyttäminen. Mitkä sitten ovat uuden itsearviointimenetelmän hyödyt, ja ovatko mahdolliset hyödyt niin suuria, että menetelmää kannattaisi suositella opetuskäyttöön? Päästäänkö uuden puhenäytteisiin perustuvan itsearviointimenetelmän avulla parempiin tuloksiin kuin esimerkiksi perinteisen paperilla suoritetun itsearvioinnin avulla?

Ensinnäkin tämän tutkimuksen tuloksiin pohjautuen voidaan todeta, että uusi puhenäytteisiin perustuva itsearviointimenetelmä osoittautui riittävän hyväksi, validiksi ja reliabeliksi, jotta sitä voidaan hyvin ajatella käytettävän esimerkiksi opetuskäytössä. Menetelmällä on tiettyjä etuja verrattuna paperilla tehtyyn itsearviointiin. Suurimpana etuna voidaan pitää sitä, että oppija pääsee operoimaan opiskelemaansa kielen kanssa konkreettisella tasolla kuunnellessaan sekä omaa puhe-suoritustaan että toisten puhe-suorituksia. Todelliseen puheeseen liittyvä vertaaminen ja itsearviointi analysoimalla sekä omia että toisten puhe-suorituksia kuvion 8 mukaisesti mahdollistaa opettavaisen, konkreettisesti perustuvan ja mielekkään tavan tutustua puhe-suoritusten sisältöön.

Tämä tutkimus valaisee sitä kuvaa tai näkemystä, joka kullakin oppijalla on omasta osaamisestaan ranskan kielen suullisen kielitaidon osalta. Tutkimus vertaa tätä kuvaa koulutettujen arvioijien arviointien avulla esille tuomaan kuvaan oppijan kielitaidosta. Tutkimuksen aikana oppijalla on ollut monta mah-

dollisuutta toisaalta osoittaa osaamisensa ja toisaalta antaa palautetta tutkimukseen liittyvän itsearvioinnin eri puolista. Erityisesti oppijoiden antama palaute on mahdollistanut sen, että on päästy tutustumaan oppijan osaamiseen myös oppijan omasta näkökulmasta käsin. Tämä onkin ollut erityisen hedelmällistä tutkimusprosessin ja vuorovaikutuksen näkökulmista.

Eräs tärkeä syy tutkimuksessa käytetyn uuden itsearviointimenetelmän hyvään toimimiseen on se, että kaikenlainen arviointi ja vertailu suhteessa toisiin on hyvin tyypillistä ihmisluonnolle. Harjoitamme vertailua jo ihan pienestä pitäen ja jatkamme tätä toimintaa jossain määrin koko elämämme ajan. Vertailu on varmasti myös huomattavasti helpompaa kuin vaikkapa luokittelu. Jos esimerkiksi tässä tutkimuksessa olisi pyydetty oppijoita luokittelemaan puhe-suoritusten eri piirteitä, tutkimuksesta saadut tulokset olisivat saattaneet olla hyvin erilaisia kuin mitä ne tällä tutkimusasetelmalla olivat. Vertailua käytetäänkin hyväksi jonkin verran nykyisin silloin, kun esimerkiksi arvioidaan puhumista tai kirjoittamista maamerkkisuoritusten avulla. Muun muassa tässä tutkimuksessa jo mainitussa Eurooppalaisessa kielitaito-indikaattoritutkimuksessa kirjoitussuoritusten arviointi tapahtui vertaamalla kirjoitussuorituksia valmiisiin eritasoihin kirjoittamisen maamerkkeihin (First European Survey on Language Competences. Final Report 2012, 107–111).

Itsearvioinnin aloittamisesta puhuttaessa on syytä muistaa, että ennen kuin varsinaiseen itsearvioinnin harjoittamiseen päästään, oppijoita täytyy tutustuttaa ja kouluttaa asiaan. Aikataulua suunniteltaessa on otettava huomioon, että usein joudutaan tekemään monenlaista asenteiden muokkaustyötä, ennen kuin kaikki oppijat saadaan mukaan itsearviointiin. Itsearvioinnin yhtenä suurimpana haasteena voidaankin pitää oppijoiden asenteiden muuttamista itsearvioinnille myönteisiksi. Muun muassa Holec totesi jo 1980-luvun alussa (1982), että oppijan autonomia voidaan ottaa käyttöön vain toteuttamalla kaksi rinnakkaista prosessia: 1. opittavan kompetenssin haltuunotto ja 2. vähittäinen tottumuksista pois opetteleminen. Vasta näiden prosessien jälkeen oppija voi vapautua kielenoppimista ja omaa oppijarooliaan koskevista ennakkoluuloistaan. Ensimmäinen askel itsearvioinnin suuntaan pitääkin sisällään sen, että saadaan vähitellen osoitettua oppijoille, miksi itsearviointia kannattaa tehdä ja mitkä ovat sen konkreettiset hyödyt. Vasta tämänkaltaisen perustyön jälkeen voidaan lähteä etenemään itsearvioinnin polulla. Yksi tehokas tapa muuttaa tottumuksia ja asenteita itsearvioinnille myönteisemmiksi on jo työssä aiemmin mainittu reflektointi. Reflektointi on usein tiedostamatonta toimintaa, mutta siitä voidaan tehdä tietoista (Lammi 2000, 161). Tietoisella reflektoinnilla voidaan muun muassa purkaa syvään juurtuneita uskomuksia esimerkiksi itsestä oppijana sekä muuttaa odotuksia, suhtautumista ja työtapoja (esim. Kohonen 1994, 40; 1999). Riskinotto, nöyryys, epäily, hämmästely, epävarmuus ja uusien näkökulmien etsiminen kuuluvat Lammen (2000, 162) mukaan kaikki reflektiiviseen ajatteluun. Tärkeintä itsearvioinnin käytössä on se, miten oppija itse kasvaa sen aikana. Itsearviointiin kannustamisen ja oppilaiden motivoinnin valossa suotavaa olisi, että kouluissa opettavat henkilöt olisivat valmiita panostamaan riittävästi siihen, että oppilaita harjoitettaisiin itsearviointiin. Toivottavaa olisi myös, että opetustehtävissä toimivat henkilöt uskoisivat riittävästi siihen, että itsear-

vioinnista saatava hyöty on molemminpuolista ja että sen säännöllisestä harjoittamisesta mitä todennäköisimmin on hyötyä oppijalle myös hänen koulun jälkeisellä opiskelu- ja työurallaan.

Kuten aiemmin tutkimuksessa kävi ilmi, tutkimuksen paperilla tehdyssä itsearvioinnissa käytetyt väittämät on laadittu DIALANG-kielitaidon arviointityökalun suullisen osuuden käyttöön. Ne eivät vielä ole julkisesti käytettävissä internetissä, vaikka muihin kielen osataitoihin hankkeessa kehitetty materiaali onkin ollut DIALANG-järjestelmän kautta yleisön käytettävissä jo vuosia. Toiveena olisikin, että DIALANG-järjestelmä sisältäisi tulevaisuudessa myös suullisen kielitaidon arvioinnin mahdollistavan osan. Tämänkin tutkimuksen tuottamien tilastollisiin ja laadullisiin analyyseihin pohjautuvien tulosten toivoisi voivan hyödyttää DIALANG-järjestelmän suullista kielitaitoa arvioivan osan edelleenkehittämistyötä.

On syytä korostaa, että tämän tutkimuksen tarkoitus ei ole yrittää osoittaa, että kuunneltaviin puhenäytteisiin perustuva itsearviointi edustaisi jollain tavalla yliverstaista tai arvokkaampaa itsearviointia kuin perinteinen paperilla tehty itsearviointi. Molemmat itsearviointimenetelmät, puhenäytteiden kuuntelemiseen perustuva ja paperilla tehty itsearviointi, voidaan nähdä toinen toistaan täydentävinä toimintoina ilman, että niitä tarvitsisi asettaa paremmuusjärjestykseen tai vastakkain. Molemmilla menetelmillä voidaan saada hankittua arvokasta tietoa oppijan osaamisen eri puolista. Tämän tutkimuksen tarkoituksena on osoittaa, että uudenlainen puhenäytteiden kuunteluun ja oman suorituksen vertailuun perustuva itsearviointi on luotettavaa, toimivaa ja motivoivaa autenttisten puhenäytteiden ansiosta. Sitä voidaan siis käyttää sekä oppijan reflektiivisten taitojen kehittämiseen että hänen osaamisensa arviointiin.

Toivottavaa olisi, että tästä työstä voisi olla hyötyä käytännön kielten opetustyötä tekeville henkilöille ja oppijakeskeisen opetuksen edistämiseksi. On todennäköistä, että oppijakeskeinen lähestymistapa kielten opetukseen kuuluu nykyhetken lisäksi kielten opetukseen myös tulevaisuudessa. Totean samoin kuin Dregemark (2009, 234) totesi englannin kielen kirjoittamista koskevan itsearviointitutkimuksensa lopuksi: tämän tutkimuksen tutkimuskysymysten vastaukset osoittavat, että puhenäytteisiin perustuva itsearviointitapa on *yksi* mahdollinen tapa toteuttaa monipuolisempia ja oikeudenmukaisempia arviointikäytänteitä (vieraiden) kielten opetustilanteissa. Itsearviointi voi oikein toteutettuna olla yksi niistä keinoista, joilla rikotaan arviointiin liittyviä korostuneita valtasuhteita ottamalla oppija mukaan oman opiskelunsa suunnitteluun, toteuttamiseen ja arviointiin.

## SUMMARY

### **Background, purpose and theoretical framework**

Oral proficiency is a key aspect in the development of our language skills at all levels of education. It is now recognized that assessment of proficiency must be as versatile and complete as possible. The so-called alternative assessment methods, such as self-assessment or peer-assessment, enable a more complete assessment of learner's knowledge. Self-assessment relates to the constructivist notion of learning in which the learner and his or her own active collaboration is brought into focus. Self-assessment of the learner's action involves learner's self-reflection and can be used to support the view that the learners are able to plan their studies, to set themselves learning goals and to monitor the implementation of the objectives. When the learner is able to manage his or her own learning process, his or her autonomy, self-reliance and self-confidence will grow. All these will be useful in various areas of lifelong learning.

The study examines oral language assessment from the point of view of self-assessment by comparing self-assessment with external assessment. The theoretical assumptions of the study are concepts related to learner centeredness such as the reflectivity and the autonomy of the learner, learning to learn and lifelong learning. The study also relates to the Common European Framework of Reference for Languages, the Finnish National Foreign Language Certificates,<sup>25</sup> as well as the EU-funded project DIALANG<sup>26</sup>, from which the self-assessment statements used in this research were borrowed.

The main purpose of this study was to test the functionality of a new kind of speech sample-based self-assessment method where adult French language learners listened to speech samples and compared their own oral performance with the samples.

The study was designed to investigate how effective the learners' use of different methods of self-assessment were as well as the extent to which self-assessments corresponded to each other and to external assessment. Qualitative analysis also examined learners' perceptions of the functionality, usability and understandability of written self-assessment statements used in the study. In addition, the study examined the characteristics of the speech samples which learners found either unusually easy or difficult when comparing their performance with the samples.

### **Methods and Data**

The study involved 209 French language learners, most of whom were university students. The research included two speaking tasks in French, which the learners did in a language laboratory, as well as four different self-assessment measures or instruments (general self-assessment on paper, situation-specific

---

<sup>25</sup> [www.oph.fi](http://www.oph.fi)

self-assessment on paper, self-assessment based on speech samples in tasks 1 and 2). The study used two self-assessment methods, the traditional paper-based self-assessment and the new method where the learners listened to the speech samples and compared their performance with these. The first two above-mentioned instruments, the general self-assessment and situation-specific self-assessment on paper, were developed for the DIALANG online diagnostic language assessment tool.

The data collection process started with the action where the learners assessed their oral French language skills against the six general skill levels of the Common European Framework (A1-C2). After this, the learners answered 42 written situation-specific self-assessment statements by selecting either a positive ("yes") or negative ("no") option, depending on the content of the statement. The learners chose between "yes" and "no" by considering whether the statement described his/her own language skill or not. This traditional self-assessment method on paper was then followed by new speech sample based self-assessment. In the new method the learners completed the speaking task and their performance was recorded. After this, the learners listened to their own performance, so that they had an image of its style and the way they succeeded in doing the task. Finally, the learners listened to speech samples by other learners performing the same task (these samples had been rated at different levels). Learners responded by selecting "yes" in the answer form if they thought their own performance was better than the one that they listened to or "no" if they thought their own speech performance was worse, or as good as the one they heard.

During the data collection process a variety of feedback was collected from learners. They gave feedback on the written self-assessment statements used for self-assessment on paper, self-assessment in general, and whether they had any difficulties in answering to the self-assessment statements on paper. If the learners responded that they had had some problems, they were asked to specify what those problems were. In addition, they were asked to write down the number of the self-assessment statement they found harder than usual.

Furthermore, learners were asked to comment on why they chose "yes" or "no" when listening to the speech samples. At this point, learners had an opportunity to express if they found it difficult to compare their own performance with some specific speech sample. The study therefore includes information on what language features the learners paid attention to while listening to the speech samples. The study also provides information on what kind of factors or features of the language may have caused difficulties for the learners while they were listening to the speech samples and comparing their own performance with them.

The results of the self-assessment method based on speech samples were then compared to the results of the traditional paper-based self-assessment and external assessment. The study used both traditional and IRT-based statistical methods. The statistical methods used included both classical test theory methods (item-total correlations and Cronbach's alphas) and the modern test theory

indicators such as Infit mean square values, which indicate the degree to which each item goes together with the other items. In addition to the quantitative statistical analyzes the study includes a qualitative part which aims to deepen the information gathered by quantitative analyses.

## Results

In this kind of empirical study where the functionality of a new method for self-assessment was tested, it was important to make sure that the instruments used in the study had a sufficiently good reliability. In general, it should be noted that the reliability of all three self-assessment instruments was good. The Cronbach's alpha coefficient measuring the reliability was .94 in both task 1 and in the paper-based self-assessment. The alpha coefficient for task 2 was .89. Overall, the proportion of items exceeding the recommended item-total value of .30 was approximately 85% in all three self-assessment instruments. However, there were significant differences in the magnitude of the correlation values between instruments. Task 1 had about 65% of its items exceeding the .60 correlation whereas the same value was 28% in the paper-based self-assessment and only 21 % for the task 2.

88–92 % of the Infit mean square values from the Rasch analysis were inside the recommended range of values (.7-1.3) in all three instruments. Again, the instrument used in task 1 was best in terms of quality with 27 % of the items having the ideal value of 1.0. The corresponding value for the instrument representing task 2 was 17,3 % and in the paper-based self-assessment 12,5 %.

The results of the study show that the statements and speech samples that belonged to the two top levels of the scale, C1 and C2, were mixed with each other. In some cases, it was also difficult to distinguish them from the statements/speech samples at the B2 level.

The functionality of the instrument representing task 2 was very different from that of the task 1. One reason for this difference may be the brevity of the speech samples in task 2 (an average of 17 seconds). Shortness of samples could have made the evaluators' placement of the samples on the CEFR levels difficult (when they rated the samples before the study) in the same way as it made the learners' comparisons of their performances against these samples difficult in the actual study.

The results showed that the traditional item-test correlation is able to highlight potential problems in the extreme ends of the proficiency scale. Rasch analysis on the other hand was able to detect potential problems in the items in the middle of the scale, as well as to confirm whether there were problems in the easiest and most difficult items situated at the extreme ends of the scale or not. Thus, in this kind of self-assessment study the most suitable combination to analyze the quality and effectiveness of the instruments seems to be Rasch analysis combined with qualitative data analysis.

The different self-assessments used in the study correlated relatively well with each other. The correlations between the situation-specific self-assessment

instrument borrowed from the DIALANG-project and the other self-assessment instruments are shown in Table 32:

TABLE 32 Correspondence of different methods of self-assessment: three highest correlation coefficients.

Method of self-assessment		Correlation coefficient	Remarks
6-point general self-assessment (A1-C2)	Situation-specific self-assessment statements (42 items)	.902	Correlation between two measures borrowed from the DIALANG-project
Speech samples (task 1, 26 items)	Situation-specific self-assessment statements (42 items)	.763	Correlation between a new method of self-assessment and a traditional self-assessment
Speech samples combined (tasks 1 & 2)	Situation-specific self-assessment statements (42 items)	.748	Correlation between a new method of self-assessment and a traditional self-assessment instrument

In general, it should be noted that greater variation in results will produce slightly better correlations if the instrument is longer than e.g. 6-point general self-assessment instrument.

The comparison of the learners' self-assessments with external assessment again produced the result that the situation-specific self-assessment instrument borrowed from DIALANG correlated most strongly with the external assessment than the other types of self-assessment (table 33).

TABLE 33 Correspondences of the self-assessments and external assessments: the three highest correlation coefficients.

Method of self-assessment	External assessment	Correlation coefficient
Situation-specific self-assessment statements (42 items)	Learner's level of speaking, combined (tasks 1 & 2)	.792
Situation-specific self-assessment statements (42 items)	Learner's level of speaking (task 1)	.768
Speech samples (task 1, 26 items)	Learner's level of speaking (task 1)	.753

The good functionality of the situation-specific self-assessment instrument borrowed from DIALANG was to be expected, because it had been empirically validated during the project. The self-assessment method based on speech sam-

ples in task 1 correlated quite well with the external assessment (Table 33, the lowest column). In comparison, the correspondence of the task 2 with the external assessment was weaker (.648), indicating that task 2 did not work as well as task 1. Reasons why task 2 did not work as well as task 1 can be the fact that it was very short (listening to the sample and comparing one's own performance was difficult in this task). In addition, some learners (especially the youngest ones) had some trouble in mastering the politeness and formality of language demanded by the task.

Learners' comments on the functionality of the situation-specific written statements focused on the statements that they found to be problematic. On the other hand, the design of the form partially directed them to pay attention to the statements with which they had difficulties in answering. The statements that were found problematic had the following features: (1) too short or too long statement or (2) the statement contained two sentences, both of which could not be answered either "yes" or "no."

A well-functioning statements, in turn, had the following features: (1) one-part statement, (2) the contents of the statement was related to the person him/herself and /or on his/her life, (3) the statement contained an example of what a word or expression in the statement means, (4) the statement contained a reference to only a limited number (1-2) of linguistic activities or (5) the content of the statement corresponded or was close to the learner's own cultural practices. Some of the learners' comments concerned the difficulty of answering because there were only two answer options ("yes" or "no"), as well as the fact that the statements resembled each other too much. Some learners also experienced difficulties in selecting "yes" or "no" when the statements described situations in which they had never been in a real life.

The study examined briefly typical features in the speech samples in which learners had had trouble answering either "yes" or "no." This was examined by the number of ticks in the "hard to decide" column. As a general observation, it can be noted that the speech samples with most ticks in the "hard to decide" box were located in the middle of the scale. In contrast, the smallest number of "hard to decide" ticks were found for speech samples representing the extreme ends of the scale. An interesting observation is that the samples that the learners found difficult to answer, however, worked well in a statistical sense.

Finally, we explored in this study the features of language that the learners paid attention to when they compared their own oral performances with speech samples. Learners commented by far the most on aspects of speech fluency. The fact that somebody is or is not fluent seems to greatly influence learners' view of a person's oral skills. Learners commented also many features related to pronunciation. When pronunciation issues were commented on in more detail, the comments were often emotionally coloured (*awful pronunciation, but smooth*). Grammar issues were mentioned especially when the speaker of the sample used a wrong word instead of another word. The comments associated with sociolinguistic language mastery were common in task 2, especially



when the sample speaker's skill level was similar to that of the commentator and was located in the middle of the proficiency scale.

### **Implications**

The results show that, at least in the research situation, which is a controlled situation, the majority of language learners succeeded fairly well in assessing their own oral French language skill level, regardless of whether or not they had prior experience in the self-assessment.

If written self-assessment statements are used, it is necessary to draw attention to the fact that one must explain to learners certain principles of skill level assessment. All learners did not seem to have a clear idea of the progression principle, that is, that if a person's oral language skill is located for example, at B2 level, it can be expected that that person is able to master the language situations belonging to the lower levels of proficiency. One must pay a lot of attention to the design of the self-assessment statements, so that the statements are understandable and as unambiguous as possible.

If one wants to use the new self-assessment method in which the performance of the learner is compared to the speech samples, attention has to be paid to the fact that the measuring instruments used are of sufficient quality and feature-rich, yet practical. In a language teaching situation one rarely has a lot of time to design measurement instruments. The pre-testing of the instrument can, of course, help in that only the speech samples representing typical characteristics of a certain skill level of performance are left in the instrument. Comparing one's own performance with the speech samples appears to be of interest to the learners. It seems to be a motivating alternative to self-assessment of the oral language skills. From a pedagogical point of view, learners can practice their language on several linguistic activities with concrete speech samples. Speech samples offer a possibility of doing "something concrete" with the language. Learners can learn a lot by using speech samples not only for pure self-assessment activities but also, for instance, for classifying and analyzing features in learners' own speech samples and in those of their peers or for placing samples onto proficiency levels.

## LÄHTEET JA LYHENTEET

Lyhenteet:

OUP = Oxford University Press

CUP = Cambridge University Press

- Alanen, K. & T. Männikkö. 2012. Suullisen kieli- ja viestintätaidon arvioinnin kehittäminen yliopistotasolla. *Kieliverkoston verkkolehti* 10.10.2012. Saatavissa: <http://www.kieliverkosto.fi/article/suullisen-kieli-ja-viestintataidon-arvioinnin-kehittaminen-yliopistotasolla/> [Viittauspäivä 1.6.2013.]
- Alanen, R. & K. Kajander 2011. Reflektio ja itsearviointi. Opettajan mielistelyä vai kriittistä oman toiminnan arviointia? Teoksessa R. Hildén & O-P. Salo (toim.) *Kielikasvatus tänään ja huomenna Opetussuunnitelmat, opettajankoulutus ja kielenopettajan arki*. Helsinki: WSOYpro, 65–82.
- Alderson, J. C. 2005. *Diagnosing Foreign Language Proficiency. The Interface Between Learning and Assessment*. Norfolk: Biddles Ltd.
- Alderson, J. C. & L.F. Bachman. 2007. Series editors' preface. Teoksessa C. Alderson, J. C. & L. F. Bachman (toim.) *Assessing language through computer technology*. Cambridge: CUP, ix–xi.
- Alderson, J.C. & A. Huhta. 2005. The development of a suite of computer-based diagnostic tests based on the Common European Framework. *Language Testing* 22, 301–320.
- Alderson, J. C. & D. MacIntyre. 2006. Implementing and evaluating a self-assessment mechanism for the web-based Language and Style course. *Language and Literature* 15, 291–306.
- ALTE. 2002. The ALTE Can do project. Saatavissa: [http://www.alte.org/attachments/files/alte\\_cando.pdf](http://www.alte.org/attachments/files/alte_cando.pdf) [Viittauspäivä 26.8.2012.]
- Andrade, H. & Y. Du. 2007. Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, 32, 2, 159–181.
- Araújo, L. 2010. *Computer-based Assessment (CBA) of Foreign Language Speaking Skills*. Luxembourg: Publications Office of the European Union
- Atjonen, P. 2007. *Hyvä, paha arviointi*. Helsinki: Tammi.
- Austin, J. L. 1962. *How to do things with words*. Oxford: OUP.
- Bachman, L. F. 1990. *Fundamental considerations in language testing*. Oxford: OUP.
- Bachman, L.F. 2004. *Statistical Analyses for Language Assessment*. Cambridge: CUP.
- Bachman, L. F. & A.S. Palmer. 1989. The construct validation of self-ratings of communicative language ability. *Language Testing* 6, 1, 14–25.
- Bachman, L. F. & A. S. Palmer. 1996. *Language testing in practice: designing and developing useful language tests*. Oxford: OUP.

- Barbot, M.-J. 1990. Métamorphoses de l'évaluation. L'évaluation dans les systèmes d'auto-apprentissage. Teoksessa L. Porcher (toim.) *Etudes de linguistique appliquée. Evaluation en didactique des langues et des cultures*.
- Baxter Magolda, M.B. 1992. *Knowing and reasoning in college: Gender related patterns in students' intellectual development*. San Francisco, CA: Jossey-Bass.
- Birdsong, T. & W. Sharplin. 1986. Peer evaluation enhances students' critical judgement. *Highway One* 9, 23–28.
- Black, P., C. Harrison, C. Lee, B. Marshall & D. Wiliam. 2003. *Assessment for Learning. Putting it into practice*. Berkshire: Open university press.
- Blanche, P. 1986. *The DLIFLC Study*. CA.: The United States Department of Defence Language Institute, Foreign Language Center (mimeo).
- Blanche, P. 1988. Self-assessment of Foreign Language Skills: Implications for Teachers and Researchers. *RELC Journal* 19, 1, 75–93.
- Blanche, P. & B. J. Merino. 1989. Self-assessment of Foreign Language Skills: Implications for Teachers and Researchers. *Language Learning*. September 39, 3, 313–340.
- Blue, G. M. 1988. Self-assessment: the limits of learner independence. *ELT Documents* 1, 100–118.
- Bond, T. G. & C.M. Fox. 2007. *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Boud, D. 1981. Towards Student Responsibility for Learning. Teoksessa D. Boud (toim.) *Developing Student Autonomy in Learning*. London: Kogan Page, 21–37.
- Boud, D. 1990. Assessment and the promotion of academic values. *Studies in Higher Education* 15, 1, 101–111.
- Boud, D. 1995. *Enhancing Learning through Self Assessment*. London: Kogan Page.
- Brantmeier, C., R. Vanderplank & M. Strube. 2012. What about me? Individual self-assessment by skill and level of language instruction. *System* 40, 1, 144–160.
- Brown, A. 2007. An investigation of the rating process in the IELTS oral interview. Teoksessa L. Taylor & P. Falvey (toim.) *Studies in language testing 19: IELTS collected papers: Research in speaking and writing assessment*. Cambridge: CUP, 98–139.
- Brown, G., A. Anderson, R. Shillcock & G. Yule. 1984. *Teaching Talk. Strategies for production and assessment*. Cambridge: CUP.
- Brown, G. & G. Yule. 1983. *Teaching the spoken language: an approach based on the analysis of conversational English*. Cambridge: CUP.
- Bygate, M. 1987. *Speaking*. Oxford: OUP.
- Campbell, K.S., D. L. Mothersbaugh, C. Brammer & T. Taylor. 2001. Peer versus Self Assessment of Oral Business Presentation Performance. *Business Communication Quarterly* 64, 3, 23–40.
- Canale, M. 1983. On Some Dimensions of Language Proficiency. Teoksessa J. Oller (toim.) *Issues in Language Testing Research*. Rowley, Mass.: Newbury House, 333–342.

- Canale, M. & M. Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1, 1–47.
- Caouette, L. 1990. *Self-Assessment in the Second Language Classroom: A Study of Core French Students' Ability to Assess Their Own proficiency*. M.Ed. Dissertation, University of Alberta, Canada.
- Carroll, J. B. 1961. *Research on Teaching Foreign Languages*. Language Laboratories: Series Preprints and Reprints, vol. III, B. Ann Arbor, University of Michigan.
- Carroll, J. B. 1964. *Language and Thought*. Englewood Cliffs, NJ, Prentice Hall.
- Carter, M.A. 1992. *Self-Assessment using Writing Portfolios*. Ohio State University. Ph.D. Dissertation.
- Carton, F. 1995. L'auto-évaluation. *La lettre de l'ACEDLE* 4. Paris: Association des Chercheurs et Enseignants Didacticiens des Langues Etrangères.
- Celce-Murcia, M., Z. Dörnyei & S. Thurrell. 1995. Communicative Competence: A Pedagogically Motivated Framework with Content Specifications. *Issues in Applied Linguistics* 6, 5–35.
- Chalkia, E. 2012. Self-assessment : an alternative method of assessing speaking skills. *Language testing and assessment issues in the Greek educational context*. Hellenic Open University, School of Humanities. Research Papers in Language Teaching and Learning 3, 1, 225–239.
- Chanquoy, L. 2005. *Statistiques appliquées à la psychologie et aux sciences humaines et sociales*. Paris: Hachette Supérieur.
- Cheng, W. & M. Warren. 2005. Peer assessment of language proficiency. *Language Testing* 22, 1, 93–121.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press.
- Christie, M. & F. Ferdos. 2006. Assessing group-work projects in higher education: Some pedagogical and ethical considerations. Teoksessa M. Christie (toim.) *Shifting perspectives in engineering education*. Chalmers University of Technology, 61–71. Saatavissa: <http://www.ituniv.se/program/ckk/>. [Viittauspäivä 23.8.2012]
- Clark, J. L. D. 1975. Theoretical and technical considerations in oral proficiency testing. Teoksessa R. L. Jones & B. Spolsky (toim.) *Testing language proficiency*. Arlington VA: Center for Applied Linguistics, 10–28.
- Clark, J. L. D. 1979. Direct versus semi-direct tests of speaking ability. Teoksessa E.J. Brière and F.B. Hinofitis (toim.) *Concepts in language testing: some recent studies*. Washington D.C: TESOL, 35–49.
- Clark, J. L. D. 1981. Language. Teoksessa T.S. Barrows, S.M. Ager, M.F. Bennett, H.I. Braun, J.L.D. Clark, L.G. Harris & S.F. Klein (toim.) *College students' knowledge and beliefs: a survey of global understanding*. New Rochelle, NY: Change Magazine Press, 25–35.
- Coombe, C. 1992. *The relationship between self-assessment estimates of functional literacy skills and basic English skills test results in adult refugee ESL learners*. Ph.D. Diss. The Ohio State University.
- Cooper, R. L. 1968. An elaborated language testing model. *Language Learning* 18, 57–73.

- Coste, D., J. Courtyllon, V. Ferenczi, M. Martins-Baltar & E. Papo. 1976. *Un niveau-seuil*. Paris: Hatier.
- Crahay, M. 2006. Dangers, incertitudes et incomplétudes de la logique de la compétence en éducation. *Revue française de pédagogie* 154, 97–110.
- Cram, B. 1995. Self-assessment: from theory to practice. Developing a workshop guide for teachers. Teoksessa G. Brindley (toim.) *Language Assessment in Action*. Research Series 1995, National Centre for English Language Teaching and Research: Macquarie University.
- Cuq, J-P. 2003. *Dictionnaire de didactique du français. Langue étrangère et seconde*. Paris: CLE International.
- Dandenault, E. 1997. *Self-assessment of Communicative Ability: Investigation of a Novel Tool for ESL Learners*. Unpublished MA thesis. Concordia University, Montreal, Canada.
- Deckert, G. D. 1990. Cultural background and self-assessment tendencies in an ESL classroom. Teoksessa V. Bickley *et al* (toim.) *Language Teaching and Learning Styles within and across Cultures*. Hong Kong: Institute of Education.
- Demeuse, M., F. Artus, D. Casanova, A. Crendal, F. Desroches & F. Renaud. 2006. *Evaluation et auto-évaluation en français langue étrangère: approche empirique du Cadre européen commun de référence pour les langues (CECR) du Conseil de l'Europe à travers le TEF et l'échelle d'auto-évaluation DIALANG*. Actes du 19e colloque international de l'Association pour le Développement des Méthodologies d'Evaluation en Education (ADMEE-Europe), Luxembourg.
- De Saint Léger, D. 2009. Self-assessment of speaking Skills and Participation in a Foreign Language Class. *Foreign Language Annals* 42, 1, 158–178.
- De Saint Léger, D. & N. Storch. 2009. Learners' perceptions and attitudes: Implications for willingness to communicate in an L2 classroom. *System* 37,2, 269–285.
- Desroches, F., D. Casanova, A. Crendal, F. Renaud, C. Mahy & M. Demeuse. 2005. *Mise en équivalence des versions papier-crayon et électronique d'un test de français langue étrangère et mise en perspective des résultats par une procédure d'auto-évaluation*. Actes du 18e colloque international de l'Association pour le Développement des Méthodologies d'Evaluation en Education (ADMEE-Europe), Reims.
- DIALANG. Saatavissa: <http://dialang.org>. [Viittauspäivä 21.2.2012]
- Dickinson, L. 1987. *Self-instruction in language learning*. Cambridge: CUP.
- Dochy, F., M. Segers & D. Sluijsmans. 1999. The Use of Self-, Peer- and Coassessment in Higher Education: A Review. *Studies in Higher Education* 24, 3, 331–51.
- Dragemark Oscarson, A. 2009. Self-assessment of Writing in Learning English as a Foreign Language. A Study at the Upper Secondary school level. Göteborg: Acta universitatis Gothoburgensis. Väitöskirja.
- Dweck, C. 1999. *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.

- Eckes, T. 2009. On common ground? How raters perceive scoring criteria in oral proficiency testing. Teoksessa A. Brown & K. Hill (toim.) *Language testing and evaluation 13: Tasks and criteria in performance assessment*. Frankfurt am Main: Peter Lang, 43–73.
- EK. Ks. Elinkeinoelämän keskusliitto.
- Ek, J. A. van. 1975. *The Threshold Level in a European Unit/Credit System for Language Learning by Adults*. Strasbourg: Council of Europe.
- Ek, J.A. van & J.L.M. Trim. 1991. *Threshold level 1990*. Cambridge: CUP.
- Ek, J.A. van & J.L.M. Trim. 1991. *Waystage 1990*. Cambridge: CUP.
- Ek, J.A. van & J.L.M. Trim. 1997. *Vantage Level*. Strasbourg: Council of Europe.
- Ekbatani, G. 2000. Moving towards learner-directed assessment. Teoksessa G. Ekbatani & H. Pierson (toim.) *Learner-directed assessment in ESL*. Mahwah, NJ: Erlbaum, 1–11.
- von Elek, T. 1982. *Test of Swedish as a second language: An experiment in self-assessment* (Working papers from the language teaching research center no 31). Gothenburg University, Language Teaching Research Center.
- 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.
- Elinkeinoelämän keskusliitto EK. 2011. *Oivallus loppuraportti*. Saatavissa: [http://www.ek.fi/ek/fi/tutkimukset\\_julkaisut/2011/5\\_touko/Oivallus-web-v4\\_final.pdf](http://www.ek.fi/ek/fi/tutkimukset_julkaisut/2011/5_touko/Oivallus-web-v4_final.pdf). [Viittauspäivä 2.5.2013.]
- Ellis, R. 2005a. Editorial. *Language Teaching Research* 9, 227–231.
- Ellis, R. 2005b. Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition* 27, 141–172.
- Elsinen, R. 2007. Kielikeskusopiskelijoiden kokemuksia ja käsityksiä kielitaidon itse- ja vertaisarvioinnista. Teoksessa O.-P. Salo, T. Nikula & P. Kalaja (toim.) *Kieli oppimisessa*. AFinLAn vuosikirja 65, 351–369.
- Elsinen, R. & T. Juurakko-Paavola. 2006. *Korkeakouluopiskelijoiden ruotsin kielen taidon arviointi*. Hämeen ammattikorkeakoulun julkaisuja 4/2006. Hämeen ammattikorkeakoulu. Hämeenlinna: Saarijärven Offset Oy.
- Erickson, G. 1996. *Lärare och elever tycker om standardprovet i engelska: Resultat av två enkätundersökningar*. Report from The Language Education Unit, Department of Education. Göteborg: Göteborg University.
- Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen eurooppalainen viitekehys* (EVK). 2004. Helsinki: WSOY.
- European Commission. 2012. *First European Survey on Language Competences*. Final Report. Saatavissa: [http://ec.europa.eu/languages/eslc/docs/en/final-report-escl\\_en.pdf](http://ec.europa.eu/languages/eslc/docs/en/final-report-escl_en.pdf) [Viittauspäivä 23.11.2012]
- Evers, R. 1981. *Self-Evaluation of Proficiency Levels in French as a second Language*. Paper read at the Sixth International Congress of Applied Linguistics in Lund, Sweden.
- EVK. Ks. Eurooppalainen viitekehys.

- Falchikov, N. & D. J. Boud 1989. Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research* 59, 4, 395–430.
- Ferguson, N. 1978. Notes and Discussion: Self-Assessment of Listening Comprehension. *International Review of Applied Linguistics in Language Teaching* 16, 149–156.
- Figueras Casanovas, N. 2001. *Developing oral tests: can we get closer to real life?* Tesi Doctoral. Universitat de Barcelona. Departament de Didàctica de la Llengua i la Literatura.
- First European Survey on Language Competences. Final Report. Ks. European Commission.
- Floropoulou, C. 2002a. *Investigating the Interface of DIALANG*. Unpublished course paper, Department of Linguistics and Modern English Language, Lancaster University.
- Floropoulou, C. 2002b. *Foreign language learners' attitudes to self-assessment and DIALANG: A comparison between Greek and Chinese Learners of English*. Unpublished MA thesis, Department of Linguistics and Modern English Language, Lancaster University.
- Forde, K. 1996. The effects of gender and proficiency on oral self- and peer-assessments. *English Language Studies Working Papers*, City University of Hong Kong 1, 34–47.
- Fulcher, G. 2003. *Testing Second Language Speaking*. Pearson ESL.
- Galaczi, E. 2010. Face-to-face and computer-based assessment of speaking: challenges and opportunities. Teoksessa L. Araújo (toim.) *Computer-based Assessment (CBA) of Foreign Language Speaking Skills*. Luxembourg: Publications Office of the European Union.
- Galloway, V. & A. Labarca. 1990. From Student to Learner: Style, Process and Strategy. Teoksessa D. Birckbichler (toim.) *New Perspectives and New Directions in Foreign Language Education*. The ACTFL Foreign Language Education Series.
- Gipps, C. 1992. *Improving the quality of student learning*. Bristol: Technical and Educational Services Ltd.
- Gipps, C. 1995. *Beyond Testing: Towards a Theory of Educational Assessment*. London: The Falmer Press.
- Gremmo, M.-J. & P. Riley. 1995. Autonomy, self-direction and self access in language teaching and learning: the history of an idea. *System*, 23, 2, 151–164.
- Grice, H.P. 1975. Logic and Conversation. Teoksessa P. Cole & J. Morgan (toim.) *Speech Acts (Syntax and Semantics)*, Volume 3. New York: Academic Press, 41–58.
- 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.
- Halliday, M.A.K. 1970. Language structure and language function. Teoksessa J. Lyons (toim.) *New horizons in linguistics*. Harmondsworth: Penguin.
- Hallitusohjelma 2011 = Pääministeri Jyrki Kataisen hallituksen ohjelma 22.6. 2011.

- Harjanne, P. 2006. "Mut ei tää oo hei midsommarista!" – ruotsin kielenviestinnällinen suullinen harjoittelu yhteistoiminnallisten skeema- ja elaborointitehtävien avulla. Helsinki: Helsingin yliopisto. Väitöskirja.
- Harris, M. 1997. Self-assessment of language learning in formal settings. *ELT journal* 51, 1, 12–20.
- Hattie, J. & H. Timperley. 2007. The Power of Feedback. *Review of Educational Research* 77, 1, 81–112.
- Heikkilä, T. 2001. *Tilastollinen tutkimus*. Helsinki: Edita.
- Heilenman, L.K. 1990. Self-Assessment of second language ability: The role of response effects. *Language Testing* 7, 2, 174–201.
- Helsingin Sanomien verkkolehti. 2013. Suullisen kielitaidon kokeet yotutkintoon 2019. Uutiset kotimaa. 11.6.2013.  
<http://www.hs.fi/kotimaa/Suullisen+kielitaidon+kokeet+yotutkintoon+2019/a1370915068107> [Viittauspäivä 16.6.2013].
- Hildén, R. 2000. *Att tala bra, bättre och bäst*. Reseach Report 217. Helsinki: University of Helsinki, Department of Teacher Education.
- Hildén, R. 2011. Muutosten tarpeet ja mahdollisuudet kielten opetuksessa. Teoksessa R. Hildén & O-P. Salo (toim.) *Kielikasvatus tänään ja huomenna. Kielikasvatus tänään ja huomenna. Opetussuunnitelmat, opettajankoulutus ja kieltenopettajan arki*. Helsinki: WSOYpro, 6–18.
- Hirsjärvi, S. & P. Remes. 1988. *Suomalaisen koulutuksen tulevaisuudenkuvaat*. Keuruu: Otava.
- Hirtt, N. 2010. En Europe, les compétences contre le savoir. *Le Monde diplomatique*, 22–23.
- Hirvonen, P. 1974. *Englannin kielen taidon mittaaminen lukion päättyessä*. Suomen sovelletun kielitieteen yhdistyksen (AFinLA) julkaisuja. Turun yliopisto. Väitöskirjatutkimus.
- Holec, H. 1979. *Autonomy and Foreign Language Learning*. Strasbourg: Council of Europe.
- Holec, H. 1981. Arguing for Self-Evaluation. *Français dans le Monde* 21, 15– 23.
- Holec, H. 1982. *Autonomie et apprentissage des langues étrangères*. Paris: Hatier.
- Holec, H. (toim.). 1988. *Autonomy and self-directed learning: Present fields of application*. Strasbourg: Council of Europe, Council for Cultural Co-operation.
- Holec, H. 1990. Apprendre à l'apprenant à s'évaluer : Quelques pistes à suivre. *Etudes de Linguistique Appliquée* 79.
- Huhta, A. 1993. Suullisen kielitaidon arviointi. Teoksessa S. Takala (toim.) *Suullinen kielitaito ja sen arviointi*. Jyväskylä: Jyväskylän yliopisto, Kasvatustieteiden tutkimuslaitos, 143–226.
- Huhta, A. 2003. Itsearviointi kielitaidon arviointimuotona. Teoksessa A. Airola (toim.) *Kokemuksia suullisen kielitaidon arviointihankkeesta*. Joensuu: Pohjois-Karjalan ammattikorkeakoulu, 20–37.
- Huhta, A. 2007. Itsearviointi osana kielitestiä – tutkimus käyttäjien suhtautumisesta tietokoneen antamaan palautteeseen kielitaidon itsearvioinnista. Teoksessa O-P. Salo, P. Kalaja & T. Nikula (toim.) *Kieli oppimisessa*. AFInLA yearbook. Publications of the Association Finlandaise de Linguistique Appliquée 65, 371–388.



- Huhta, A. 2008. Diagnostic and formative assessment. Teoksessa B. Spolsky & F. Hult (toim.) *Handbook of Educational Linguistics*. Blackwell, 469–482.
- Huhta, A. 2010. *Innovations in diagnostic assessment and feedback: an analysis of the usefulness of the Dialang language assessment system*. Jyväskylän Yliopisto. Väitöskirja.
- Huhta, A. (tulossa). *Concept-mapping – an approach to understanding the concept of self-assessment*.
- Huhta, A. & Figueras, N. 2004. Using the CEF to promote language learning through diagnostic testing. Teoksessa K. Morrow (toim.) *Insights from the Common European Framework*. Oxford: OUP, 65–76.
- Huhta, A. & T. Suontausta. 1993. Suullisen kielitaidon testausmenetelmiä. Teoksessa S. Takala (toim.) *Suullinen kielitaito ja sen arviointi*. Jyväskylä: Jyväskylän yliopisto, Kasvatustieteiden tutkimuslaitos,, 227–266.
- Huhta, A. & M. Tarnanen. 2009. Assessment practices in the Finnish comprehensive school - what is the students' role in them? Teoksessa S. May (toim.) *LED2007: Refereed conference proceedings of the 2nd International Conference on Language, Education and Diversity*. Hamilton, New Zealand: Wilf Malcolm Institute of Educational Research (WMIER), University of Waikato.
- Huttunen, I. 1990. *Towards learner autonomy and increased communication*. Oulu: Reports from the Faculty of Education, University of Oulu 73.
- Huttunen, I. & J. Huttunen. 1998. *Vieraat kielet: Oppiminen, opettaminen, arviointi yleiseurooppalainen viitekehys luonnos 2 ehdotuksesta viitekehyykseksi*. Oulu: Oulun Yliopistopaino.
- Huttunen, I. & H. Jaakkola. 2003. *Vieraiden kielten yhteinen eurooppalainen viitekehys: oppiminen, opetus ja arviointi*. Esitelmä professori Tellan jatkokoulutusseminaarissa 14.4.2003. Helsinki.
- Huusko, M. 2008. Itsearviointi kehittävän arvioinnin menetelmänä. Teoksessa E. Korkeakoski & H. Silvennoinen (toim.) *Avaimia koulutuksen arvioinnin kehittämiseen*. Koulutuksen arviointineuvoston julkaisuja 31. Jyväskylä: Jyväskylän yliopistopaino, 127–138.
- Hymes, D. 1972. On communicative competence. Teoksessa J.B. Pride & J. Holmes (toim.) *Sociolinguistics*. Harmondsworth: Penguin.
- Iwashita, N., A. Brown, T. McNamara & S. O'Hagan. 2008. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 1, 24–49.
- Jaakkola, H., U. Viita-Leskelä, S.-L. Sävy & K. Komsu. 2002. How to promote learning to learn in first foreign language classes in Finland. Teoksessa *Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Case Studies*. Strasbourg: Council of Europe.
- Janssen-van Dielen, A-M. 1989. The development of a test of Dutch as a second language: the validity of self-assessment by inexperienced subjects. *Language Testing* 6, 1–13.
- Janssen-van Dielen, A-M. 1992. *Zelfbeoordeling en tweede-taalleren ("Self-Assessment in second language learning")*. Ph.D. Dissertation. Katholieke Universiteit, Nijmegen, The Netherlands.

- Juurakko-Paavola, T. 2005. Ammattikorkeakouluopiskelijoiden ruotsin kielen kirjallinen taito. Teoksessa T. Juurakko-Paavola (toim.) *Ammattikorkeakouluopiskelijoiden kielitaito ja sen arviointikäytännöt*. Hämeen ammattikorkeakoulun julkaisuja 1/2005. Hämeenlinna: Saarijärven Offset Oy, 199–270.
- Juurakko, T. & S. Takala. 2001. Ammattikorkeakouluopiskelijoiden käsityksiä AMKKIA-pilottikokeista. Teoksessa T. Juurakko (toim.) *Kielitaidon arvioinnin yhdenmukaistaminen ammattikorkeakouluissa (AMKKIA)*. Hämeenlinna: Hämeen ammattikorkeakoulu, 153–168.
- Kaftandjieva, F. & S. Takala. 2002. Council of Europe scales of language proficiency: a validation study. Teoksessa *Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Case Studies*. Strasbourg: Council of Europe Publishing.
- Kahn, S. 2012. Intentions des politiques et réalités du terrain en Belgique. Teoksessa J.L. Villeneuve (toim.) *Le socle commun en France et ailleurs*. Paris: Le Manuscrit, 111–112.
- Kantelinen, R. & R. Hildén. 2012. Eurooppalainen kielisalkku vihdoinkin suomalaisen perusopetukseen! *Kieliverkoston verkkolehti* 10.10.2012. Saatavissa: <http://www.kieliverkosto.fi/article/eurooppalainen-kielisalkku-vihdoin-suomalaiseen-perusopetukseen/> [Viittauspäivä 1.6.2013.]
- Kenyon, D. M & V. Malabonga 2001. Comparing examinee attitudes toward computer-assisted and other oral proficiency assessments. *Language Learning and Technology* 5, 2, 60–83.
- Kenyon, D. M. & M. Malone. 2010. Investigating examinee autonomy in a computerized test of oral proficiency. Teoksessa L. Araújo (toim.) *Computer-based Assessment (CBA) of Foreign Language Speaking Skills*. Luxembourg: Publications Office of the European Union.
- Kieli ja sen kielioipit*. 1994. Opetusministeriö. Helsinki: Painatuskeskus Oy.
- King, P.M. & K.S. Kitchener. 2004. Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist* 39, 5–18.
- Knoch, U. 2010. Investigating the effectiveness of individualized feedback to rating behavior – a longitudinal study. *Language Testing* 28, 2, 179–200.
- Kohonen, V. 1994. Salkkuarvioinnin mahdollisuuksista opettajan ammatillisen kehittymisen tukena. Teoksessa P. Linnakylä, P. Pollari & S. Takala (toim.) *Portfolio arvioinnin ja oppimisen tukena*. Jyväskylä: Jyväskylän yliopisto, kasvatustieteiden tutkimuslaitos, 33–50.
- Kohonen, V. 1997. Authentic assessment as an integration of language learning, teaching, evaluation and the teacher's professional growth. Teoksessa A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (toim.) *Current developments and alternatives in language assessment*. Proceedings of the LTRC 96. Jyväskylä, University of Jyväskylä: Centre for Applied Language Studies, 7–22.
- Kohonen, V. 1999. Authentic assessment. Teoksessa J. Arnold (toim.) *Affect in language learning*. Cambridge: CUP, 279–284.

- Kohonen, V. n.d. *Agency and autonomy through reflection and self/peer-assessment in FL education*. University of Tampere, Department of Teacher Education.
- Saatavissa: <<http://learnerautonomy.org/ViljoKohonen.pdf>> [Viittauspäivä 5.8.2012.]
- Kohonen, V. 2000. Eurooppalainen kielisalkku kielenopetuksen kehittämissä: suuntaviittoa ja mahdollisuuksia. Teoksessa V. Kohonen & U. Pajukanta (toim.) *Eurooppalainen kielisalkku – kokemuksia EKS-projektin alkutai-paleelta*. Tampereen yliopiston opettajankoulutuslaitoksen julkaisuja A23/2000, 9–44.
- Krokkfors, L. & S. Patrikainen. Hyvä, paha opetus. Teoksessa J. Husu & R. Jyrhämä (toim.) *Suoraa puhetta. Kollegiaalisesti opetuksesta ja kasoitukselta*. Juva: PS-kustannus, 29–45.
- Lado, R. 1961. *Language Testing: The Construction and Use of Foreign Language Tests*. London: Longman.
- Lammi, K. 2000. Reflektio ja sen opettaminen. Teoksessa V. Kohonen & U. Pajukanta (toim.) *Eurooppalainen kielisalkku – kokemuksia EKS-projektin alkutai-paleelta*. Tampere: Tampereen Yliopisto, opettajankoulutuslaitos, 161–173.
- Language learning for European citizenship*. 1997. Final report of the Project Group (1989–96). Strasbourg: Council of Europe Publishing.
- Lantolf, J. & S. Thorne. 2006. *Sociocultural theory and the genesis of second language development*. Oxford: OUP.
- Latomaa, S. 1996. Self-Reported second language proficiency among four immigrant groups in the Nordic Countries. *Jyväskylä Cross Language Studies* 17, 169–178.
- Laval, C., F. Vergne, P. Clément & G. Dreux. 2011. *La nouvelle école capitaliste*. Paris: La Découverte.
- LeBlanc, R. & G. Painchaud. 1985. Self-assessment as a second language placement instrument. *TESOL Quarterly* 19, 4, 673–87.
- Lee-Ellis, S. 2009. The development and validation of a Korean C-Test using Rasch Analysis. *Language Testing*, 26, 2, 245–274.
- Lewkowicz, J.A. & J. Moon. 1985. Evaluation, A Way of Involving the Learner. Teoksessa J. C. Alderson (toim.) *Lancaster Practical Papers in English Language Education*. Vol. 6, Evaluation. Oxford: Pergamon Press, 45–80.
- Linacre J.M. 2002. What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions* 16:2. Saatavissa: <<http://www.rasch.org/rmt/rmt162f.htm>> [Viittauspäivä 15.5.2013.]
- Linacre, J. M. 2011. *A User's Guide to FACETS. Rasch Model Computer Program*.
- Lindblom-Ylänne, S., H. Pihlajamäki & T. Kotkas. 2006. Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education* 7, 51–62.
- Little, D. 1991. *Learner Autonomy 1: Definitions, Issues and Problems*. Dublin: Authentik.
- Little, D. 2005. The Common European Framework and the European

- Language Portfolio: involving learners and their judgements in the assessment process. *Language Testing*, 22, 3, 321–336.
- Little, D. 2009. Language learner autonomy and the European language portfolio: two L2 English examples. *Language Teaching: Surveys and Studies* 42, 222–233.
- LOPS 1994. Ks. Lukion opetussuunnitelman perusteet.
- LOPS 2003. Lukion opetussuunnitelman perusteet. Saatavissa: [http://www.oph.fi/saadokset\\_ja\\_ohjeet/opetussuunnitelmien\\_ja\\_tutkintojen\\_perusteet/lukiokoulutus](http://www.oph.fi/saadokset_ja_ohjeet/opetussuunnitelmien_ja_tutkintojen_perusteet/lukiokoulutus). [Viittauspäivä 19.12.2012]
- Lucas U. & P. L. Tan. 2013. Developing a capacity to engage in critical reflection: students' 'ways of knowing' within an undergraduate business and accounting programme. *Studies in Higher Education*, 38, 1, 104–123.
- Lukiokoulutuksen suullisen kielitaidon arviointityöryhmän muistio. Työryhmämuistioita ja selvityksiä 2006:26. Helsinki: Opetusministeriö.
- Lukion opetussuunnitelman perusteet. 1994. Helsinki: Opetushallitus.
- Luoma, S. 2004. *Assessing speaking*. Cambridge Language Assessment Series. Cambridge: CUP.
- Luoma, S. & M. Tarnanen. 2003. Creating a self-rating instrument for second language writing: from idea to implementation. *Language Testing*, 20, 4, 440–465.
- Lussier, D. & C. E. Turner. 1995. *Le point sur l'évaluation en didactique des langues*. Anjou (Québec): Centre éducatif et culturel Inc.
- Luukka, M.-R., S. Pöyhönen, A. Huhta, P. Taalas, M. Tarnanen & A. Keränen. 2008. *Maailma muuttuu – mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytännöt koulussa ja vapaa-ajalla*. Jyväskylän yliopisto: Soveltavan kielentutkimuksen keskus.
- Luukka, M.-R. & S. Pöyhönen. 2007. Kielikoulutuspoliittisen projektin keskeiset suositukset. Teoksessa S. Pöyhönen & M.-R. Luukka (toim.) *Kohti tulevaisuuden kielikoulutusta. Kielikoulutuspoliittisen projektin loppuraportti*. Jyväskylä: Jyväskylän Yliopisto, soveltavan kielentutkimuksen keskus, 453–479.
- 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 & R. C. Gardner. 1994. The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning* 44, 283–305.
- 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.
- Malabonga, V., D. M. Kenyon & H. Carpenter 2005. Self-assessment, preparation and response time on a computerized oral proficiency test. *Language Testing* 22, 1, 59–92.
- Matsuno, S. 2009. Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26, 1, 75–100.

- MacDonald, B. & D. Boud. 2003. The Impact of Self-assessment on Achievement: the effects of self-assessment training on performance in external examinations. *Assessment in Education*, 10, 2, 209–220.
- McNamara, T. 1996. *Measuring Second Language Performance*. Harlow, Essex: Addison Wesley Longman.
- Meisalo, V., E. Sutinen & J. Tarhio. 2003. *Modernit oppimisympäristöt. Tieto- ja viestintäteknikka opetuksen ja opiskelun tukena*. Pieksämäki: Tietosanoma Oy.
- Mezirow, J. 1991. *Transformative dimensions of adult learning*. San Francisco, CA: Jossey-Bass.
- Miller, L. & R. Ng. 1994. *Peer assessment of oral language proficiency*. Perspective: working papers on the department of English, City polytechnic of Hong Kong 6, 41–56.
- Miller, L. & R. Ng. 1996. Autonomy in the classroom: peer assessment. 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, 133–146.
- Moirand, S. 1982. *Enseigner à communiquer en langue étrangère*. Hachette.
- Mondock, S.L. 1997. Portfolios--The Story behind the Story. *The English Journal* 86, 1. National Council of Teachers of English.
- Mothe, J. C. 1990. Evaluer les compétences en langue étrangère. *Etudes de linguistique appliquée*. Oct-Décembre. L'évaluation en didactique des langues des cultures.
- Murau, A. 1993. Shared writing: students' perceptions and attitudes of peer review. *Working Papers in Educational Linguistics* 9, 71–79.
- North, B. 1992. *Transparency and coherence in language learning in Europe*. Strasbourg: Council of Europe.
- North, B. 1995. The Development of a Common Framework Scale of Descriptors of Language Proficiency based on a Theory of Measurement. *System* 23, 445–465.
- North, B. 2002. Developing descriptor scales of language proficiency for the CEF common reference levels. In *Common European Framework of Reference for Languages: Learning, Teaching and Assessment. Case Studies*. Strasbourg: Council of Europe Publishing.
- North, B. & G. Schneider. 1998. Scaling descriptors for language proficiency scales. *Language Testing* 15, 217–262.
- Nunan, D. 1988. *The learner-centred curriculum: a study in second language teaching*. Cambridge: CUP.
- Ockey, G. 2011. Self-consciousness and Assertiveness as Explanatory Variables of L2 Oral Ability: A Latent Variable Approach. *Language Learning*, 61, 3, 968–989.
- Orsmond, P., S. Merry & K. Reiling. 1997. A Study in Self-Assessment: Tutor and Students' Perceptions of Performance Criteria. *Assessment & Evaluation in Higher Education* 22, 357–67.
- Orsmond, P., S. Merry & K. Reiling. 2000. The Use of Student Derived Marking Criteria in Peer and Self Assessment. *Assessment and Evaluation in Higher Education* 25, 1, 23–38.

- Oscarson, M. 1978. *Approaches to Self-assessment in Foreign Language Learning*. Strasbourg: Council of Europe.
- Oscarson, M. 1980. *Approaches to self-assessment in foreign language learning*. Oxford: Pergamon Press.
- Oscarson, M. 1981. Subjective and Objective Assessment of Foreign Language Performance. Teoksessa J. A. S. Read (toim.) *Directions in Language Testing*. Singapore: University Press, 225–239.
- Oscarson, M. 1984. *Self-Assessment of Foreign Language Skills: A Survey of Research and Development Work*. Strasbourg: Council of Europe, Council for Cultural Co-operation.
- 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. 1999. Estimating Language Ability by Self-Assessment – A review of some of the issues. In *Papers on Language – Learning, Teaching, Assessment. Festschrift till Torsten Lindblad*. Rapport från Institutionen för pedagogik och didaktik 1999/02. Göteborg : Göteborgs universitet.
- Oxford, R. 1990. *Language Learning Strategies. What every teacher should know*. New York: Newbury House.
- Patri, M. 2002. The influence of peer feedback on self- and peer-assessment of oral skills. *Language Testing* 19, 109–131.
- Peirce, B. N., M. Swain & D. Hart. 1993. Self-Assessment, French Immersion, and locus of Control. *Applied Linguistics* 14, 25–42.
- Perrenoud, P. 2011. *Quand l'école prétend préparer à la vie : Développer des compétences ou enseigner d'autres savoirs?* Issy-les-Moulineaux : ESF.
- Poehner, M.E. & J.P. Lantolf. 2005. Dynamic assessment in the language classroom. *Language Teaching Research* 9, 233–265.
- Pollitt, A. & N. L. Murray. 1996. What raters really pay attention to. Teoksessa M. Milanovic & N. Saville (toim.) *Studies in language testing 3: Performance testing, cognition and assessment*. Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem. Cambridge: CUP, 74–91.
- POPS 1994. Ks. Peruskoulun opetussuunnitelman perusteet. 1994. Helsinki: Opetushallitus.
- POPS 2004. Perusopetuksen opetussuunnitelman perusteet. Saatavissa: [http://www.oph.fi/download/139848\\_pops\\_web.pdf](http://www.oph.fi/download/139848_pops_web.pdf). [Viittauspäivä 19.12.2012]
- Rainey, K. 1990. Teaching technical writing to non-native speakers. *Technical Writing Teacher* 17, 131–135.
- Rasch, G. 1979. Probabilistic Models for Some Intelligence and Achievement Tests. Danish Institute for Educational Research, Copenhagen.
- Roever, C. 2001. Web-based Language Testing. *Language Learning and Technology* 5, 2, 84–94.

- Ross, S. 1998. Self-assessment in second language testing: a meta-analysis and analysis of experiential factors. *Language Testing* 15, 1, 1–20.
- Räisänen, A. 2005. Kehittävään arviointiin. Teoksessa H.K. Lyytinen & A. Räisänen (toim.) *Kehittämissuuntaa arvioinnista*. Jyväskylä: Koulutuksen arviointineuvosto 6, 109–127.
- Saari, S. 2002. *Opettajankoulutuksen arviointi- ja kehittämisdiskurssi koulutuspoliittisessa kontekstissa*. Tampere: Acta universitatis tamperensis 893.
- Saleva, M. 1997. *Now They're Talking. Testing Oral Proficiency in a Language Laboratory*. Studia Philologica Jyväskyläensia 43. Jyväskylä: University of Jyväskylä. Väitöskirjatutkimus.
- Salo, O-P. & R. Hildén. 2011. Hiljaa hyvä tulee – vai tuleeko? Kieltenopettajien käsityksiä opiskelija-autonomian edistämisestä yleissivistävässä koulutuksessa. Teoksessa R. Hildén & O-P. Salo (toim.) *Kielikasvatus tänään ja huomenna. Opetussuunnitelmat, opettajankoulutus ja kielenopettajan arki*. Helsinki: WSOYpro, 19–40.
- Sato, T. 2011. The contribution of test-takers' speech content to scores on an English oral proficiency test. *Language Testing*, 29, 2, 223–241.
- Savignon, S.J. 1983. *Communicative Competence: Theory and Classroom Practice. Texts and Contexts in Second Language Learning*. Reading: Addison-Wesley Publishing Co.
- Schachter, J. 1990. Communicative competence revisited. Teoksessa B. Harley, P. Alien, J. Cummins & M. Swain (toim.) *The development of second language proficiency*. New York: CUP.
- Schärer, R. 1983. Identification of learner needs at Eurocentre. Teoksessa R. Richterich (toim.), *Case studies in identifying language needs*. Oxford: Pergamon Press, 106–116.
- Searle, J. R. 1969. *Speech Acts. An Essay in the Philosophy of Language*. Cambridge: CUP.
- Shepard, L. A. 1994. *The challenges of assessing young children appropriately*. Saatavissa: [http://www.cse.ucla.edu/products/parents/cresst\\_challengeshepard.pdf](http://www.cse.ucla.edu/products/parents/cresst_challengeshepard.pdf). [Viittauspäivä 18.12.2012]
- Shrauger, J. S. & T. M. Osberg. 1981. The Relative Accuracy of Self-Predictions and Judgements by Others in Psychological Assessment. *Psychological Bulletin* 90, 2, 322–351.
- 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.
- Smith, K. 1997. *Self-Assessment as an Alternative Assessment Approach in the Learning of English as a Foreign Language in Israel*. Ph.D. Dissertation. Bar-Ilan University, Israel.
- Smith K. & R.B. Baldauf. 1982. The concurrent validity of self-rating with interviewer rating on the Australian Second Language Proficiency Scale. *Educational and Psychological Measurement* 42, 4, 1117–1124.
- Spolsky, B. 1995. *Measured Words. The development of objective language testing*. Oxford: OUP.

- Springer, C. 1996. *La didactique des langues face aux défis de la formation de adultes*. Collection Autoformation et Enseignement Multimédia. Paris: OPHRYS.
- Stiggins, R. 1991. Assessment literacy. *Phi Delta Kappan*, March, 534–539.
- Strong-Krause, D. 1997. *How Effective Is Self-Assessment for ESL Placement?* Paper presented at TESOL '97, Orlando, Florida, March 1997.
- Suomen Lukiolaisten Liitto ry. 2005. Kannanotto 14.2.2005. Suulliset kokeet ylioppilastutkinnon seuraava uudistus.
- Takala, S. 1973. *Kielitaidon mittaaminen*. Pops-70, opas 3 b. Kouluhallitus, kokeilu- ja tutkimustoimisto. Helsinki: Valtion painatuskeskus.
- Takala, S. 1977. Oppilaiden näkemyksiä kielten opiskelusta. Teoksessa S. Takala (toim.) *Piirteitä vieraiden kielten opetuksesta*. Jyväskylä: Jyväskylän yliopisto, kasvatustieteiden tutkimuslaitos, 63–65.
- Takala, S. 1993. Suullisen kielitaidon kokeen työryhmän (SUKO) raportti. Teoksessa S. Takala (toim.) *Suullinen kielitaito ja sen arviointi*. Jyväskylä: Jyväskylän yliopisto, kasvatustieteiden tutkimuslaitos, 35–61.
- Takala, S. 1994. Arviointi – ongelma ja mahdollisuus. Teoksessa P. Linnakylä, P. Pollari & S. Takala (toim.) *Portfolio arvioinnin ja oppimisen tukena*. Jyväskylä: Jyväskylän yliopisto, kasvatustieteiden tutkimuslaitos, 1–8.
- Tarnanen, M. & A. Huhta. 2011 Foreign language assessment and feedback practices in Finland. Teoksessa D. Tsagari & I. Csepes (toim.) *Classroom-Based Language Assessment* (Language Testing and Evaluation Series, R. Grotjahn, & G. Sigott (general eds). Frankfurt am Main: Peter Lang, 129–146.
- Tarnanen, M., A. Huhta & K. Pohjala. 2007. Mitä on osaaminen? Kielitaidon arviointi vastaajana. Teoksessa S. Pöyhönen & M-R Luukka (toim.) *Kohti tulevaisuuden kielikoulutusta. Kielikoulutuspoliittisen projektin loppuraportti*. Jyväskylä: Jyväskylän Yliopisto, soveltavan kielitutkimuksen keskus, 381–412.
- Taylor, L. 2011. *Examining Speaking: Research and Practice in Assessing Second Language Speaking*. Cambridge: CUP.
- 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.
- Threshold Level. Ks. Ek, J.A. van & J.L.M. Trim ja Ek, J.A. van.
- Topping, K. 1998. Peer assessment between students in colleges and universities. *Review of Educational Research* 68, 249–276.
- Trim, J. L. M. 1980. *Developing a Unit/Credit Scheme of Adult Language Learning*. Strasbourg: Council of Europe.
- Trim, J.L.M. 2001. *The Work of the Council of Europe in the field of Modern Languages, 1957–2001*. European Day of Languages at the premises of the European Centre for Modern Languages, Graz.
- Tudor, I. 1996. *Learner-centredness as language education*. University of Cambridge Local Examination Syndicate 1998. The multilingual glossary



- of language testing terms. *Studies in Language Testing* 6. Cambridge: CUP.
- Tudor, I. & F. Nivelles. 1991. Assessing Self-Assessment: the Case for an Integrative Profiling of L2 Learners' Strategic Abilities. *Rapport d'Activités de l'Institut phonétique* 27, Université Libre de Bruxelles, 67–90.
- Un niveau-seuil. Ks. D. Coste, J. Courtillon, V. Ferenczi, M. Martins-Baltar & E. Papo.
- Underhill, N. 1987. *Testing spoken language: a handbook of oral testing techniques*. Cambridge: CUP.
- Université Laval 1996. Apprendre au 21e siècle. Enoncé de perspectives sur les apprenantes et les apprenants. Saatavissa: <<http://www.tact.fse.ulaval.ca/fr/html/vision2.html>> [Viittauspäivä 31.7.2012.]
- Uusikylä, K. & P. Atjonen 2005. *Didaktiikan perusteet*. Porvoo: WSOY.
- Valtionhallinnon kielitutkintojen perusteet 2005. Helsinki: Opetushallitus.
- Van Moere, A. 2010. Automated spoken language testing: test construction and scoring model development. Teoksessa L. Araújo (toim.) *Computer-based Assessment (CBA) of Foreign Language Speaking Skills*. Luxembourg: Publications Office of the European Union.
- Van Moere, A. 2012. A psycholinguistic approach to oral language assessment. *Language Testing*, 29, 3., 325–344.
- Vantage Level. Ks. Ek, J.A. van & Trim, J.L.M.
- Vygotski, L.S. 1982. *Ajattelu ja kieli*. Espoo: Weilin Göös.
- Waystage. Ks. Ek, J.A. van & Trim, J.L.M.
- Weir, C. 1990. *Communicative language testing*. Prentice Hall International.
- Weir, C. 1993. *Understanding and Developing Language Tests*. Prentice Hall Europe.
- Wenden, A. 1991. *Learner Strategies for Learner Autonomy*. Hemel Hempstead: Prentice Hall.
- Widdowson, H.G. 1983. *Learning purpose and language use*. Oxford: OUP.
- Wilkins, D. A. 1973. *Second-language learning and teaching*. London: Edward Arnold.
- Wilkins, D. A. 1976. *National Syllabuses*. Oxford: OUP.
- Willing, K. 1989. *Teaching how to learn*. NCELTR. Macquarie University.
- Wilson, K. M. 1996. *Validity of Global Self-Ratings of ESL Speaking Proficiency Based on an FSI/ILR-Referenced Scale: An Empirical Assessment*. In collaboration with R. Lindsay. Princeton, NJ: Educational Testing Service.
- [www.winsteps.com/winman/index.htm?diagnosingmisfit.htm](http://www.winsteps.com/winman/index.htm?diagnosingmisfit.htm) 2012. A User's Guide to Winsteps. Rasch-Model Computer Programs. [Viittauspäivä 27.10.2012.]
- Wood, R. 1991. *Assessment and testing : a survey of research commissioned by the University of Cambridge Local Examinations Syndicate*. Cambridge: CUP.
- Yang, R. 2003. *Investigating how test-takers use the DIALANG feedback*. Unpublished MA thesis, Department of Linguistics and Modern English Language, Lancaster University.
- Yleisten kielitutkintojen perusteet 2011. Helsinki: Opetushallitus.

- Yli-Renko, K. 1985. *Lukion saksan kielen opetuksen tavoitteet*. Helsinki: Helsingin yliopiston opettajankoulutuslaitos. Väitöskirjatutkimus.
- Yli-Renko, K. 1989. *Suullinen kielitaito ja sen mittaaminen lukion päättövaiheessa*. Helsingin yliopisto. Opettajankoulutuslaitos.
- Yli-Renko, K. 1991. Suullisen kielitaidon oppiminen lukiossa: oppilaiden näkökulma. Teoksessa K. Yli-renko & L. Salo-Lee (toim.) *Oral communication proficiency and its learning in senior secondary school*. Turku: University of Turku. Department of Education 147, 25–58.
- Zareva, A. 2005. What is New in the New TOEFL-iBT 2006 Test Format? *Electronic Journal of Foreign Language Teaching* 2, 2, 45–57.

### **1. Lait, asetukset, määräykset ja päätökset**

- Asetus suomen ja ruotsin kielen taidon osoittamiseksi suoritettavista tutkinnoista 313/1922.
- Asetus suomen ja ruotsin kielen taidon osoittamisesta 442/1987.
- Laki valtion virkamiehiltä vaadittavasta kielitaidosta 149/1922.
- Lukiolaki 629/1998.
- Opetushallituksen määräys 10/011/2009.
- Perusopetuslaki 628/1998.
- Päätös sisältävä johtosäännön suomen ja ruotsin kielessä suoritettavia tutkintoja varten asetetuille tutkintolautakunnille 82/1923.
- SK 82/1923. Ks. Päätös sisältävä johtosäännön suomen ja ruotsin kielessä suoritettavia tutkintoja varten asetetuille tutkintolautakunnille.

**LIITE 1      Tehtävä 1****Tehtävä 1: Kuvasta kertominen**

Kuvaile alla olevaa henkilöä mahdollisimman yksityiskohtaisesti.

Voit kertoa henkilöön liittyen esimerkiksi seuraavia asioita:

- mitä hänellä on päällään (vaatteiden värit, materiaalit)
- mikä aika päivästä on
- mihin hän on menossa
- mistä hän on tulossa
- missä hän työskentelee
- minkä tyyppisestä ihmisestä on kysymys
- mistä hän pitää tai ei pidä
- mitä harrastaa (jne. käytä vapaasti mielikuvitustasi!)

**Tehtävän valmisteluaika:** 1,5 minuuttia

**Puhe aika:** valmistelun jälkeen korkeintaan 1,5 minuuttia



**LIITE 2      Tehtävä 2**

**Tehtävä 2: Tilanteessa reagointi**

Lue ensin alla oleva kuvaus tilanteesta.

Tilannekuvaus (eli mitä on jo tapahtunut):

Olet hakenut asemalta ranskalaisen tärkeän vieraan, Julien Fabre-Perrault'n. Olette jo esittäytyneet toisillenne. Nyt olette palanneet työpaikallasi.

Suorita nyt viivan alla oleva tehtävä.

**Tehtävän valmistelu-aika:** 40 sekuntia

**Puhe-aika:** valmistelun jälkeen 30 sekuntia

---

**Suoritettava tehtävä:**

Esittele toisilleen työtoverisi Kristiina Saukkonen-Riipinen ja juuri saapunut vieras.  
Kerro, että työtoverisi puhuu hyvin ranskaa.

**LIITE 3 Osallistujan taustatiedot****OSALLISTUJAN TAUSTATIEDOT**

Tähän tutkimukseen osallistuvien henkilöiden henkilö-, tausta- ja suoritustietoja käsitellään luottamuksellisesti. Tutkimuksen aikana tehtäviä suorituksia ja niiden arviointeja käsitellään nimettöminä.

Osallistujanumero \_\_\_\_\_

Sukunimi \_\_\_\_\_

Etunimi \_\_\_\_\_

Osoite \_\_\_\_\_

Sähköposti \_\_\_\_\_

Oppilaitos, missä opiskelen \_\_\_\_\_

Ikä \_\_\_\_\_ v.

**Vastaa alla oleviin kohtiin suluissa ilmoitetulla tavalla :**

1. **Äidinkielenesi** (Rastita)  suomi  ruotsi

2. **Kuinka monta vuotta olet opiskellut ranskan kieltä (luvut vuosia) ?** (Ympyröi)

peruskoulussa	1	2	3	4	5	6	enemmän
lukiossa	1	2	3	4			
yliopistossa	1	2	3	4	5	6	enemmän
muualla	1	2	3	4	5	6	enemmän

(missä ? \_\_\_\_\_)

3. **Minkä ikäisenä aloitit ranskan kielen opiskelun ?** \_\_\_\_\_

4. **Kuinka olet nykyisin tekemisissä ranskan kielen kanssa ?** (Rastita)

en mitenkään

puhun ranskaa

kirjoitan ranskaksi (esim. kirjeitä)

kuuntelen ranskaa (esim. televisiosta)  luen ranskaksi

5. **Kuinka usein käytät ranskan kieltä (työssä, opinnoissa, vapaa-aikana) ?**  
(Rastita)

- |   |  |
|---|--|
| <input type="checkbox"/> (melkein) päivittäin     | <input type="checkbox"/> kerran kahdessa viikossa    |
| <input type="checkbox"/> muutaman kerran viikossa | <input type="checkbox"/> kerran kuussa tai harvemmin |

6. **Oletko aikaisemmin tehnyt itsearviointia ?** (Rastita ja/tai vastaa)

- en
- kyllä, mihin itsearviointi liittyi (esim. jonkun kielen arviointiin, omiin työskentelytapoihin)?

---

---

**Kiitos !**

## LIITE 4 Puhumisen itsearviointilomake

### PUHUMISEN ITSEARVIOINTILOMAKE

Osallistujanumero\_\_\_\_\_

Nimi\_\_\_\_\_

Oppilaitos\_\_\_\_\_

Arvioi ensin oma **ranskan** kielen puhumisen taitosi. Valitse **yksi** alla olevasta kuudesta puhumisen tasokuvauksesta ; valitse siis se kuvaus, joka kuvaa **parhaiten** sinun taitoasi puhua ranskaa. Merkitse valitsemasi kuvaus rastilla (x) kuvauksen eteen.

	Osaan käyttää yksinkertaisia ilmauksia ja lauseita kuvaamaan asuinpaikkaani ja tuntemiani ihmisiä. Osaan jonkin verran keskustella, jos keskustelukumppani auttaa minua toistamalla sanottavansa tai sanomalla asiat uudestaan hitaammin. Osaan kysyä yksinkertaisia kysymyksiä ja vastata niihin, jos ne käsittelevät tuttuja jokapäiväisiä asioita.	1
	Osaan käyttää yksinkertaisia ilmauksia ja lauseita kuvaamaan perhettäni ja ystäviäni ja muita ihmisiä, asuinpaikkaani (missä ja miten asun), koulutustani ja nykyistä tai aikaisempaa työtäni. Osaan keskustella jokapäiväisissä tilanteissa, joissa puhutaan lyhyesti tutuista asioista, mutta en osaa ranskaa tarpeeksi pitääkseni keskustelua yllä.	2
	Osaan kuvata yksinkertaisesti tapahtumia, kokemuksiani, unelmiani, toiveitani ja tavoitteitani. Osaan lyhyesti selittää mielipiteitä ja suunnitelmia. Osaan kertoa tarinan tai kuvata kirjan tai elokuvan juonta ja omia reaktioitani. Osaan valmistautumatta puhua minua kiinnostavista ja jokapäiväisistä asioista kuten perheestäni ja ystäväistäni, harrastuksista, työstä, matkustamisesta ja ajankohtaisista tapahtumista.	3
	Osaan esittää selkeän ja yksityiskohtaisen kuvauksen useista minua kiinnostavista aiheista. Osaan esittää mielipiteeni ja eri vaihtoehtojen hyviä ja huonoja puolia. Osaan aktiivisesti osallistua tutuissa tilanteissa käytäviin keskusteluihin ja ilmaista näkökantojani.	4
	Osaan käsitellä vaikeita aihealueita selkeästi, yksityiskohtaisesti ja eri näkökohdat huomioon ottaen. Osaan kehitellä aiheita ja saattaa asioita päätökseen. Osaan puhua sujuvasti ja spontaanisti. Minun tarvitsee hakea sanoja vain harvoin. Osaan käyttää ranskan kieltä tehokkaasti sekä yksityiselämässä että työssä. Osaan ilmaista itseäni tarkasti ja pystyn sopeuttamaan sanottavani muiden sanomisiin.	5
	Osaan kuvata ja selittää asioita selkeästi, joustavasti ja tilanteisiin sopivasti. Osaan rakentaa sanottavani niin, että kuulija voi helposti ymmärtää tärkeät koh-	6

	dat. Minulle on erittäin helppoa ottaa osaa keskusteluun, ja tunnen eri tilanteisiin sopivat ranskan kielen ilmaukset, myös puhekielen eri muodot. Osaan puhua sujuvasti ja ilmaista itseäni erittäin tarkasti. Osaan kiertää ongelmakohdat kuuli- jan edes huomaamatta.	
--	---	--

Alla on väittämiä, jotka liittyvät taitoosi puhua **ranskaa**. Käy läpi **jokainen väittämä** ja rastita (x)

- **KYLLÄ**      jos väittämä sopii sinuun  
tai  
- **EI**            jos väittämä ei sovi sinuun.

Kyllä	Ei	VÄITTÄMÄ	
		Osaan keskustella jokapäiväisistä käytännön asioista yksinkertaisesti, jos keskustelukumppani puhuu selkeästi ja hitaasti.	1
		Osaan puhua ranskaa luontevasti, sujuvasti ja tehokkaasti tutuissa tilanteissa.	2
		Osaan käyttää erilaisissa tilanteissa ranskaa joustavasti ja tehokkaasti mukaan lukien tunteiden ilmaisun, humoristisen kielenkäytön ja viittaukset tunnettuihin kaunokirjallisiin ja muihin lähteisiin.	3
		Selviän vieraistakin tilanteista julkisissa kulkuvälineissä, esimerkiksi jos joudun kysymään kanssamatkustajalta, missä minun pitäisi jäädä pois.	4
		Osaan muotoilla kutsuja ja anteeksipyyntöjä ja vastata niihin.	5
		Osaan palata takaisin vaikeisiin kohtiin ja muotoilla uudestaan sanottavani niin, että toiset tuskin huomaavat sitä.	6
		Osaan selittää ja perustella suunnitelmiani, aikomuksiani ja tekemisiäni.	7
		Osaan keskusteluissa ilmaista, selittää ja tukea mielipiteitäni antamalla asianmukaisia selityksiä, perusteluja ja kommentteja.	8
		Osaan vastata haastattelussa esitettyihin, yksinkertaisiin ja suoriin henkilökohtaisia asioita koskeviin kysymyksiin, kun ne esitetään hitaasti ja selkeästi.	9
		Osaan keskustella siitä, mitä tehdä, minne mennä ja osaan sopia tapamisista.	10
		Osaan sanoa mielipiteeni, kun minua puhutellaan yksinkertaisessa virallisessa tilaisuudessa, jos voin tarvittaessa pyytää toistamaan tärkeitä asiat.	11



		Osaan kysyä itseeni liittyviä ja toisia ihmisiä koskevia kysymyksiä, jotka liittyvät asuinpaikkaani, tuntemiini ihmisiin ja tavaroihini. Osaan myös vastata tällaisiin kysymyksiin.	12
		Osaan ilmaista ajatuksiani ja mielipiteitani selvästi ja tarkasti sekä esittää monimutkaisia perusteluja vakuuttavasti.	13
		Osaan kysyä toisten ihmisten työstä ja vapaa-ajasta sekä vastata toisten vastaaviin kysymyksiin.	14
		Pystyn esittämään vaikean asian vakuuttavasti ja selkeästi kuuntelijoille, jotka eivät tunne aiheita. Osaan sopeuttaa sanottavani kuuntelijointani ajatellen.	15
		Osaan ranskaa niin hyvin, että pystyn käsittelemään ristiriitatilanteita (esim. väärä pysäköintisakko, asunnon vaurioitumiseen liittyvä korvauskiista, onnettomuuteen liittyvä vastuukysymys).	16
		Tulen toimeen useimmissa tilanteissa, esimerkiksi kun järjestelen matkaa ulkomaisessa matkatoimistossa tai tilanteissa, joita kohtaan matkalla ollessani.	17
		Puhun ranskaa sujuvasti ja selkeästi niin että kieli ei aiheuta mitään ongelmia sosiaaliseen ja henkilökohtaiseen elämäni.	18
		Osaan ilmaista yksinkertaisia numeroita, määriä, hintoja ja aikoja.	19
		Osaan tervehtiä, kysyä kuulumisia ja vastata yksinkertaisiin asioihin, joita minulle kerrotaan.	20
		Puhun ranskan kieltä melkein yhtä hyvin kuin äidinkieltäni.	21
		Osaan aloittaa, ylläpitää ja lopettaa tavallisen keskustelun asioista, jotka ovat tuttuja tai joista olen henkilökohtaisesti kiinnostunut.	22
		Osaan ylläpitää keskustelua, mutta minua voi joskus olla vaikea ymmärtää, kun yritän sanoa tarkasti, mitä haluan.	23
		Osaan useita aihealueita käsiteltäessä ilmaista itseäni sujuvasti ja spontaanisti ilman suurempia vaikeuksia.	24
		Osaan keskustella tutuista asioista silloinkin kun minulla ei ole ollut mahdollisuutta valmistautua puhumaan.	25
		Kun en löydä tarvitsemaani sanaa, osaan käyttää yksinkertaisempaa sanaa, jolla on samansuuntainen merkitys ja pyytää toista sanomaan oikean sanan.	26
			27

		Osaan tehdä ostoksia tilanteissa, joissa voin osoittamalla tai muilla eleillä tukea sanottavaani.	
		Pystyn ilmaisemaan itseäni niin hyvin, että jopa syntyperäiset kielenpuhujat luulevat minun asuneen maassa kauan aikaa.	28
		Selviän lyhyistä sosiaalisista tilanteista, mutta pystyn harvoin pitämään keskustelua yllä.	29
		Osaan pyytää ihmisiltä esineitä ja asioita ja vastaavasti antaa pyydettyjä esineitä ja asioita.	30
		Osaan tehdä ehdotuksia ja vastata niihin.	31
		Osaan esittää yksinkertaisia kysymyksiä ja vastata niihin, esittää yksinkertaisia väittämiä ja ottaa niihin kantaa, kun kyseessä ovat jokapäiväiset asiat.	32
		Minulla on hyvä sanavarasto, jonka turvin selviydyn kohdatessani vaikeuksia, joita vieraat sanat mahdollisesti aiheuttavat.	33
		Pystyn selviytymään helposti väittelystä, vaikka siinä käsiteltäisiin abstrakteja ja vaikeita asioita, joita en tunne.	34
		Pystyn esittämään mielipiteen ajankohtaisesta asiasta ja kertomaan eri vaihtoehtojen eduista ja haitoista.	35
		Osaan puhua syntyperäiselle kielenpuhujalle siten, etten ole tahtomattani huvittava, ärsyttävä tai vaadi heitä käyttäytymään toisin kuin toisen syntyperäisen kanssa.	36
		Pystyn osallistumaan aktiivisesti keskusteluun, jossa käsitellään tavallisimpia aiheita.	37
		Osaan tuottaa selkeän, sujuvan ja loogisesti rakentuneen puheen.	38
		Osaan ilmaista tunnetiloja kuten hämmästys, onnellisuus, suru, mielenkiinto ja välinpitämättömyys ja reagoida niihin.	39
		Osaan kertoa melko yksityiskohtaisesti kokemuksistani ja kuvata tunteita ja reaktioita.	40
		Osaan esittäytyä ja esitellä muita ihmisiä ja käyttää tavanomaisia tervehdys- ja hyvästelyilmauksia.	41
		Osaan kuvailla ja määritellä konkreettisia asioita, joille en juuri sillä hetkellä satu muistamaan sopivaa sanaa (esimerkiksi pysäköintisakko, luottokortti tai vakuutuskirja).	42

Oliko Sinulla mitään erityisiä vaikeuksia vastata kysymyksiin? Jos oli, millaisia?  
Olivatko esimerkiksi jotkut väittämät tavallista hankalampia  
(merkitse väittämän numero)?

---

---

---

---

---

---

---

---

**LIITE 5 Osallistujan vastauslomake, tehtävä 1****OSALLISTUJAN VASTAUSLOMAKE****Tehtävä 1: Kuvasta kertominen**

Osallistujanumero: \_\_\_\_\_

Nimi: \_\_\_\_\_

Oppilaitos: \_\_\_\_\_

Kuulet nauhalta 26 suoritusta juuri tekemästäsi tehtävä 1:stä. Tehtävänäsi on nyt verrata omaa suoritustasi **jokaiseen** nauhalta kuuluvaan suoritukseen vastaamalla **kyllä** tai **ei** seuraavaan kysymykseen:

**Suoriuditko paremmin tehtävästä kuin nauhalta kuulemasi henkilö?**

Valitse vaihtoehto **kyllä**, jos suoriudit mielestäsi paremmin tehtävästä kuin kuulemasi henkilö.

Valitse vaihtoehto **ei**, jos suoriudit mielestäsi heikommin tai korkeintaan yhtä hyvin tehtävästä kuin kuulemasi henkilö.

Jos keksit, **kommentoi lyhyesti** kyllä/ei -vastauksiasi kohtia seuraaville viivoille (esim. mihin kiinnitit huomiota verratessasi omaa suoritustasi nauhalta kuulemaasi suoritukseen?).

**Merkitse rastilla** (x) vaikea päättää? -kohtaan ne suoritukset, joiden kohdalla sinulla oli **vaikeuksia päättää**, selviydyitkö tehtävästä paremmin vai heikommin kuin kuulemasi henkilö.

→ Vastaa seuraavalle sivulle!

**Suoriuditko paremmin tehtävästä kuin nauhalta kuulemasi henkilö?**

Näytteen numero		Miksi?	Vaikea päättää?
1	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (1)_____	_____
2	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (2)_____	_____
3	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (3)_____	_____
4	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (4)_____	_____
5	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (5)_____	_____
6	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (6)_____	_____
7	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (7)_____	_____
8	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (8)_____	_____
9	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (9)_____	_____
10	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (10)_____	_____
11	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (11)_____	_____
12	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (12)_____	_____
13	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (13)_____	_____
14	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (14)_____	_____
15	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (15)_____	_____
16	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (16)_____	_____
17	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (17)_____	_____
18	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (18)_____	_____
19	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (19)_____	_____
20	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (20)_____	_____
21	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (21)_____	_____
22	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (22)_____	_____
23	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (23)_____	_____
24	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (24)_____	_____
25	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (25)_____	_____
26	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (26)_____	_____

**LIITE 6 Osallistujan vastauslomake, tehtävä 2****OSALLISTUJAN VASTAUSLOMAKE****Tehtävä 2: Tilanteessa reagointi**

Osallistujanumero: \_\_\_\_\_

Nimi: \_\_\_\_\_

Oppilaitos: \_\_\_\_\_

Kuulet nauhalta 23 suoritusta juuri tekemästäsi tehtävä 2:sta. Tehtävänäsi on nyt verrata omaa suoritustasi **jokaiseen** nauhalta kuuluvaan suoritukseen vastaamalla **kyllä** tai **ei** seuraavaan kysymykseen:

**Suoriuditko paremmin tehtävästä kuin nauhalta kuulemasi henkilö?**

Valitse vaihtoehto **kyllä**, jos suoriudit mielestäsi paremmin tehtävästä kuin kuulemasi henkilö.

Valitse vaihtoehto **ei**, jos suoriudit mielestäsi heikommin tai korkeintaan yhtä hyvin tehtävästä kuin kuulemasi henkilö.

Jos keksit, **kommentoi lyhyesti** kyllä/ei -vastauksiasi kohtia seuraaville viivoille (esim. mihin kiinnitit huomiota verratessasi omaa suoritustasi nauhalta kuulemaasi suoritukseen?).

**Merkitse rastilla (x)** vaikea päättää? -kohtaan ne suoritukset, joiden kohdalla sinulla oli **vaikeuksia päättää**, selviydyitkö tehtävästä paremmin vai heikommin kuin kuulemasi henkilö.

→ Vastaa seuraavalle sivulle!

**Suoriuditko paremmin tehtävästä kuin nauhalta kuulemasi henkilö?**

Näytteen numero		Miksi?	Vaikea päättää?
1	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (1)_____	_____
2	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (2)_____	_____
3	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (3)_____	_____
4	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (4)_____	_____
5	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (5)_____	_____
6	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (6)_____	_____
7	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (7)_____	_____
8	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (8)_____	_____
9	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (9)_____	_____
10	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (10)_____	_____
11	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (11)_____	_____
12	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (12)_____	_____
13	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (13)_____	_____
14	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (14)_____	_____
15	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (15)_____	_____
16	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (16)_____	_____
17	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (17)_____	_____
18	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (18)_____	_____
19	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (19)_____	_____
20	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (20)_____	_____
21	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (21)_____	_____
22	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (22)_____	_____
23	<input type="checkbox"/> kyllä	<input type="checkbox"/> ei (23)_____	_____

**LIITE 7 Arviointiasteikko, tehtävä 1****SUULLISEN KIELITAIDON ARVIOINTIASTEIKKO**

(Tehtävä 1: Kuvasta kertominen)

<b>C2</b>	Pystyy tuottamaan taidokkaan ja mieleenpainuvan kuvauksen. Pystyy ilmaisemaan itseään hyvin sujuvasti, luontevan pakottomasti ja epäröimättä. Pysähtyy ainoastaan pohtimaan juuri oikeaa sanaa ilmaistakseen ajatuksensa. Hallitsee hyvin erittäin laajan sanaston, johon kuuluu idiomaattisia ilmauksia. Sanaston käyttö johdonmukaisesti oikea ja asianmukainen. Osaa taitavasti vaihdella intonaatiota ja asettaa lausepainon oikein ilmaistakseen hienompia merkitysvivahteita.
<b>C1</b>	Pystyy ilmaisemaan itseään sujuvasti ja spontaanisti. Hallitsee hyvin laajan sanaston, joka mahdollistaa puuttuvien sanojen vaivattoman kiertämisen. Vain vähän selvästi havaittavaa ilmausten hakemista. Hallitsee hyvin idiomaattiset ilmaukset. Vain satunnaisia, vähäisiä lipsahduksia sanaston hallinnassa. Osaa vaihdella intonaatiota ja asettaa lausepainon oikein ilmaistakseen hienompia merkitysvivahteita.
<b>B2</b>	Pystyy tuottamaan yhtämittaista puhetta suhteellisen tasaisesti. Vaikka ehkä osoittaakin epäröintiä etsiessään rakenteita ja ilmauksia, huomattavan pitkät tauot ovat harvinaisia. Omaa laajahkon sanavaraston ja pystyy jossain määrin vaihtelevaan muotoiluun välttääkseen jatkuvaa toistoa. Vain joitakin sekaannuksia ja vääriä sanavalintoja esiintyy. On omaksunut selkeän, luonnollisen ääntämyksen ja intonaation.
<b>B1</b>	Pystyy ilmaisemaan itseään suhteellisen helposti. Huolimatta joistain puheen muotoiluun liittyvistä ongelmista (tauot ja ilmausten rakenteiden hankaluudet) pystyy jatkamaan puhettaan ja ilmaisemaan ajatuksensa kohtalaiseen sujuvasti. Omaa riittävän sanavaraston ilmaistakseen itseään, joskin joutuu käyttämään jonkin verran kiertoilmauksia asioiden ilmaisuun. Osoittaa hyvää perustason sanaston hallintaa, vaikkakin sekaannuksia voi esiintyä kompleksisempia ajatuksia ilmaistaessa. Pystyy kuvaamaan henkilön tunnetiloja ja reaktioita. Ääntäminen on selvästi ymmärrettävää, vaikka vierasperäinen korostus on joskus ilmeinen ja ei-kohdekielenomaista ääntämistä esiintyy silloin tällöin.
<b>A2</b>	Ääntäminen on yleensä niin selkeää, että se on ymmärrettävää huolimatta selvästi havaittavasta vierasperäisestä korostuksesta. Kuulija saattaa joutua kuitenkin pyytämään toistoa aika ajoin. Puhuja pystyy rakentelemaan yksinkertaisia lauseita kohtalaiseen helposti, vaikka puheessa esiintyykin selvää epäröintiä. Omaa riittävän sanavaraston tehtävästä selviytyäkseen. Sanasto on kuitenkin kapeaa.
<b>A1</b>	Syntyperäiset ranskan kielen puhujat, jotka ovat tottuneet olemaan tekemisissä ranskaa käyttävien suomalaisten kanssa, pystyvät ymmärtämään puhujan puhetta, joka koostuu pääasiassa suppeista, ennalta opituista sanoista ja ilmauksista. Ymmärtäminen edellyttää kuulijalta tietyn asteista ponnistelua. Puhuja pysähtyy usein etsimään ilmausta.



**LIITE 8 Arviointiasteikko, tehtävä 2**

**SUULLISEN KIELITAIDON ARVIOINTIASTEIKKO**

(Tehtävä 2: Tilanteessa reagointi)

<b>C2</b>	Pystyy ilmaisemaan itseään hyvin sujuvasti, luontevan pakottomasti ja epäröimättä. Pysähtyy ainoastaan pohtimaan juuri oikeaa sanaa ilmaistakseen ajatuksensa. Reagoi tilanteessa erittäin luontevasti. Hallitsee erittäin hyvin sekä tilanteen vaatimat kohteliaisuusmuodot että sen vaatiman muodollisuusasteen. Hallitsee hyvin erittäin laajan sanaston, johon kuuluu idiomaattisia ilmauksia. Sanaston käyttö johdonmukaisesti oikea ja asianmukainen. Osaa taitavasti vaihdella intonaatiota ja asettaa lausepainon oikein ilmaistakseen hienompia merkitysvivahteita.
<b>C1</b>	Pystyy ilmaisemaan itseään sujuvasti ja spontaanisti. Reagoi tilanteessa luontevasti. Hallitsee hyvin sekä tilanteen vaatimat kohteliaisuusmuodot että sen vaatiman muodollisuusasteen. Hallitsee hyvin laajan sanaston, joka mahdollistaa puuttuvien sanojen vaivattoman kiertämisen. Vain vähän selvästi havaittavaa ilmausten hakemista. Hallitsee hyvin idiomaattiset ilmaukset. Vain satunnaisia, vähäisiä lipshauksia sanaston hallinnassa. Osaa vaihdella intonaatiota ja asettaa lausepainon oikein ilmaistakseen hienompia merkitysvivahteita.
<b>B2</b>	Pystyy kommunikoimaan siinä määrin sujuvasti ja spontaanisti, että kommunikaatio toisen osapuolen kanssa on täysin mahdollista ilman, että se aiheuttaa suurta ponnistusta kummallekaan osapuolelle. Reagoi tilanteessa kohtalaisen luontevasti. Hallitsee melko hyvin sekä tilanteen vaatimat kohteliaisuusmuodot että sen vaatiman muodollisuusasteen. Omaa laajahkon sanavaraston ja pystyy jossain määrin vaihtelevaan muotoiluun välttääkseen jatkuvaa toistoa. Vain joitakin sekaannuksia ja vääriä sanavalintoja esiintyy; ne eivät kuitenkaan estä kommunikaatiota. On omaksunut selkeän, luonnollisen ääntämyksen ja intonaation.
<b>B1</b>	Pystyy ilmaisemaan itseään suhteellisen helposti. Huolimatta joistain puheen muotoiluun liittyvistä ongelmista, jotka johtuvat tauoista ja ilmausten rakentelun hankaluuksista, pystyy jatkamaan puhetta ja ilmaisemaan ajatuksensa kohtalaisen sujuvasti. Tilanteessa reagointi ei aina kovin luontevaa. Epäluontevuutta tilanteen vaatimien kohteliaisuusmuotojen ja sen vaatiman muodollisuusasteen hallinnassa. Omaa riittävän sanavaraston ilmaistakseen itseään, joskin joutuu käyttämään jonkin verran kiertoilmauksia asioiden ilmaisuun, jotka koskevat esim. työtä. Osoittaa hyvää perustason sanaston hallintaa, vaikkakin sekaannuksia voi esiintyä kompleksisempia ajatuksia ilmaistaessa. Ääntäminen on selvästi ymmärrettävää, vaikka vierasperäinen korostus on joskus ilmeinen ja ei-kohdekielenomaista ääntämistä esiintyy silloin tällöin.
<b>A2</b>	Ääntäminen on yleensä niin selkeää, että se on ymmärrettävää huolimatta selvästi havaittavasta vierasperäisestä korostuksesta. Todellisessa kommunikaatiotilanteessa keskustelukumppani saattaisi kuitenkin joutua pyytämään toistoa aika ajoin. Puhuja pystyy rakentelemaan yksinkertaisia lauseita niin, että selviytyy vuorovaikutustilanteesta kohtalaisesti huolimatta selvästä epäröinnistä. Tilanteessa reagointi ei kovin luontevaa. Tilanteen vaatimien kohteliaisuusmuotojen ja sen vaatiman muodollisuusasteen hallinnassa puutteita. Omaa riittävän sanavaraston

	tehtävän edellyttämästä kommunikaatiotilanteesta selviytyäkseen. Sanasto on kuitenkin kapeaa.
<b>A1</b>	Syntyperäiset ranskan kielen puhujat, jotka ovat tottuneet olemaan tekemisissä ranskaa käyttävien suomalaisten kanssa, pystyvät ymmärtämään puhujan puhetta, joka koostuu pääasiassa suppeista, ennalta opituista sanoista ja ilmauksista. Ymmärtäminen edellyttää kuulijalta tietyn asteista ponnistelua. Puhuja pysähtyy usein etsimään ilmausta ja paikkailemaan kommunikaatiota. Tilanteessa reagointi kömpelöä, ei luontevaa. Ei hallitse lainkaan tilanteen vaatimia kohteliaisuusmuotoja eikä sen vaatimaa muodollisuusastetta.

**LIITE 9 Palautelomake**

Osallistujanumero \_\_\_\_\_

Nimi \_\_\_\_\_

Osallistuit ranskan kielen suullisen kielitaidon arviointia ja itsearviointia koskevaan väitöskirjatutkimukseeni syksyn -1999 - kevään 2000 aikana. Suoritit kielistudiossa 2 ranskan kielen puhumistehtävää, josta saat tässä viestissä lupamaani palautetta siitä, mille tasolle ranskan kielen arvioijat ovat suorituksesi arvioineet.

Yhteensä 3 kielitaidon arviointikoulutuksen suorittanutta arvioijaa on kuunnellut molemmat tehtäväsuorituksesi, joista he ovat sitten kukin itsenäisesti antaneet tasoarvion käyttäen työssään Euroopan Neuvoston kielitaidon arviointiasteikkoihin perustuvia 6-portaisia asteikoita. Saamasi tasoarvio on näiden kolmen arvioijan antama keskiarvo suorituksestasi.

Alla näet erillisen arvion sekä tehtävästä 1 että tehtävästä 2 sekä muistin virkistämiseksi oman itsearviointisi ranskan kielen puhumisen yleistasosta. Henkilökohtaisen arvion jälkeen selostan lyhyesti Euroopan Neuvoston arviointiasteikkojen ideologiaa sekä otan esille muutamia kielitaidon arviointiin liittyviä seikkoja, jotka saattavat olla hyvinkin merkityksellisiä saamasi arvion kannalta.

\*\*\*\*\*

OMA ITSEARVIOINTISI RANSKAN KIELEN PUHUMISEN TASOSTA  
(merkitsit tämän itsearviointilomakkeen etusivulle)

\_\_\_\_\_

SUORITUSTESI SAAMAT TASOARVIOT

Tehtävä 1 (Kuvasta kertominen, François) \_\_\_\_\_

Tehtävä 2 (Tilanteessa reagointi, esittely) \_\_\_\_\_

TASOARVIOITASI KUVAAVAT SANALLISET MÄÄRITELMÄT

\*\*\*\*\*

Euroopan Neuvoston arviointiasteikoista

Euroopan Neuvoston eräänä tärkeänä tehtävänä on luoda kaikkialle Eurooppaan yhtenäiset kielitaidon arviointiasteikot. Koska eri kouluilla ja oppilaitoksilla on kaikkialla Euroopassa käytössään lukemattomia erilaisia ja eri asioita tarkoittavia kielitaidon arviointitapoja, Euroopan Neuvoston työ yleiseurooppalaisten kielitaidon arviointiasteikkojen kehittämiseksi on ollut erityisen tärkeää. Yhteisten arviointiasteikkojen ideana on siis se, että yhtä hyvin suomalainen kielenopettaja kuin vaikkapa missä tahansa Euroopan maassa oleva hänen kollegansa tai vaikkapa työnantaja jakaisivat yhteisen näkemyksen siitä, millaista kielitaitoa esimerkiksi Euroopan Neuvoston taso B1 edustaa. Muun muassa tällä tavalla pyritään edistämään ihmisten helpompaa liikkuvuutta Euroopan sisällä (esim. työnhaku ulkomailta).

Euroopan Neuvoston kehittämät kielitaidon arviointiasteikot jaetaan kolmeen päätasoon: A, B ja C tasot. Kukin näistä kolmesta päätasosta on edelleen jaettu kahteen alatasoon, jolloin tässäkin tutkimuksessa käytetyt 6 arviointitasoa ovat alhaalta ylöspäin seuraavat:

A1 A = Perustason kielenkäyttäjät

A2

B1 B = Itsenäiset kielenkäyttäjät

B2

C1 C = Taitavat kielenkäyttäjät

C2

#### Tasoarvioista

Saamaasi tasoarviota tarkastellessasi on syytä pitää mielessä, että nämä tasoarviot on annettu tietyistä tehtävistä tietynä ajankohtana. On siis hyvinkin mahdollista, että ranskan kielen taitosi voi mielestäsi olla yleisesti ottaen eri tasolla kuin tässä tutkimuksessa saamasi arviot. Suullinen kielitaito vaihtelee jokaisella aina jonkin verran tilanteesta toiseen (varsinkin jos et ole vielä kovin pitkällä ranskan opinnoissasi). Saamaasi tasoarvioon on saattanut vaikuttaa alentavasti mm. seuraavat seikat:

- studiotilanteen ”jännittävyys”, tehtävätilanteen ”epäaitous”, rajoitettu puhe-aika, tavalliseen kommunikaatiotilanteeseen liittyvän non-verbaliikan puuttuminen, oma psykososiaalinen tilasi (väsymys, jännittyneisyys, keskittymiskyvyn puute yms.), aikaisemman kokemuksen puuttuminen tehtävässä esiintyvistä tilanteista.

Yleisenä kommenttina todettakoon vielä lopuksi, että suurin osa osallistujista selviytyisi todennäköisesti hieman paremmin vastaavasta todellisesta kielenkäyttötilanteesta kuin testitilanteesta.

Mikäli haluat lisätietoa Euroopan Neuvoston työstä kielitaidon arvioinnin alalla ja olet kiinnostunut kehittämään myöskin itsenäisesti omaa kielitaitoasi eteenpäin, kehotan Sinua tutustumaan alla olevaan Internet -osoitteeseen. Sieltä löydät halutessasi muita eritasoisia kielitaitoa kuvaavia sanallisia määritelmiä edellä mainituista taitotasoista. Muistuttaisin myös, että vuoden 2001 lopulta lähtien käytössäsi on web-pohjainen kielitaidon arviointityökalu DIALANG, jonka avulla pystyt seuraamaan kielitaitosi kehittymistä kaikissa EU-kielissä.

<http://culture.coe.fr/lang> (Euroopan neuvoston sivut)  
<http://www.sprachlabor.fu-berlin.de/dialang/>  
(DIALANG)

Jos taas haluat kysyä minulta vielä jotain tutkimukseen tai sen arviointiin liittyvää, ota yhteys sähköpostitse (sähköpostiosoite).

Kiitos vielä kerran osallistumisesta tutkimukseen ja antoisia hetkiä ranskan kielin parissa myös tulevaisuudessa!

LIITE 10 ITSEARVIOINTI PAPERILLA, YHTEENVEDOT OPPIJOISTA  
JA ITSEARVIOINTIVÄITTÄMISTÄ (väittäjä 29 poistettu)

INPUT: 210 Person 42 Item REPORTED: 209 Person 40 Item 80 CATS WINSTEPS 3.72.3

SUMMARY OF 209 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	22.6	41.0	.55	.58	.97	-.2	1.22	.2
S.D.	8.3	.1	2.64	.08	.40	1.2	2.27	1.5
MAX.	40.0	41.0	7.63	1.19	2.56	4.5	9.90	9.9
MIN.	4.0	40.0	-5.97	.52	.31	-2.7	.08	-1.2
REAL RMSE	.62	TRUE SD	2.57	SEPARATION	4.11	Person	RELIABILITY	.94
MODEL RMSE	.58	TRUE SD	2.58	SEPARATION	4.43	Person	RELIABILITY	.95
S.E. OF Person MEAN = .18								

DELETED: 1 Person  
Person RAW SCORE-TO-MEASURE CORRELATION = 1.00  
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .94

SUMMARY OF 40 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD
MEAN	113.1	208.9	.00	.29	.97	-.3	1.97	.8
S.D.	67.5	.2	3.65	.13	.20	1.6	2.93	3.5
MAX.	207.0	209.0	7.64	.75	1.68	5.5	9.90	9.9
MIN.	3.0	208.0	-6.76	.20	.70	-2.6	.14	-2.3
REAL RMSE	.34	TRUE SD	3.64	SEPARATION	10.78	Item	RELIABILITY	.99
MODEL RMSE	.32	TRUE SD	3.64	SEPARATION	11.25	Item	RELIABILITY	.99
S.E. OF Item MEAN = .59								

LACKING RESPONSES: 1 Item  
DELETED: 1 Item  
UMEAN=.0000 USCALE=1.0000  
Item RAW SCORE-TO-MEASURE CORRELATION = -.98  
8358 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 4163.13 with 8110 d.f. p=1.0000  
Global Root-Mean-Square Residual (excluding extreme scores): .2737

LIITE 11 ITSEARVIOINTI PAPERILLA, YHTEENVEDOT OPPIJOISTA  
JA ITSEARVIOINTIVÄITTÄMISTÄ

INPUT: 210 Person 42 Item REPORTED: 209 Person 41 Item 82 CATS WINSTEPS 3.72.3

SUMMARY OF 209 MEASURED Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	23.3	42.0	.51	.53	.98	-.2	1.23	.2	
S.D.	8.1	.1	2.18	.06	.37	1.2	1.86	1.2	
MAX.	40.0	42.0	5.96	.87	2.39	3.9	9.90	9.9	
MIN.	5.0	41.0	-4.97	.49	.33	-2.9	.14	-.9	
REAL RMSE	.57	TRUE SD	2.10	SEPARATION	3.67	Person	RELIABILITY	.93	
MODEL RMSE	.53	TRUE SD	2.11	SEPARATION	3.97	Person	RELIABILITY	.94	
S.E. OF Person MEAN = .15									

DELETED: 1 Person

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00

CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .93

SUMMARY OF 41 MEASURED Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	113.5	209.0	.00	.28	.94	-.8	1.85	.2	
S.D.	66.7	.2	3.21	.13	.41	2.3	3.03	3.1	
MAX.	207.0	209.0	6.74	.74	3.32	9.9	9.90	9.9	
MIN.	3.0	208.0	-6.18	.18	.67	-3.7	.12	-2.1	
REAL RMSE	.32	TRUE SD	3.19	SEPARATION	10.02	Item	RELIABILITY	.99	
MODEL RMSE	.31	TRUE SD	3.19	SEPARATION	10.38	Item	RELIABILITY	.99	
S.E. OF Item MEAN = .51									

LACKING RESPONSES: 1 Item

UMEAN=.0000 USCALE=1.0000

Item RAW SCORE-TO-MEASURE CORRELATION = -.98

8567 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 4869.83 with 8318 d.f. p=1.0000

Global Root-Mean-Square Residual (excluding extreme scores): .2925

## LIITE 12 TEHTÄVÄ 1: YHTEENVETO OPPIJOISTA

INPUT: 210 Person 26 Item REPORTED: 209 Person 26 Item 52 CATS WINSTEPS 3.72.3

### SUMMARY OF 205 MEASURED (NON-EXTREME) Person

	TOTAL		MEASURE	MODEL	INFIT		OUTFIT	
	SCORE	COUNT			ERROR	MNSQ	ZSTD	MNSQ
MEAN	12.4	26.0	-.21	.75	.93	-.1	.97	.1
S.D.	6.9	.3	2.98	.24	.59	1.2	1.72	1.2
MAX.	25.0	26.0	7.62	1.42	4.86	8.0	9.90	9.9
MIN.	1.0	23.0	-6.73	.53	.12	-2.1	.02	-1.0
REAL RMSE	.87	TRUE SD	2.85	SEPARATION	3.27	Person	RELIABILITY	.91
MODEL RMSE	.79	TRUE SD	2.88	SEPARATION	3.65	Person	RELIABILITY	.93
S.E. OF Person MEAN = .21								

MAXIMUM EXTREME SCORE: 3 Person  
 MINIMUM EXTREME SCORE: 1 Person  
 DELETED: 1 Person

### SUMMARY OF 209 MEASURED (EXTREME AND NON-EXTREME) Person

	TOTAL		MEASURE	MODEL	INFIT		OUTFIT	
	SCORE	COUNT			ERROR	MNSQ	ZSTD	MNSQ
MEAN	12.5	26.0	-.11	.77				
S.D.	7.1	.3	3.21	.29				
MAX.	26.0	26.0	9.31	1.99				
MIN.	.0	23.0	-8.21	.53	.12	-2.1	.02	-1.0
REAL RMSE	.91	TRUE SD	3.08	SEPARATION	3.41	Person	RELIABILITY	.92
MODEL RMSE	.83	TRUE SD	3.11	SEPARATION	3.75	Person	RELIABILITY	.93
S.E. OF Person MEAN = .22								

Person RAW SCORE-TO-MEASURE CORRELATION = .97  
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .94

### SUMMARY OF 26 MEASURED (NON-EXTREME) Item

	TOTAL		MEASURE	MODEL	INFIT		OUTFIT	
	SCORE	COUNT			ERROR	MNSQ	ZSTD	MNSQ
MEAN	100.5	208.8	.00	.28	.98	-.2	2.12	1.0
S.D.	54.2	.4	3.29	.12	.19	1.5	2.57	2.7
MAX.	201.0	209.0	7.70	.66	1.41	3.4	9.90	8.3
MIN.	6.0	208.0	-6.43	.20	.65	-3.0	.34	-1.8
REAL RMSE	.31	TRUE SD	3.28	SEPARATION	10.64	Item	RELIABILITY	.99
MODEL RMSE	.30	TRUE SD	3.28	SEPARATION	10.94	Item	RELIABILITY	.99
S.E. OF Item MEAN = .66								

UMEAN=.0000 USCALE=1.0000  
 Item RAW SCORE-TO-MEASURE CORRELATION = -.98  
 5324 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 2907.99 with 5094 d.f. p=1.0000  
 Global Root-Mean-Square Residual (excluding extreme scores): .2883



## LIITE 13 TEHTÄVÄ 2: YHTEENVETO OPPIJOISTA

INPUT: 210 Person 23 Item REPORTED: 209 Person 23 Item 46 CATS WINSTEPS 3.72.3

### SUMMARY OF 208 MEASURED (NON-EXTREME) Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	12.0	23.0	.17	.67	.95	-.1	1.07	.2
S.D.	5.2	.5	2.15	.14	.46	1.2	1.64	1.0
MAX.	22.0	23.0	5.41	1.18	4.67	6.3	9.90	7.0
MIN.	1.0	16.0	-5.34	.59	.37	-2.4	.07	-.8
REAL RMSE	.73	TRUE SD	2.02	SEPARATION	2.78	Person	RELIABILITY	.89
MODEL RMSE	.68	TRUE SD	2.04	SEPARATION	2.98	Person	RELIABILITY	.90
S.E. OF Person MEAN = .15								

MAXIMUM EXTREME SCORE: 1 Person  
DELETED: 1 Person

### SUMMARY OF 209 MEASURED (EXTREME AND NON-EXTREME) Person

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	12.0	23.0	.21	.68				
S.D.	5.2	.5	2.19	.16				
MAX.	23.0	23.0	6.87	1.92				
MIN.	1.0	16.0	-5.34	.59	.37	-2.4	.07	-.8
REAL RMSE	.74	TRUE SD	2.07	SEPARATION	2.80	Person	RELIABILITY	.89
MODEL RMSE	.70	TRUE SD	2.08	SEPARATION	2.99	Person	RELIABILITY	.90
S.E. OF Person MEAN = .15								

Person RAW SCORE-TO-MEASURE CORRELATION = .99  
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .89

### SUMMARY OF 23 MEASURED (NON-EXTREME) Item

	TOTAL SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT	
					MNSQ	ZSTD	MNSQ	ZSTD
MEAN	109.1	208.7	.00	.24	1.01	-.2	1.39	.2
S.D.	58.5	.5	2.56	.08	.18	1.2	1.54	1.5
MAX.	203.0	209.0	5.13	.48	1.43	1.7	7.96	4.6
MIN.	9.0	208.0	-5.32	.18	.72	-3.1	.47	-2.0
REAL RMSE	.27	TRUE SD	2.55	SEPARATION	9.50	Item	RELIABILITY	.99
MODEL RMSE	.25	TRUE SD	2.55	SEPARATION	10.23	Item	RELIABILITY	.99
S.E. OF Item MEAN = .55								

UMEAN=.0000 USCALE=1.0000  
Item RAW SCORE-TO-MEASURE CORRELATION = -.99  
4776 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 3143.52 with 4546 d.f. p=1.0000  
Global Root-Mean-Square Residual (excluding extreme scores): .3186

## LIITE 14 FACETS-ANALYYSIEN TULOKSET ARVIOIJISTA, TEHTÄVÄ 1 (1/2)

**Analyysi 1: Kaikki arvioijat** (arvioijat 2 ja 3 arvioivat kaikki suoritukset, arvioijat 1a, 1b, 1c, 1d ja 1e vain osan suorituksista)

Arvioija Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Measure	Model S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
70	32	2.2	3.12	2.94	.63	.64	-.7	.45	-.5	1.17	.38	64.5	70.7	5 arv.1a
149	40	3.8	3.03	2.19	.36	1.47	1.7	1.86	1.9	.38	.84	36.5	41.5	2 arv.1b
115	40	2.9	2.81	.26	.38	1.00	.0	.93	.0	1.00	.90	45.7	60.9	1 arv.1c
591	210	2.9	2.63	-.57	.18	.70	-2.7	.55	-1.2	1.28	.92	66.2	66.4	6 arv.2
587	210	2.9	2.60	-.71	.19	.57	-4.1	.42	-1.7	1.38	.92	67.5	66.4	7 arv.3
160	44	3.6	2.40	-1.51	.33	1.96	3.5	1.79	2.6	.09	.49	38.6	57.9	3 arv.1d
85	45	2.1	2.18	-2.61	.50	1.67	2.3	.98	.4	.18	.68	60.5	74.0	4 arv.1e
251.0	88.7	2.9	2.68	.00	.37	1.15	.0	.99	.2		.73			Mean (Count: 7)
215.8	76.8	.6	.31	1.83	.15	.51	2.6	.56	1.5		.21			S.D. (Population)
233.1	83.0	.7	.33	1.98	.16	.56	2.8	.61	1.6		.22			S.D. (Sample)

**Analyysi 2: Arvioijat 1b ja 1d poistettu**

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Measure	Model S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
70	32	2.2	3.07	3.74	.79	.76	-.2	1.29	.6	.96	.38	64.5	70.6	5 arv.1a
115	40	2.9	2.95	.49	.43	1.26	1.5	1.00	.3	.60	.90	45.7	66.6	1 arv.1c
591	210	2.9	2.86	-.51	.26	.73	-2.1	.27	-1.0	1.27	.94	74.3	79.7	6 arv.2
587	210	2.8	2.83	-.78	.26	.73	-2.1	.40	-.7	1.22	.94	75.0	79.6	7 arv.3
85	45	2.1	2.35	-2.94	.56	1.88	2.7	1.33	1.0	-1.00	.68	60.5	75.7	4 arv.1e
289.6	107.4	2.6	2.81	.00	.46	1.07	.0	.86	.0		.77			Mean (Count: 5)
244.9	83.9	.4	.24	2.18	.20	.45	2.0	.44	.8		.22			S.D. (Population)
273.8	93.8	.4	.27	2.43	.22	.50	2.2	.49	.9		.24			S.D. (Sample)

**LIITE 14 (JATKUU) FACETS-ANALYYSIEN TULOKSET ARVIOIJISTA, TEHTÄVÄ 1 (2/2)**

**Analyysi 3: Arvioijat 1b, 1d ja 1e poistettu**

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
70	32	2.2	3.05	3.19	.86	.84	.0	2.90	1.3	.88	.38	64.5	70.7	5 arv.1a
115	40	2.9	2.99	-.28	.44	1.32	2.0	1.01	.4	.29	.90	45.7	67.5	1 arv.1c
591	210	2.9	2.97	-1.28	.29	.84	-1.4	.23	-.7	1.23	.94	76.4	83.0	6 arv.2
587	210	2.8	2.96	-1.63	.30	.77	-1.9	.20	-.8	1.28	.95	77.1	82.8	7 arv.3
340.8	123.0	2.7	2.99	.00	.47	.94	-.3	1.09	.1		.79			Mean (Count: 4)
248.8	87.0	.3	.03	1.91	.23	.22	1.6	1.10	.9		.24			S.D. (Population)
287.2	100.5	.3	.04	2.20	.27	.25	1.8	1.27	1.1		.28			S.D. (Sample)

**Lopullinen analyysi: Vain arvioija 2 ja arvioija 3 mukana**

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Model Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
591	210	2.7	2.92	.23	.34	.85	-1.1	.17	-1.3	1.19	.95	82.8	90.0	6 arvioija 2
587	210	2.7	2.88	-.23	.34	.87	-.9	.17	-1.1	1.19	.95	82.8	90.0	7 arvioija 3
589.0	210.0	2.7	2.90	.00	.34	.86	-1.1	.17	-1.3		.95			Mean (Count: 2)
2.0	.0	.0	.02	.23	.00	.01	.1	.00	.1		.00			S.D. (Population)
2.8	.0	.0	.03	.33	.00	.01	.1	.00	.2		.00			S.D. (Sample)

Model, Populn: RMSE .34 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability (not inter-rater) .00  
 Model, Sample: RMSE .34 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability (not inter-rater) .00  
 Model, Fixed (all same) chi-square: .9 d.f.: 1 significance (probability): .34  
 Inter-Rater agreement opportunities: 174 Exact agreements: 144 = 82.8% Expected: 156.5 = 90.0%

## LIITE 15 FACETS-ANALYYSIEN TULOKSET ARVIOIJISTA, TEHTÄVÄ 2 (1/2)

### Facets-analyysien tulokset arvioijista Tehtävässä 2

**Analyysi 1: Kaikki arvioijat** (arvioijat 2 ja 3 arvioivat kaikki suoritukset, arvioijat 1a, 1b, 1c, 1d ja 1e vain osan. suorituksista)

Arvioija Measurement Report (arranged by mN).

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Model Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
114	40	3.1	3.42	2.88	.36	1.61	2.0	1.96	2.6	.40	.83	25.0	48.4	5 1a
140	40	3.5	3.19	1.95	.32	1.27	1.1	1.27	1.0	.75	.82	47.5	56.0	2 1b
640	210	3.1	3.05	1.21	.15	.59	-4.4	.55	-3.6	1.37	.84	58.9	51.7	6 arv.2
627	210	3.1	3.00	.91	.15	.61	-4.2	.56	-3.8	1.36	.81	58.6	51.8	7 arv.3
95	40	2.5	2.73	-.51	.35	2.86	5.7	2.85	5.1	-.92	.51	33.8	53.0	1 1c
91	45	2.1	2.72	-.55	.35	.83	-.6	.93	-.2	1.11	.66	59.5	58.8	4 1e
118	44	2.7	1.68	-5.90	.33	2.10	3.7	2.20	3.4	-.01	.57	9.1	4.1	3 1d
260.7	89.9	2.9	2.83	.00	.29	1.41	.5	1.47	.7		.72			Mean (Count: 7)
236.3	76.0	.4	.52	2.67	.09	.78	3.6	.82	3.2		.13			S.D. (Population)
255.2	82.1	.5	.56	2.88	.10	.84	3.9	.88	3.5		.14			S.D. (Sample)

### Analyysi 2: Arvioijat 1c ja 1d poistettu

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Model Measure	S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
114	40	3.1	3.11	2.33	.46	2.62	3.6	4.23	4.2	-.56	.83	25.0	52.4	5 1a
140	40	3.5	3.03	1.21	.41	2.01	3.1	1.81	1.5	.14	.82	47.5	66.0	2 1b
640	210	3.1	3.00	.06	.19	.56	-4.5	.33	-3.6	1.41	.92	68.3	69.4	6 arv.2
627	210	3.0	2.99	-.42	.19	.60	-3.9	.44	-2.7	1.35	.91	68.3	68.8	7 arv.3
91	45	2.1	2.77	-3.19	.50	1.35	1.0	2.45	2.3	.56	.66	59.5	67.3	4 1e
322.4	109.0	3.0	2.98	.00	.35	1.43	-.1	1.85	.3		.83			Mean (Count: 5)
254.5	82.5	.5	.11	1.86	.13	.80	3.5	1.44	3.0		.10			S.D. (Population)
284.6	92.2	.5	.13	2.08	.15	.90	3.9	1.61	3.4		.11			S.D. (Sample)

LIITE 15 (JATKU) FACETS-ANALYYSIEN TULOKSET ARVIOIJISTA, TEHTÄVÄ 2 (2/2)

Analyysi 3: Arvioijat 1c, 1a, 1b ja 1d poistettu

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Measure	Model S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
640	210	3.1	3.01	1.71	.29	.73	-1.9	.20	-1.5	1.25	.93	77.3	82.7	6 arv.2
627	210	3.1	3.00	.64	.29	.78	-1.6	.36	-.8	1.19	.93	79.0	84.0	7 arv.3
91	45	2.1	2.99	-2.35	.56	1.66	1.4	3.73	2.3	.28	.66	59.5	68.3	4 1e
452.7	155.0	2.8	3.00	.00	.38	1.06	-.7	1.43	.0		.84			Mean (Count: 3)
255.8	77.8	.5	.01	1.72	.13	.43	1.5	1.63	1.7		.13			S.D. (Population)
313.3	95.3	.6	.01	2.11	.16	.52	1.9	2.00	2.1		.16			S.D. (Sample)

Lopullinen analyysi: Vain arvioija 2 ja arvioija 3 mukana

Total Score	Total Count	Obsvd Average	Fair-M Avrage	Measure	Model S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Corr. PtBis	Exact Obs %	Agree. Exp %	N Arvioija
640	210	3.1	3.01	.69	.33	.84	-1.0	.18	-1.4	1.18	.94	82.1	88.9	6 arv.2
627	210	3.1	3.00	-.69	.34	.87	-.8	.18	-.7	1.18	.94	82.1	88.9	7 arv.3
633.5	210.0	3.1	3.01	.00	.34	.85	-.9	.18	-1.1		.94			Mean (Count: 2)
6.5	.0	.0	.01	.69	.00	.01	.1	.00	.3		.00			S.D. (Population)
9.2	.0	.0	.01	.98	.00	.02	.1	.00	.5		.00			S.D. (Sample)

Model, Populn: RMSE .34 Adj (True) S.D. .61 Separation 1.80 Strata 2.74 Reliability (not inter-rater) .76  
 Model, Sample: RMSE .34 Adj (True) S.D. .92 Separation 2.74 Strata 3.98 Reliability (not inter-rater) .88  
 Model, Fixed (all same) chi-square: 8.5 d.f.: 1 significance (probability): .00  
 Inter-Rater agreement opportunities: 195 Exact agreements: 160 = 82.1% Expected: 173.4 = 88.9%

LIITE 16 "KYLLÄ"-VASTAUSTEN PROSENTTIOSUUEDET, ITSEARVIOINTI PAPERILLA (1/4)

Väittäjä	"Kyllä"-vastausten %-osuus	N	Puuttuvat vastaukset	Std. Deviation	VÄITTÄMÄ
1	97	209	1	,167	Osaan keskustella jokapäiväisistä käytännön asioista yksinkertaisesti, jos keskustelukumppani puhuu selkeästi ja hitaasti.
2	48	209	1	,501	Osaan puhua ranskaa luontevasti, sujuvasti ja tehokkaasti tutuissa tilanteissa.
3	16	209	1	,370	Osaan käyttää erilaisissa tilanteissa ranskaa joustavasti ja tehokkaasti mukaan lukien tunteiden ilmaisun, humoristisen kielenkäytön ja viittaukset tunnettuihin kaunokirjallisiin ja muihin lähteisiin.
4	77	209	1	,422	Selviän vieraistakin tilanteista julkisissa kulkuvälineissä, esimerkiksi jos joudun kysymään kanssamatkustajalta, missä minun pitäisi jäädä pois.
5	70	209	1	,460	Osaan muotoilla kutsuja ja anteeksipyyntöjä ja vastata niihin.
6	12	209	1	,325	Osaan palata takaisin vaikeisiin kohtiin ja muotoilla uudestaan sanottavani niin, että toiset tuskin huomaavat sitä.
7	64	209	1	,481	Osaan selittää ja perustella suunnitelmiani, aikomuksiani ja tekemisiäni.
8	39	209	1	,488	Osaan keskusteluissa ilmaista, selittää ja tukea mielipiteitäni antamalla asianmukaisia selityksiä, perusteluja ja kommentteja.
9	97	209	1	,180	Osaan vastata haastattelussa esitettyihin, yksinkertaisiin ja suoriin henkilökohtaisia asioita koskeviin kysymyksiin, kun ne esitetään hitaasti ja selkeästi.
10	94	209	1	,242	Osaan keskustella siitä, mitä tehdä, minne mennä ja osaan sopia tapaamisista.
11	76	209	1	,428	Osaan sanoa mielipiteeni, kun minua puhutellaan yksinkertaisessa virallisessa tilaisuudessa, jos voin tarvittaessa pyytää toistamaan tärkeät asiat.

LIITE 16 (JATKUU) "KYLLÄ"-VASTAUSTEN PROSENTTIOSUUEDET, ITSEARVIOINTI PAPERILLA (2/4)

	"Kyllä"- vastausten %-osuus	N	Puuttuvat vastaukset	Std. Deviation	VÄITTÄMÄ
12	93	209	1	,251	Osaan kysyä itseeni liittyviä ja toisia ihmisiä koskevia kysymyksiä, jotka liittyvät asuinpaikkaani, tuntemiini ihmisiin ja tavaroihini. Osaan myös vastata tällaisiin kysymyksiin.
13	10	209	1	,301	Osaan ilmaista ajatuksiani ja mielipiteitäni selvästi ja tarkasti sekä esittää monimutkaisia perusteluja vakuuttavasti.
14	90	209	1	,295	Osaan kysyä toisten ihmisten työstä ja vapaa-ajasta sekä vastata toisten vastaaviin kysymyksiin.
15	09	209	1	,288	Pystyn esittämään vaikean asian vakuuttavasti ja selkeästi kuuntelijoille, jotka eivät tunne aihetta. Osaan sopeuttaa sanottavani kuuntelijoitani ajattelun.
16	17	209	1	,374	Osaan ranskaa niin hyvin, että pystyn käsittelemään ristiriitatilanteita (esim. väärä pysäköintisakko, asunnon vaurioitumiseen liittyvä korvauskiista, onnettomuuteen liittyvä vastuukysymys).
17	61	209	1	,489	Tulen toimeen useimmissa tilanteissa, esimerkiksi kun järjestelen matkaa ulkomaisessa matkatoimistossa tai tilanteissa, joita kohtaan matkalla ollessani.
18	16	209	1	,370	Puhun ranskaa sujuvasti ja selkeästi niin että kieli ei aiheuta mitään ongelmia sosiaaliseen ja henkilökohtaiseen elämäni.
19	99	209	1	,119	Osaan ilmaista yksinkertaisia numeroita, määriä, hintoja ja aikoja.
20	1	209	2	,000	Osaan tervehtiä, kysyä kuulumisia ja vastata yksinkertaisiin asioihin, joita minulle kerrotaan.
21	1	209	1	,119	Puhun ranskan kieltä melkein yhtä hyvin kuin äidinkieltäni.
22	65	208	2	,478	Osaan aloittaa, ylläpitää ja lopettaa tavallisen keskustelun asioista, jotka ovat tuttuja tai joista olen henkilökohtaisesti kiinnostunut.

LIITE 16 (JATKU) "KYLLÄ"-VASTAUSTEN PROSENTTIOSUUEDET, ITSEARVIOINTI PAPERILLA (3/4)

	"Kyllä"- vastausten %-osuus	N	Puuttuvat vastaukset	Std. Deviation	VÄITTÄMÄ
23	64	209	1	,482	Osaan ylläpitää keskustelua, mutta minua voi joskus olla vaikea ymmärtää, kun yritän sanoa tarkasti, mitä haluan.
24	22	209	1	,415	Osaan useita aihealueita käsiteltäessä ilmaista itseäni sujuvasti ja spontaanisti ilman suurempia vaikeuksia.
25	69	209	1	,462	Osaan keskustella tutuista asioista silloinkin kun minulla ei ole ollut mahdollisuutta valmistautua puhumaan.
26	68	209	1	,466	Kun en löydä tarvitsemaani sanaa, osaan käyttää yksinkertaisempaa sanaa, jolla on samansuuntainen merkitys ja pyytää toista sanomaan oikean sanan.
27	97	209	1	,167	Osaan tehdä ostoksia tilanteissa, joissa voin osoittamalla tai muilla eleillä tukea sanottavaani.
28	07	209	1	,259	Pystyn ilmaisemaan itseäni niin hyvin, että jopa syntyperäiset kielenpuhujat luulevat minun asuneen maassa kauan aikaa.
29	62	209	1	,486	Selviän lyhyistä sosiaalisista tilanteista, mutta pystyn harvoin pitämään keskustelua yllä.
30	93	209	1	,251	Osaan pyytää ihmisiltä esineitä ja asioita ja vastaavasti antaa pyydettyjä esineitä ja asioita.
31	83	209	1	,879	Osaan tehdä ehdotuksia ja vastata niihin.
32	88	209	1	,325	Osaan esittää yksinkertaisia kysymyksiä ja vastata niihin, esittää yksinkertaisia väittämiä ja ottaa niihin kantaa, kun kyseessä ovat jokapäiväiset asiat.
33	25	209	1	,436	Minulla on hyvä sanavarasto, jonka turvin selviydyn kohdatessani vaikeuksia, joita vieraat sanat mahdollisesti aiheuttavat.
34	05	209	1	,214	Pystyn selviytymään helposti väittelystä, vaikka siinä käsiteltäisiin abstrakteja ja vaikeita asioita, joita en tunne.



**LIITE 16 (JATKUU) "KYLLÄ"-VASTAUSTEN PROSENTTIOSUUEDET, ITSEARVIOINTI PAPERILLA (4/4)**

	"Kyllä"- vastausten %-osuus	N	Puuttuvat vastaukset	Std. Deviation	VÄITTÄMÄ
35	31	209	1	,462	Pystyn esittämään mielipiteen ajankohtaisesta asiasta ja kertomaan eri vaihtoehtojen eduista ja haitoista.
36	29	209	1	,456	Osaan puhua syntyperäiselle kielenpuhujalle siten, etten ole tahtomattani huvittava, ärsyttävä tai vaadi heitä käyttäytymään toisin kuin toisen syntyperäisen kanssa.
37	52	209	1	,501	Pystyn osallistumaan aktiivisesti keskusteluun, jossa käsitellään tavallisimpia aiheita.
38	23	209	1	,422	Osaan tuottaa selkeän, sujuvan ja loogisesti rakentuneen puheen.
39	70	209	1	,458	Osaan ilmaista tunnetiloja kuten hämmästys, onnellisuus, suru, mielenkiinto ja välinpitämättömyys ja reagoida niihin.
40	37	208	2	,484	Osaan kertoa melko yksityiskohtaisesti kokemuksistani ja kuvata tunteita ja reaktioita.
41	99	209	1	,98	Osaan esittäytyä ja esitellä muita ihmisiä ja käyttää tavanomaisia tervehdys- ja hyvästelyilmauksia.
42	51	209	1	,501	Osaan kuvailla ja määritellä konkreettisia asioita, joille en juuri sillä hetkellä satu muistamaan sopivaa sanaa (esimerkiksi pysäköintisakko, luottokortti tai vakuutuskirja).

LIITE 17 "KYLLÄ"-VASTAUSTEN PROSENTTIOSUUDET,  
TEHTÄVÄ 1

Näyte	"Kyllä"- vastausten %-osuus	N	Puuttuvat vastaukset	Std. Deviation
1	65	209	1	,478
2	33	209	1	,470
3	88	209	1	,325
4	50	209	1	,501
5	14	209	1	,351
6	41	209	1	,493
7	97	208	2	,181
8	50	209	1	,501
9	3	209	1	,180
10	92	208	2	,267
11	44	209	1	,498
12	3	209	1	,167
13	59	209	1	,493
14	26	209	1	,441
15	33	209	1	,473
16	30	208	2	,459
17	95	209	1	,224
18	41	209	1	,493
19	56	209	1	,498
20	38	209	1	,487
21	36	209	1	,481
22	73	209	1	,444
23	28	208	2	,452
24	50	209	1	,501
25	29	208	2	,454
26	76	208	2	,425

LIITE 18 "KYLLÄ"-VASTAUSTEN PROSENTTIOSUUDET,  
TEHTÄVÄ 2

Näyte	"Kyllä"- vastausten %-osuus	N	Puuttuvat vastaukset	Std. Deviation
1	97	209	1	,167
2	69	208	2	,463
3	33	209	1	,470
4	15	209	1	,361
5	79	209	1	,409
6	33	209	1	,473
7	47	208	2	,500
8	20	209	1	,398
9	51	209	1	,501
10	82	208	2	,383
11	44	209	1	,497
12	70	209	1	,460
13	69	208	2	,465
14	88	209	1	,325
15	18	209	1	,387
16	90	209	1	,295
17	37	209	1	,485
18	4	209	1	,203
19	90	208	2	,296
20	56	208	2	,498
21	7	209	1	,251
22	66	209	1	,476
23	38	208	2	,487